# Counterfactual Dependence Is Not Sufficient for Causation

Jens Jäger

**Abstract**

The project of reducing causation to counterfactual dependence—initiated over half a century ago by David Lewis—remains influential today. Over time, the approach has evolved: initial and relatively simple, yet counterexample-prone, reductions have given way to increasingly sophisticated analyses of causation in terms of structural equation models. This paper argues that all extant counterfactual dependence analyses of causation—including those in terms of structural equation models—are false. For none correctly handle synchronic laws, which generate counterfactual dependence without causal dependence.

## Contents

**1 Previous Arguments Against Sufficiency** **2**

**2 Synchronic Laws** **7**

2.1 Warm-Up: Mirror World . . . . . 7

2.2 Gauss's Law . . . . . 8

2.3 Time Loops . . . . . 15

**3 Troubles for Counterfactualist Accounts of Causation** **22**

3.1 Against Lewis (1973a) and Hall (2007) . . . . . 22

3.2 Against Model-Theoretic Definitions of Causation . . . . . 23

3.3 No Easy Fix . . . . . 28

**4 Conclusion** **31**

**A Proof: Counterfactualist SEM Accounts & Sufficiency** **31**

# Introduction

MORE than half a century has passed since David Lewis's seminal "Causation" (1973). To this day, the project of analyzing actual causation in terms of counterfactual dependence remains alive and, by many accounts, well. While the project encompasses a variety of approaches, those approaches are widely regarded to be united by one common principle. Here are Beckers and Vennekens (2017, p. 2, my emphasis):

> "The currently most prominent approaches to defining actual causation are those within the counterfactual dependence tradition, which started with Lewis (1973a). *All of these approaches take as their starting point the assumption that counterfactual dependence is sufficient for causation*, but not necessary (Hitchcock (2001); Woodward (2003); Hall (2004; 2007); Halpern and Pearl (2005); Halpern (2016); Weslake (2015) ...)."

The italicized principle is appealing. If I hadn't hit the bullseye, I wouldn't have won. If my laptop didn't have enough juice, I'd soon be sitting in front of a black screen. If you weren't reading this sentence, you wouldn't know what it says. So, one concludes, my hitting the bullseye *causes* my winning, my laptop's ample charge *causes* its continued operation, your reading the sentence *causes* your knowing what it says. Generalizing:

> **Sufficiency:** Necessarily, if $(c, e)$ is a suitable pair of actual events such that $e$ wouldn't have occurred if $c$ hadn't occurred, then $c$ causes $e$.

The "necessarily" modal witnesses the fact that counterfactualist reductivist consider the sufficient condition part of a *definition* of causation. The restriction to "suitable" event pairs, meanwhile, fends off otherwise easy counterexamples to the principle. For example, my hitting the bullseye doesn't *cause* my hitting the board's center circle—they are the same event. Yet if I hadn't hit the bullseye, I wouldn't have hit the board's center circle. One standard requirement on "suitability" is therefore that $c$ and $e$ be "distinct" events, meaning here that neither is a part of the other, and that neither's occurrence logically entails the other's occurrence (cf. Lewis (1973; 1979)). We'll soon see other suitability requirements.

First we have to deal with another class of easy counterexamples. Counterfactual conditionals are notoriously context-sensitive. Lewis (1979, p. 458) famously distinguishes between "standard" and "backtracking" contexts: suppose that earlier today you returned Susy's book, as you promised you'd do. Given that Susy is known to take promises seriously, the following conditional seems true:

(1) If I hadn't returned the book to her today, Susy would be disappointed in me.

Given (1), **Sufficiency** entails that your returning the book was a cause of your staying in Susy's good graces. So far so good. But now consider that you're extremely reliable and honest, known to never break a promise. With this in mind, you might have reasoned as follows:

(2) If I hadn't returned the book today, that would have been because Susy and I agreed on a later return date to begin with. So Susy wouldn't have been disappointed in me.

You could reasonably assert either (1) or (2), but never their conjunction. This indicates a context shift between (1) and (2). Following Lewis (1979), call the interpretation of the conditional triggered by (1) the "standard" interpretation and the that triggered by (2) the "backtracking" interpretation. Now, given (2), **Sufficiency** entails that your returning the book today is a cause of your agreeing on today as the return date. But that's absurd: you have no such retrocausal powers. For this reason, it's important that the conditional in **Sufficiency** always be evaluated on its standard interpretation.

This paper argues that **Sufficiency**, and those counterfactualist approaches which rely on it, face an existential threat: the possibility of *synchronic laws*. These are laws relating simultaneous distinct events.[1] In Section 1, I explore previous challenges to **Sufficiency** and why, despite them, versions of **Sufficiency** have endured. In Section 2, I then provide two examples of synchronic laws: the first is Gauss's law of classical electrodynamics; the second involves constraints imposed by closed time-like curves on their pasts. I argue that each undermines even weak versions of **Sufficiency**. Section 3 argues that this immediately rebuts several reductivist theories, including Lewis (1973a) and Hall (2007). In Appendix A, I prove that, on the standard counterfactualist reduction of structural equations (due to Hitchcock (2001)), all prominent structural equation accounts entail **Sufficiency**. In Section 4, I conclude that all extant counterfactualist reductions of causation are false.

# 1 Previous Arguments Against Sufficiency

**Sufficiency** has faced previous challenges. Some have argued that *omissions* aren't causes (e.g. Beebee, 2004). The pressure is especially acute for far-flung omissions: Julius Caesar's failure to water my plant isn't a cause of the plant's death, yet my plant's death arguably counterfactually depends on it: if he had watered it, it wouldn't have died. In reply, the **Sufficiency** lover may bite the bullet, while trying to blunt its impact in various ways:

[1] Relativistically speaking: space-like separated distinct events.

she might attempt to partially reduce omissions to "positive" events, i.e. *commissions* (Bernstein, 2014), or to explain the appearance of non-causation as mere infelicity (Schaffer, 2005). But, failing that, she can retreat and strengthen her notion of "suitability": she may stipulate that an event pair is suitable only if its first element is a *positive* event.[2]

A demand for proportionality in causation poses another challenge to **Sufficiency**. I greet my neighbor loudly, and she startles. My *greeting loudly* causes the startle, but my greeting *simpliciter* doesn't—my neighbor isn't *that* jumpy. Yet, if I hadn't greeted her *simpliciter*, my neighbor wouldn't have startled. So counterfactual dependence isn't sufficient for causation. One might resist this line by reinterpreting the demand for proportionality: one might appeal to pragmatics, insisting that mentions of "$A$ causes $B$" tend to carry an implicature that $A$ is a *maximally specific* cause of $B$. Or one could try to separate causation from explanation and shift the burden of proportionality over to the explanatory side (cf. Weslake (2017)). But, failing that, the **Sufficiency** advocate can retreat and strengthen her notion of "suitability": she may stipulate that suitable event pairs consist of *proportional* events.

A third challenge emerges from David Lewis's own semantics for counterfactuals in deterministic worlds. On his miracles-based semantics, intended to model the "standard" resolution of the counterfactual conditional, the near past is still counterfactually different. According to Lewis, counterfactual antecedents are preferentially brought about by *small* miracles (cf. (Lewis, 1979)), but small miracles need time to snowball into big change. So, where antecedents dictate big differences to actuality, they must include a significant delay between the occurrence of the miracle and the antecedent event—a delay during which the counterfactual world differs from actuality. Bennett (2003) calls this delay a "ramp". The need for counterfactual ramps yet again raises the specter of retrocausation.

To illustrate by example: you throw a ball at me; I notice just in time and catch it. If I hadn't caught the ball, surely that wouldn't be because at the moment of impact a miracle instantly twisted my arm, missing the ball. Instead, some macroscopic change would have occurred—perhaps I would have noticed the ball only later, or you threw it a littler harder, or a gust of wind deflected the ball outside of my reach. According to Lewis, any of these changes would be brought about by a small miracle—e.g. changes in neuronal firing patterns, or microscopic meteorological changes—which subsequently needs time to effect the big change. But surely my catching the ball causes neither my actual neuronal firing pattern nor the actual earlier atmospherical state.

Lewis (1979) offers a response. **Sufficiency** says that $c$ causes $e$ if $\neg O(e)$ holds in all clos-

[2] This restriction could also be subsumed under a ban on "overly disjunctive" events, as e.g. (Lewis, 1986b, p. D) discusses; see below.

est worlds where $\neg O(c)$. But merely asserting $\neg O(e)$ leaves much undetermined:typircally there's a myriad of ways in which $e$ might fail to occur. Lewis's hope is thus that $\neg O(e)$ leaves open the specific content of the counterfactual ramping period:

> "[W]e should sacrifice the independence of the immediate past to provide an orderly transition from actual past to counterfactual present and future. That is not to say, however, that the immediate past depends on the present in any very definite way. There may be a variety of ways the transition might go, hence there may be no true counterfactuals that say in any detail how the immediate past would be if [some given event hadn't occurred]."[3] (Lewis, 1979, p.463)

This response works for our example. As we saw, there are all sorts of reasons I might not have caught the ball—heightened alertness, a harder throw, an altered wind pattern. The hope is that, for any actual positive event $e^*$ preceding $c$, there is a closest possible way of filling out the ramping period in which $e^*$ still occurs. Then $\neg O(c) \Box\!\!\rightarrow \neg O(e^*)$ is false for any such $e^*$.[4]

Vihvelin (1995) identifies two kinds of threats to this response. The first arises when the antecedent event is an *omission*—an omission's negation typically entails definite ramping period. But this is no threat to the weakening of **Sufficiency** to positive events. The second threat stems from overly detailed past events. Let $c$ be *my catching the ball at* $t$, for some time $t$, and let $e$ be the totality of all events during some open time interval bounded, to the future, by $t$. Given that $e$'s non-occurrence would require a miracle to occur shortly before $t$, we have $\neg O(c) \Box\!\!\rightarrow \neg O(e)$. But my catching the ball doesn't *cause* $e$.

One possible response follows Lewis (1986a) in banning overly "fragile" events. Those are events with extremely detailed essences—intuitively, events which could have very easily failed to occur. Lewis's justification for the ban is that our standard way of denoting event propositions, "standard nominalizations", isn't nearly detailed enough to pick out these events. Inspired by Lewis's move, the **Sufficiency** advocate may strengthen "suitability" yet again, additionally requiring that $e$ not be *overly fragile.*

[3]The original quote ends with "...if the past were different". This isn't what Lewis needs here and is also less plausible than my substitution. Sometimes that antecedent may pick up on the non-occurrence of an omission, and thereby engender a rather specific immediate past. For example: "I didn't do my homework yesterday. In a different past—i.e., if the past had been different—I would have done it, and thus my teacher would have lauded me just now." (The original quote thus wouldn't permit the first response to Vihvelin's (1995) objection below.)

[4]The hope requires a counterfactual semantics which permits multiple maximally close antecedent worlds. On a semantics like Stalnaker's (1968), in which a given counterfactual antecedent selects a *unique* closest antecedent world, there's a unique closest ramping period. In this case some positive event $e^*$ preceding $c$ will satisfy $\neg O(c) \Box\!\!\rightarrow \neg O(e^*)$. In response, the **Sufficiency** advocate should insist that, even though the conditional is true, it's not *determinately* true. She may then retreat to a weakened version of **Sufficiency** which requires *determinate* truth of $\neg O(c) \Box\!\!\rightarrow \neg O(e^*)$ for causation.

Lewis's response is unlikely to convince everyone: it doesn't conclusively show that *all* non-occurrences of robust positive events permit eclectic swaths of equally-close ramping periods. Moreover, Lewis doesn't independently characterize the distinction between the robust and the fragile. Some counterfactual reductivists have therefore proposed an alternative: retreating to a variant of **Sufficiency** where the counterfactual conditional doesn't require ramping periods. I'll explain this alternative strategy in the following footnote.[5] Those in the structural equations tradition might have yet other means at their disposal to solving the ramping problem, e.g. via judicious choices of variables. My own arguments will work for any such restrictions on variable choice.

Now, a *fourth* challenge results from *rejecting* counterfactual miracles, in favor of a view on which counterfactual worlds have the same laws. If the laws are deterministic, this requires the counterfactual worlds to be different at all times, including *past* times (cf. Bennett, 1984; Loewer, 2007; Albert, 2015; Dorr, 2016). Now, in a world like ours, with continuous laws, the past will generally only have to differ *microscopically*, until very close to the antecedent's time (cf. Dorr (2016)). Still, ubiquitous retrocausation is implausible even if the effects are microscopic. In my view, the cleanest way for the **Sufficiency** advocate to circumvent the problem is not to get entangled in it to begin with: instead,

[5] Glynn's (2013) account is an example of this. It's a combination of two ideas. First, it stipulates a technical meaning of the counterfactual conditional—let's denote it $\blacksquare\!\!\rightarrow$ —according to which, when $A$ is a proposition purely about particular matters of fact at the instant (or brief interval) $t$, "$A \blacksquare\!\!\rightarrow B$" is evaluated using only miracles *at* $t$. That is, $A \blacksquare\!\!\rightarrow B$ is true iff $B$ is true at all worlds closest among those where (i) $A$ is true, (ii) no miracles occur outside of $t$, (iii) everything prior to $t$ is as it actually is, and (iv) everything after $t$ evolves according to the actual laws of nature. Second, say that *c causes e* if there is some truth $T$ solely about $t$ such that $\neg O(c)$ and $T$ are metaphysically compossible and $\neg O(c) \wedge T \blacksquare\!\!\rightarrow \neg O(e)$. Intuitively, the role of $T$ is to suppress any unwanted effects which the $\neg O(c)$-realizing miracle would otherwise bring about—*viz.* consequences which affect $e$ via causal routes bypassing $c$. Generally, $T$ requires a highly complex miracle. For example, my bus is stuck in traffic, and my being on the bus right now ($c$) causes my being late to the meeting ($e$). Moreover, if I wasn't on the bus right now, I'd be on my bike ($\neg O(c)$), speeding through grid-locked traffic, and arriving on time. On the technical meaning of the counterfactual conditional introduced above, if I was on my bike right now, this would be because a miracle had quasi-instantly teleported me from the bus onto the bike. But such a miracle would have all sorts of undesired byproducts, affecting my arrival time independently of my newly acquired ability to cycle there. For concreteness, suppose the miracle would leave me extremely startled—so much so that I'd crash, thus not arriving on time. Still, we want to say that my being on the bus ($c$) causes my being late ($e$). Glynn's account achieves this as follows. Some $T$ entail that I'm calm and not disoriented, thus negating the unintended consequence. Moreover, Glynn makes it plausible that we can find a $T$ to do this for all unintended consequences (including consequences that arise from the miracles needed to bring about the various suppressors). If so, then if $\neg O(c) \wedge T$, then I'd arrive on time despite the sudden teleport. The account thereby secures the desired causal relation (that my being on the bus causes my being late) without the need for counterfactual ramping periods. The idea would then be to change **Sufficiency** as follows:

> **Sophisticated Sufficiency:** Necessarily, if $(c, e)$ is a suitable pair of actual events such that, for some truth $T$, $\neg O(c) \wedge T \blacksquare\!\!\rightarrow \neg O(e)$, then $c$ causes $e$.

Importantly, however, **Sophisticated Sufficiency** still falls to my counterexample, as I'll explain in fn. 30.

they should stick with Lewis's miracles-based semantics (or variants of it which avoid ramping periods). Indeed, this is what the counterfactual dependence tradition in fact does (cf. Lewis (1979), Hitchcock (2001), Hall (2007), and Glynn (2013)).

(Alternatively, the **Sufficiency** advocate could strengthen her notion of "suitability", restricting it to pairs whose second element (the putative effect) is a *macroscopic* event. The downside is that this is vulnerable to cases with non-continuous laws, where any counterfactual differences would, as a matter of law, have to be macroscopic. For an easy example, imagine a universe with Newtonian gravity but discretized masses. In response, the **Sufficiency** advocate could argue that the case for a fixed-law counterfactual semantics is weaker in such worlds, because one of its key motivations—that only *microscopic* adjustments to the past are required to accommodate a counterfactual antecedent—no longer applies. Be that as it may—since miracle-based semantics are standard within the counterfactualist tradition, I'll just stick with those.)

A final constraint on suitability—obvious enough that it is often left implicit—is that $(c, e)$ is suitable only if some counterfactuals with antecedent $\neg O(c)$ are false. On the standard view, this is just to say that $c$'s occurrence is metaphysically *contingent*. Without this constraint, **Sufficiency** plausibly makes causation too cheap: some future spacetime region's being identical to itself—an event whose non-occurrence plausibly makes counterfactuals with it as antecedent vacuously true—doesn't cause (say) the Big Bang.[6]

**Sufficiency** emerges weakened but still makes substantive predictions: the canonical examples of causation, which also motivate counterfactual reductions, tend to involve positive, proportional, not overly fragile, and contingent events—stone throws, hurling boulders, hurricanes, poisonings, and the like. So, the proponent of **Sufficiency** might still think of herself as occupying a true and substantive position. Unfortunately, as I'll argue, that appearance is illusory: even in its weakened form, **Sufficiency** is false. This is because, in the presence of *synchronic laws*, counterfactual dependence fails to track causal dependence.

[6] Alternatively, one could include the anti-vacuity constraint not as a restriction on suitability, but as a restriction on what counts as an *event*. But this would render pairs $(c, e)$ ineligible even when only $e$'s non-occurrence is vacuity-inducing—I see no good reason for demanding this extra strength.

# 2 Synchronic Laws

## 2.1 Warm-Up: Mirror World

Before I get to the two main examples, I'd like to start with a particularly simple toy example, illustrating the concept of a synchronic law. Due to its toy nature, the present case may well raise some objections—objections which I'll point out, and which (I argue) won't arise with the main examples.

Imagine a variant of John Hawthorne's (2007) *Mirror World*: a world split into two halves which are, *by law*, mirror images of each other.[7]

> **Mirror Law:** At all times, the matter configuration in one half is a mirror image of the matter configuration in the other half.

Since it relates things at the *same* time, I call a law like Mirror Law *synchronic*.

Additionally, suppose each side evolves according to some local dynamical law, say Newton's Second Law supplemented by a local force law. (As it relates things at different times, we may call Newton's Second Law, and dynamical laws in general, *diachronic*.)

Here is a simple argument against **Sufficiency**. Suppose I clap my hands at time $t$ (and so does my mirror image). First premise: since Mirror Law is a law, the following counterfactual is true:

> (**M**) If I hadn't clapped my hands at $t$, my mirror image wouldn't have clapped at $t$ either.

Given (**M**), **Sufficiency** implies that my clapping at $t$ *causes* my mirror image's clapping at $t$. But—second premise—my clapping at $t$ *doesn't* cause my mirror image's clapping at $t$. Rather, my mirror image's clapping at $t$ is already fully caused by events in its (local) past (via Newton's Second Law), and there is no causal overdetermination. Final premise: Mirror World is possible. So, counterfactual dependence, even between suitable and occurrent events, isn't necessarily sufficient for causation. **Sufficiency** is false.

In conversation, I've encountered three different objections to this case. The first denies premise three: Mirror World is *not* possible. The most common version of this views notes that, given Newton's Second Law, Mirror Law is entailed by symmetric initial conditions, and that the latter is *merely accidental*, not a law, and that this implies that Mirror Law isn't a law. The second is directed at premise two: even if Mirror Law is a law, this could only happen if there *is* non-local causation. Its status as a law must be enforced by some sort of

[7]The difference to Hawthorne's example consists in the nomic necessity of the connection between the sides. This variant is also discussed by Luzon (2024).

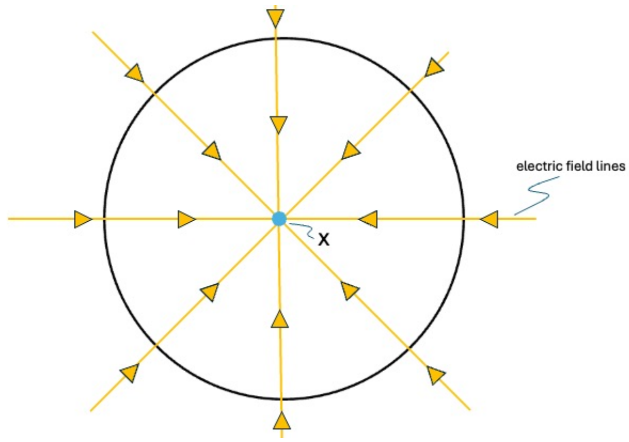



Figure 1: A 2D sketch of (three-dimensional) GAUSS

spooky action at a distance. The third objection consists of a generally dismissive attitude toward toy examples: violating any one of the premises shouldn't count against a theory, because our intuitions about toy examples are unreliable or empirically inaccessible or otherwise unimportant. Spoils to the victor!

There are plausible rejoinders to each objection.[8] But I needn't press those further. None of them arise in the next example.

## 2.2 Gauss's Law

Consider a world—GAUSS—with a single stationary proton at spatial location $x$.[9,10] Besides the particle, there is a static electric field, radiating outward from the proton—see fig. 1 for a 2D sketch. Everything is governed by Maxwell's laws of electrodynamics. Otherwise the world is empty.

Maxwell's laws entail the following synchronic law:

**Gauss's Law:** At all times, the *electric flux* through the boundary of any (spatial)

[8] Regarding the first: even if Mirror Law wasn't a *law*, wouldn't the counterfactual (M) still sound good? The second: this response would have to admit not only ubiquitous causal overdetermination but plausibly also ubiquitous causal cycles—each side causes the other. Also, Gauss's Law (see below) shows that synchronic laws *can* exist without action at a distance. The third: plenty of highly influential cases are highly "exotic", in any ordinary sense of the word—think of Putnam's (1973) Twin Earth, Thomson's (1971) violinist, Chalmer's (1996) p-zombies, Elga's (2000) Sleeping Beauty, Hawthorne's (2007) own mirror world. I find it doubtful that there are principled criteria by which to dismiss (my) Mirror World as too remote without simultaneously dismissing much of the most influential work in contemporary philosophy. Perhaps you're ready to take that step—many won't.

[9] In the following, any references to spatial location, time, sphericalness, isotropy, and other Lorentz non-invariant properties, are relative to a fixed reference frame co-moving with the proton.

[10] To avoid discontinuity in the resulting electric field, assume that the proton has some very small but non-zero spatial extent and that the electric field density falls off continuously near $x$.

volume is proportional to the total electric charge enclosed within the volume.

Intuitively, the electric flux through a boundary is the difference between how much electric field, at the boundary, points *out* of the enclosed volume, versus how much points *into* the enclosed volume. It supervenes entirely on the electric field configuration *at* the boundary: no change in electric flux through a boundary without changing the electric field somewhere on the boundary.[11]

Gauss's Law now says that, if the total electric charge inside some volume is positive—that is, there is more positive electric charge inside the volume than negative electric charge—then the electric flux through the volume's boundary is positive. Equally, if the total electric charge is zero—positive and negative electric charge are exactly balanced—the electric flux is zero. If it is negative, the electric flux is negative. Gauss's Law requires that the electric flux through the boundary "mirror", as it were, the value of the total electric charge within, just like Mirror Law requires that my counterpart's actions mirror mine.

In addition to Gauss's Law (and its analogue for the magnetic field) Maxwell's laws comprise two *diachronic* laws.[12] They make it so that, in a Maxwellian universe, everything is nomically determined by its local past—specifically, by (any spatial cross-section of) its *past light-cone*. Intuitively, an event's past light-cone is the union of all possible spacetime trajectories via which a material particle could reach the event in question.[13]One implication of these laws (together with Gauss's Law) deserves highlighting:

> **Charge Conservation:** Electric charge is conserved—charge is neither spontaneously created nor annihilated.[14]

Here is our argument against **Sufficiency**. Pick some arbitrary time $t$, and some arbitrary (hollow) spatial sphere $S$, at all times centered on the proton. The following counterfactual is true:

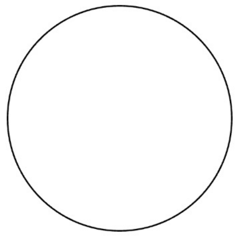
> **($G_1$)** If the proton wasn't present at $x$ at $t$, there would be no charged particles at $t$.

[11]Mathematically, the electric flux through an oriented spatial surface is the integral, over the surface, of the scalar product of electric field and the surface's normal vector.

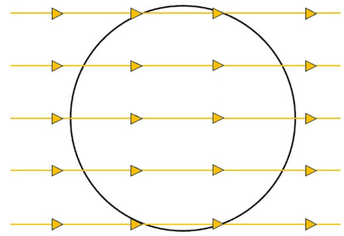
[12]They are *Faraday's Law*—relating the magnetic field's time derivative to the electric field's curl—and *Ampère's Law*—relating the electric field's time derivative to the magnetic field's curl and the electric current. The details won't matter in the following.

[13]Throughout I'll thus understand an event's "light-cone" as a subset of the *manifold*, not its tangent bundle; moreover I mean its *solid* light-cone—a 4D region—not the boundary of a 4D region. What I call "past light-cone" physicists also commonly call "causal past". In the present context, this terminology would potentially be confusing.

[14]Formally, the net *electric current* through the boundary of any volume equals the (temporal) change in electric charge inside the volume. This is an immediate consequence of Gauss's Law and Ampère's Law.



(a) A charge-free solution to Maxwell's laws (2D sketch).



(b) Another charge-free solution to Maxwell's laws (2D sketch)

But **Charge Conservation** would still hold at all times after $t$. So, from $(\mathbf{G_1})$,

> $(\mathbf{G_2})$ If the proton wasn't present at $x$ at $t$, there would be no charged particles present *at all times after* $t$; in particular, $S$ would enclose zero total electric charge at all times after $t$.

Likewise, **Gauss's Law** would still be true. So, from $(\mathbf{G_2})$:

> $(\mathbf{G_3})$ If the proton wasn't present at $x$ at $t$, there would be *zero* electric flux through $S$ at all times after $t$.

See figs. 2a and 2b for sketches of two solutions with zero electric flux through $S$. Since electric flux supervenes on electric field, in both cases the electric field is different from actuality. (For the purpose of the argument, we needn't settle here which of the two—if any—is closest to actuality.)

Given $(\mathbf{G_3})$, **Sufficiency** implies that the proton's presence at $x$ at $t$ *causes* the positive electric flux through $S$ at all times after $t$. But $S$ was arbitrary here. In particular, if we choose some time $t^+$ after $t$, let $S$ be so large that no light signal sent from the proton at $x$ at $t$ could reach $S$ by $t^+$. It follows that the proton at $t$ has faster-than-light causal influence on the electric field at $S$ at $t^+$. But that's false. So **Sufficiency** is false.

Let $p_G$ be the proton's presence at $x$ at $t$, and let $e_G$ be the electric flux through $S$ and $t^+$ being positive. In symbols, $(\mathbf{G_3})$ thus reads:

$$\neg O(p_G) \mathrel{\Box\!\!\rightarrow} \neg O(e_G).$$

The core of the previous argument can be validly expressed as follows (where the modalities are metaphysical):

$1_G$. If **Sufficiency** is true, $(p_G, e_G)$ is a suitable pair of occurring events at GAUSS, $\neg O(p_G) \mathrel{\Box\!\!\rightarrow} \neg O(e_G)$ in GAUSS, and GAUSS is possible, then $p_G$ causes $e_G$ in GAUSS.

$2_\text{G}$. $(p_\text{G}, e_\text{G})$ is a suitable pair of events at GAUSS.

$3_\text{G}$. **Dependence**$_\text{G}$: $\neg O(p_\text{G}) \mathrel{\Box\!\!\rightarrow} \neg O(e_\text{G})$ in GAUSS.

$4_\text{G}$. **Possibility**$_\text{G}$: GAUSS is possible.

$5_\text{G}$. **Non-Causation**$_\text{G}$: $p_\text{G}$ does *not* cause $e_\text{G}$ in GAUSS.

$\therefore$ **Sufficiency** is false.

Premise $1_\text{G}$ follows from the fact that a necessary truth is true in all possible worlds. What about Premise $2_\text{G}$? We've seen several demands the **Sufficiency** advocate might place on "suitability". $p_\text{G}$ and $e_\text{G}$ are certainly *distinct*, occurring as they do in separate spacetime regions. They are also *proportional* to each other, both being simple configurations of physically fundamental properties. Likewise, neither event is overly "fragile": both the proton's being located at $x$ at $t$ and the electric flux through $S$ at $t^+$'s being positive are specifiable by ordinary nominalizations (as we just did), and so aren't "fragile" in Lewis's intended sense. More generally, particle locations and electric fluxes seem like prime examples of physical facts we'd like to *causally explain*—any theory of causation which excluded them would be seriously incomplete.

I'll now turn to premises $3_\text{G}$ – $5_\text{G}$.

### 2.2.1 Defending *Dependence*$_\mathbf{G}$

We've already laid out the positive argument for **Dependence**$_\text{G}$: starting from $(\mathbf{G_1})$, we appeal to **Charge Conservation** to get to $(\mathbf{G_2})$, and then via **Gauss's Law** to $(\mathbf{G_3})$ (which is what **Dependence**$_\mathbf{G}$ says).

In conversations I've encountered three objections, two concerning the appeal to **Gauss's Law** in the last step, and one against the very first step, the assumption of $(\mathbf{G_1})$. They all claim that I'm misconstruing what goes on in the counterfactual, proton-less world. I'll take them in turn.

The first objection holds that, if the proton wasn't present at $x$ at $t$, the electric field would be 0 at $x$ at $t$, but *otherwise unchanged* (except perhaps for some local smoothing). Thereafter a sphere of vanishing electric field, centered on $x$, would expand at light speed. This restores local counterfactual dependence, as the electric field outside of $(t, x)$'s future light-cone would remain unchanged.

Now, the first thing to note is that, because there would be no charges after $t$, this objection violates **Gauss's Law** at all times after $t$. It thus posits an temporally infinite counterfactual miracle, spanning $t$'s entire future. But the miracle's infinity isn't the

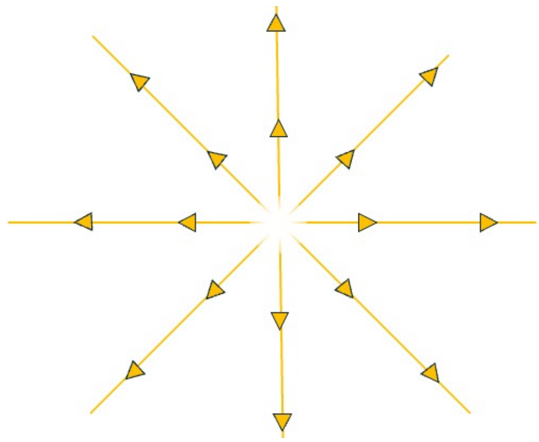


Figure 2: The future of the universe if the proton is deleted at $(t, x)$ and Gauss's Law suspended forever after.

problem. For **Dependence**$_G$ *also* requires an infinite counterfactual miracle, a spatial one: to *preserve* Gauss's law in the absence of charges, the electric field would have be different from actuality at every spatial radius from $x$ at $t$. Infinite miracles are required either way. What's still problematic, though, is that the counterfactualist tradition often explicitly rejects miracles reaching beyond the antecedent's time of occurrence. For example, Lewis (1979) wants the counterfactual world, post antecedent, to evolve according to the actual laws of nature. Glynn's (2013) account even explicitly confines miracles to the antecedent's time of occurrence. So, a temporally infinite future miracle seems incompatible with the letter of many counterfactualist approaches in a way that a spatially infinite miracle doesn't.

But there is a more devastating problem for the present objection: the envisioned scenario—zero electric field at $(t, x)$, with a sphere of vanishing field expanding at light-speed thereafter—has no basis at all in Maxwell's laws. Once we suspend **Gauss's Law**—as the counterfactual scenario demand— the remaining three Maxwell equations entail that the electric field would simply be static forever after $t$. That is, with its configuration at $t$ being as in figure 2, it would maintain that configuration forever after $t$. To see this, one has to dig into the equations, which I'll do in the following footnote.[15] So, the objection is unmotivated.

[15] For the purpose of defining derivatives, let $t$ be an (arbitrarily small) open interval. From the fact that **Charge Conservation** holds at all times after $t$ and the fact that there are no charges present in $t$, it follows that there are no charges present at all times after $t$ (cf. fn. 18); *a fortiori*, there is no *electric current* at any time after $t$. Moreover, by hypothesis, the electric field is radially symmetric around $x$ at all times, and hence its curl vanishes at all times. Since the magnetic field vanishes prior to $t$, we thus have, by Faraday's law ($\text{curl}(\mathbf{E}) = -\frac{\partial \mathbf{B}}{\partial t}$) that it always vanishes. Hence, via Ampere's law for vanishing current ($\text{curl}(\mathbf{B}) = \frac{\partial \mathbf{E}}{\partial t}$), the electric field is static after $t$.

This leads us straight into the second objection (also against the usage of **Gauss's Law** in the third step). What's the problem with thinking that the electric would indeed be static? After all, it would still avoid non-local counterfactual dependence: the electric field outside of $(t, x)$'s future light-cone would be unchanged. Now, granted, it inherits the same problems about future miracles as the previous objection does. But its conclusion can at least be motivated from the remaining Maxwellian laws. Unfortunately, it is independently problematic for the counterfactualist. For it rules out non-local counterfactual dependence only at the price of ruling out virtually *all* counterfactual dependence. Outside perhaps of a small neighborhood around $x$, the electric field is counterfactually invariant to the proton's presence. This makes it hard to see how a counterfactualist account of causation could deliver the eminently plausible result that the proton's presence at $x$ at $t$ is a cause of the electric field *in* $(t, x)$*'s future light-cone*.[16]

Onto the third objection. It proposes to reject $(\mathbf{G_1})$ in favor of the following subjunctive:

> $(\mathbf{G_1^*})$ If the proton wasn't present at $x$ at $t$, there'd be some positive charge elsewhere at $t$.

One variant of this objection has the proton located somewhere else at $t$. Another variant has it replaced by a tiny hollow sphere of positive charge, centered on $x$ at $t$, and expanding at light speed thereafter. (In this scenario, the sphere of vanishing electric field, expanding at light speed, is consistent with **Gauss's Law**.)

But $(\mathbf{G_1^*})$ already gives up the game for **Sufficiency**. For it entails

> $(\mathbf{G_1^{**}})$ If the proton wasn't present at $x$ at $t$, it would not be the case that $x$'s complement is neutrally charged everywhere at $t$.

From $(\mathbf{G_1^{**}})$ and **Sufficiency** (and the suitability of the relevant event pair[17]), it thus follows that the proton's presence at $x$ at $t$ causes $x$'s complement's charge neutrality at $t$. But that's false—on Maxwellian electrodynamics, (subluminal) electric charges only affect their future light-cones (see also 2.2.3). So, **Sufficiency** is false. [18]

[16]Now, of course, everyone already concedes that counterfactual dependence isn't necessary for causation. But necessity violations occur specifically in cases of preemption or overdetermination—in cases where, besides a given actual cause, there is a backup (actual or non-actual) cause, ready to cause the effect in the absence of the former. The present case is not like that.

[17]The pair is (the proton's being present at $x$ at $t$, $x$'s complement's being charge-neutral at $t$). This is clearly suitable: the events are distinct, they are proportional to each other, the former is positive, and the latter not overly detailed.

[18]There's a technicality about the second step—from **Charge Conservation** to $(\mathbf{G_2})$—worth addressing. If $t$ was a closed interval—e.g., a single instant—**Charge Conservation** wouldn't have the desired implication: in a world with charged particles at all times outside of $t$, and none *at* $t$, **Charge Conservation** could still hold at all times outside of $t$. (Since diachronic laws involve temporal derivatives, for such a law to "hold"

### 2.2.2 Defending *Possibility*$_\mathbf{G}$

Maxwell's laws of electrodynamics are clearly metaphysically possible (and possibly laws), as is the existence of a single proton and a static electric field. Now, are these things also *com*possible? A nomic reductivist might complain: from the perspective of a Humean best-system account of lawhood—the most popular nomic reductivist approach—Maxwell's laws are too complicated to be laws at GAUSS. Due to its simplicity, the actual particle-field configuration in GAUSS can be specified by a very simple system of propositions. Such a system purchases much more strength than Maxwell's laws for comparable or less complexity.

But the Humean's worries can be assuaged: simply add to GAUSS complex electrodynamical systems far, far away from our particle, with kinematics compatible with Maxwell's laws. The Humean will agree that, in this new world, Maxwell's laws are laws. Yet, the argument against **Sufficiency** goes through just as smoothly in the new world as in GAUSS.

### 2.2.3 Defending *Non-Causation*$_\mathbf{G}$

**Non-Causation** reflects received scientific opinion on Maxwellian electrodynamics. Two representative quotes, from standard textbooks, on Maxwellian electrodynamics:

> "The displacement between causally related events is always timelike." (Griffiths, 1981, p. 531)
>
> "[I]f any change takes place in one of the interacting bodies, it will influence the other bodies only after the lapse of a certain interval of time. *It is only after this time interval that processes caused by the initial change begin to take place in the second body*." (Landau and Lifschitz, 1994)

There is a powerful argument for the received opinion: similarly to Mirror World, the actual electric field configuration over $S$ at $t^+$ is completely determined by (any spatial cross-section of) its past light-cone—a light-cone doesn't doesn't contain the proton-at-$t$. Thus, to deny Premise 4 would be to stipulate ubiquitous and particularly egregious causal

at a time means for it to be true in some open neighborhood around that time. But in the imagined world, where $t$ is closed, every time outside of $t$ has an open neighborhood in which the temporal derivative of charge density equals the current's divergence—**Charge Conservation** holds everywhere outside of $t$.) To remedy this, let $t$ simply be a small open interval, rather than a single instant. Then **Charge Conservation** holds "for all times after $t$" only if it holds in an open interval around $t$'s future boundary. Then, if $t$ contains no charges, its future doesn't contain any, either. For convenience, I'll continue to treat $t$ as an instant in the main text; but nothing substantive will hinge on this.

overdetermination—particularly egregious, because theoretically costly (since non-local) while at the same time explanatorily void.

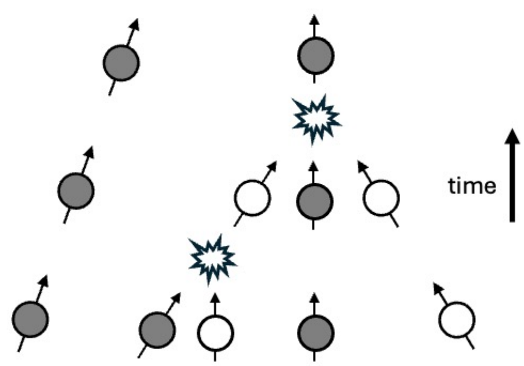
## 2.3 Time Loops

So much for Gauss's law. Our second example involves synchronic laws of a different character, grounded in part in the world's global topology. The topological feature in question are *time loops*—roughly, trajectories through spacetime which travel back in time to their starting point.[19]

Consider a world full of marbles, some grey and others white. Upon contact, the marbles fuse with each other, and the shade of the outgoing marble (i.e., the fusion product) is determined by the shade of the incoming marbles, as follows:

- If exactly an **odd** number of incoming marble are white, the outgoing marble is white.
- Otherwise, the outgoing marble is grey.

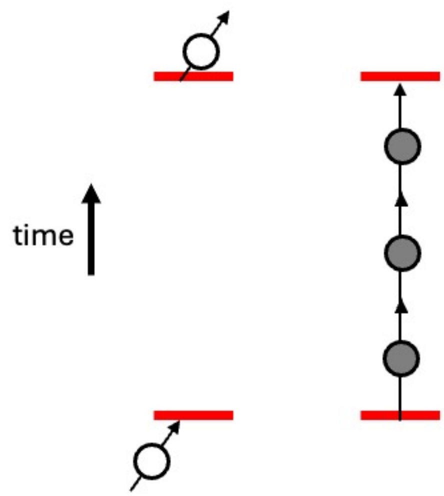
Outside of fusions, a marble's shade is always preserved. Here is a 2D sketch of a world abiding by these rules (with fusion events indicated by jagged bubbles):



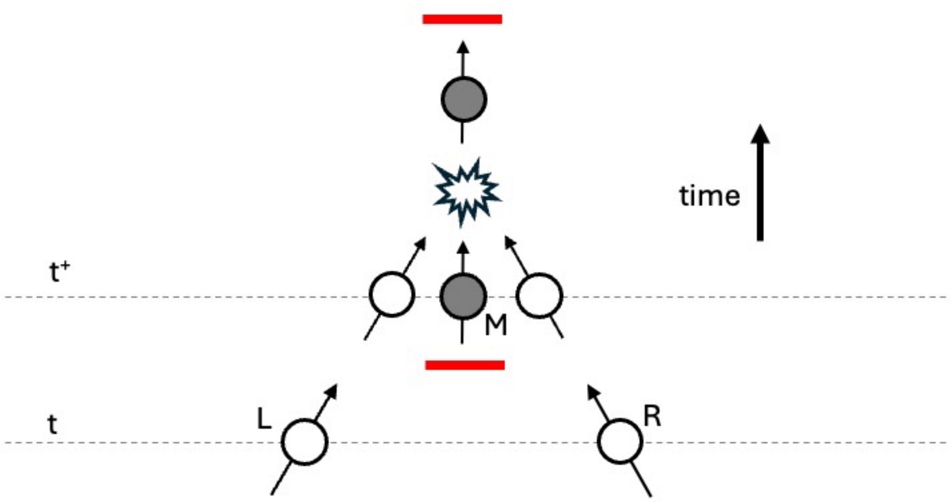

We can construct a *wormhole* by starting with an ordinary spacetime and then "identifying" two regions of space at distinct times. More specifically, the resulting topology is such that any trajectory which enters the earlier region from the past exits at the later region, and any trajectory which enters the later region from the past exits at the earlier region. A 2D illustration of both sort of processes:[20]

[19]In a relativistic spacetime, these are *closed causal curves* (where "*causal*" has its technical meaning here: *having everywhere either time-like or null tangent vector*).

[20]Depending on the initial spacetime structure, the choice of spatial regions may be highly non-unique. For example, if the underlying spacetime is Minkowskian, then, given any pair of (non-intersecting) duplicate bounded space-like surfaces, any other pair of such surfaces with the same boundaries will generate the




Now consider LOOP, a marble world with a wormhole, as follows:




That is, two white marbles approach the region between the wormhole ends. As they pass the earlier end, a grey marble exits. The three marbles eventually collide, fusing into a single marble, later entering the later wormhole end. Let's introduce three binary variables $L$, $R$, and $M$, to represent the shades of the left and right marble at $t$ and the shade of middle marble at $t^+$, respectively (0 = grey, 1 = white). So, actually, $L = R = 1$ and $M = 0$. The fusion law implies the following relationship (where $\overline{\phi}$ abbreviates $(1-\phi)$):

$$M = LRM + \overline{L}\overline{R}M + \overline{L}R\overline{M} + L\overline{R}\overline{M} \tag{1}$$

same wormhole. In this case, the surfaces indicated in the sketch represent but one arbitrary choice among the set of these pairs. To preserve manifoldness (and thus keep derivatives everywhere well-defined) the regions' (2D) boundaries are also removed.

The equation has exactly four solutions:

| $L$ | $R$ | $M$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 0 | 1 |
| 1 | 1 | 0 |
| 1 | 1 | 1 |

So, we have the following law:

> **Agreement Law:** The left marble is white iff the right marble is white, and grey iff the right marble is grey.

Relating distinct but simultaneous events, **Agreement Law** is a synchronic law.

Again, we have a straightforward argument against **Sufficiency**, as follows. Since **Agreement Law** is a law, the following subjunctive holds:

> **(L)** If the left marble had been grey at $t$ ($L = 0$), the right marble would have been grey at $t$ ($R = 0$).

But the actual events ($L = 1$ and $R = 1$) clearly form a suitable event pair—they are distinct, proportional, $L = 1$ is positive, and $R = 1$ isn't overly detailed. So, by **Sufficiency**, the left marble's being white at $t$ *causes* the right marble's being white at $t$. But that's false. So, **Sufficiency** is false.

As a valid argument:[21]

$1_L$. If **Sufficiency** is true, $(L = 1, R = 1)$ is a suitable pair of occurring events at LOOP, $\neg O(L = 1) \mathrel{\Box\!\!\rightarrow} \neg O(R = 1)$ in LOOP, and LOOP is possible, then $L = 1$ causes $R = 1$ in LOOP.

$2_L$. $(L = 1, R = 1)$ is a suitable pair of occurring events at LOOP.

$3_L$. **Dependence**$_L$: $\neg O(L = 1) \mathrel{\Box\!\!\rightarrow} \neg O(R = 1)$ in LOOP.

$4_L$. **Possibility**$_L$: LOOP is possible.

$5_L$. **Non-Causation**$_L$: $L = 1$ does *not* cause $R = 1$ in LOOP.

∴ **Sufficiency** is false.

Premise $1_L$ again follows from the fact that a necessary truth is true in all possible worlds. Premise $2_L$ we've already covered. Let's take the remaining premises in order.

[21] Where the modalities are again metaphysical.

### 2.3.1 Defending *Dependence*$_L$

**Dependence**$_L$ follows from the conjunction of three counterfactual conditionals:

(a) If $L = 0$, the positions and velocities of all particles at $t$ would be the same.

(b) If $L = 0$, then the dynamics in $t$'s future would be the same.

(c) If $L = 0$, then the spacetime structure in $t$'s future would be the same.

By the identity rule (i.e., $\vdash A \;\Box\!\!\rightarrow A$) and Agglomeration,[22] (a), (b), and (c) jointly entail

(d) If $L = 0$, then it would be that $L = 0$, the positions and velocities of all particles at $t$ would be unchanged, and the spacetime structure and dynamics in $t$'s future would be the same.

But (d)'s consequent logically entails that $R = 0$. Hence:

(d) If $L = 0$, then it would be that $R = 0$.

Why believe premises (a)–(c)? Start with (a). The **Sufficiency** lover can't plausibly hold that, if $L = 0$, then some particle's position or velocity would be different. For she would then be committed to saying that particles' current position or velocity are *caused* by $L = 0$. If this includes particles other than the left particle itself, this reintroduces synchronic action at a distance. In any case, it also runs into the familiar overdetermination problem: particles' positions and velocities are, at all times, fully causally explained by their past positions and velocities. Particle shade, by contrast, has no causal influence on particle kinematics.

Premise (b) follows from canonical formulations of a miracle-based semantics (a semantics which, recall, **Sufficiency** advocates should and do embrace). On the canonical formulation due to Lewis (1979)—as well as alternatives like Glynn (2013)—miracles are confined to *no later* than the time of the antecedent.[23] This is for good reasons: miracles placed after the antecedent time can easily prevent relevant effects. In 1983, Stanislav Petrov prevented a nuclear war by correctly judging an incoming missile warning to be a false alarm. But we can easily generate a counterfactual world in which Petrov's judging

[22] I.e. the rule $(A \;\Box\!\!\rightarrow B) \wedge (A \;\Box\!\!\rightarrow C) \vdash (A \;\Box\!\!\rightarrow (B \wedge C))$, part of any standard logic of counterfactuals, including Lewis (1973b) and Stalnaker (1968).

[23] Now, Elga (2000b) has shown (in my view conclusively) that Lewis's (1979) particular "hierarchy of importance" fails to produce the desired asymmetry of miracles. So I'm focusing here on what Lewis professes to rather than actually delivers—that is, I grant that there's *some*, as yet unspecified, semantics which produces the desired asymmetry of miracles. There is no analogous problem with Glynn's (2013) semantics, which confines miracles to exactly the time of the antecedent.

otherwise nonetheless didn't lead to nuclear war, e.g. by suppressing the signal from nuclear button to missiles. If this world is among the closest button-pressing worlds, counterfactual dependence accounts of causation will struggle to vindicate that Petrov's correct judgment prevented nuclear war. But it's hard to see a principled case for permitting post-antecedent miracles in LOOP, which doesn't also carry over to this case.

The argument for Premise (c) is much the same, for topology changes have the same history-altering power as miracles do—the signal in the nuclear cable could be swallowed up by an aptly placed temporary singularity in the cable. Again, it's hard to a principled reason for avoiding this if post-antecedent topology changes are allowed in LOOP.

### 2.3.2 Defending *Possibility*$_{\mathbf{L}}$

As far as worlds with time loops are concerned, LOOP is nothing special: *if* time loops are possible, so is LOOP. But I think there are strong reasons to think that time loops *are* (metaphysically) possible, and little reason to think they aren't.

The strongest reason for the possibility of time loops is the existence of well-understood spacetime models that contain closed time-like curves (time loops in the language of Lorentzian manifolds). One way to articulate this is via positive conceivability—roughly, conceivability that doesn't merely involve the absence of contradiction but also the presence of a "positive picture" of the scenario (Chalmers, 2002). The positive conceivability-possibility link holds that if a situation can be positively conceived, it is metaphysically possible. (Emphasis on *positive* conceivability sidesteps counterexamples that trouble simpler links; for instance, while the falsity of unprovable mathematical truths—Goldbach's conjecture, perhaps—may be naively conceivable, their truth is nonetheless necessary.) Fully interpreted spacetime models[24] with closed time-like curves are paradigm cases of positive conceivability, presenting as they do precise and detailed positive pictures.[25] This is a strong positive reason for the possibility of time loops.

Moreover, initial worries *against* time loops have now been (to my mind) convincingly refuted. One of the more influential worries has traditionally stemmed from paradoxes involving *ability*. Let *autoinfanticide* be the act of a future self's (permanently) killing her own infant self. You can't possibly commit autoinfanticide. But if time loops are possible,

[24] Such as Gödel (1949), Carter (1968), or van Stockum (1938).

[25] "Fully interpreted" is doing work here. For consider the debate on haecceitism in spacetime: do diffeomorphically equivalent Lorentzian manifolds represent genuinely distinct possibilities? (See Norton, Pooley, and Read (2023) for an overview of the debate) This question arises because it's not fully settled what swaps of mathematical points are supposed to represent. But the representational aspects of LOOP which matter to us—the meaning of spacetime trajectories looping back to their origin—*are* fully interpreted. That's all we need.

then (it seems) you *can* possibly commit autoinfanticide: simply travel back in time, gun in hand. It would thus seem that, if time loops are possible, then a contradiction is possibly true: that you both can and can't commit autoinfanticide. So time loops aren't possible.

I find the standard reply to this, due to Lewis (1976), convincing. Seemingly contradictory yet individually acceptable utterances typically indicate a *context shift*. And so it is here: what you "can" do is highly context-sensitive. To quote Lewis's example: compared to a (non-human) ape, I *can* speak Finnish: I have sufficiently developed articulators. But compared to a Finnish speaker, I *can't* speak Finnish: I don't know any Finnish vocabulary or grammar. Similarly, I *can* kill the infant, insofar as "I have what it takes": I have a loaded gun, I'm a good shot, etc. But I *can't* kill the infant, considering that the infant is my younger self. There's no contradiction here: the second statement is evaluated on different contextual facts.

Another objection concerns the "bootstrapping" aspect of time loops, with some arguing that their inexplicability renders them impossible (Al-Khalili, 1999). However, as Lewis (1976, p. 148) notes, inexplicability does not imply impossibility. The universe's initial state (if it has one), outcomes of stochastic processes, and God are all arguably inexplicable, yet possible. For other versions of the no-bootstrapping worry, and rebuttals against them, see also Effingham (2020, Ch. 5.2.2).

### 2.3.3 Defending *Non-Causation*$_{\mathbf{L}}$

**Non-Causation**$_{\mathbf{L}}$ is motivated by a similar thought as its GAUSSian analogue: the right particle's shade at $t$ is already fully causally explained by its shade at preceding times. To posit additional causal influence from the *left* particle at $t$ would be to posit ubiquitous and egregious causal overdetermination.

The situation in LOOP adds an additional twist. If you deny **Non-Causation**$_{\mathbf{L}}$, you must think that present causal facts hinge on what happens in the far future—in particular, if the future contains time loops or not. This violates the following plausible constraint on causation: what causes what up to a time is intrinsic to the world's history up to that time. More precisely:

> **Weak Intrinsicness:** Let $w$ and $w'$ be worlds with the same laws and with identical histories up to time $t$.[26] Then, for any events $c$ and $e$ in $w$ occurring up to time $t$, if it's true at $w$ that $c$ causes $e$, then it's true in $w'$ that $c$ causes $e$.

[26] Here "identical" means *numerically identical*—i.e., $w$ and $w'$ overlap up to $t$. One can also formulate a version of this principle in terms of qualitative duplication, which will be friendly to those who think that worlds don't overlap. That principle will just be slightly more cumbersome to state, but the relevant conceptual content the same.

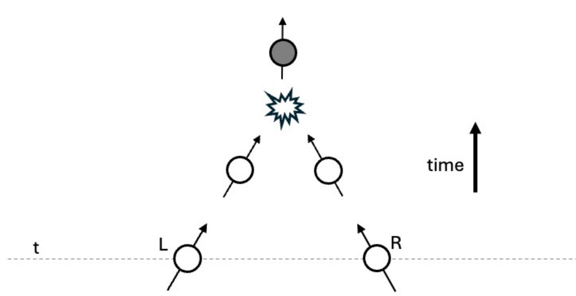
Others before me have defended similar principles. Hall (2004) proposes the following stronger principle:

> **Intrinsicness (Hall):** Let world $w$ contain $S$, a "structure of events that consists of $e$, together with all of its causes [in $S$] back to some arbitrary earlier time $t$" (ibid., p. 239). Let $c$ be some cause of $e$ in $w$. Then, if $w'$ has the same laws as $w$ and contains $S$,[27] then $c$ causes $e$ in $w'$.

Hall's Intrinsicness principle is stronger because it merely requires that $w$ and $w'$ share a small subset of $e$'s history, namely $e$'s causes up to some prior time $t$.

As Hall (2007) elsewhere points out, his own principle produces awkward results in certain canonical scenarios. But my weaker principle avoids these results. Consider *Switching*: suppose a cable carries current to a lamp, where I have the option to flip a switch that redirects the current via a different cable, to the same lamp. Clearly, my decision to leave the switch where it is doesn't cause the lamp to light—current would reach the lamp either way. Yet in an identical world where the alternative cable is grounded (so only the original cable connected to the lamp), my leaving the switch where it is *does* cause the lamp to light. But since the lamp, left cable, and switch are identical in both worlds, and constitute the causes of the lamp's lighting in the alternative world, Hall's **Intrinsicness** principle wrongly predicts that my decision to leave the switch be causes the lamp to light. **Weak Intrinsicness** avoids this error by taking into account the complete temporal history of my decision, including how both cables are connected. Similar comments apply to cases of *Threat Cancelation* (cf. Hall (2007)). So Hall's cases undermine **Intrinsicness** but not **Weak Intrinsicness**.

But if you accept **Weak Intrinsicness**, you should accept **Non-Causation$_\mathbf{L}$**. For consider the following loop-free world, sharing LOOP's history up to shortly after $t$:




In this world, $L = 1$ obviously doesn't cause $R = 1$ But then it follows by **Weak Intrinsicness** that $L = 1$ doesn't cause $R = 1$ in LOOP either.

[27] Hall also considers further strengthenings, where $w'$ merely contains a structure "similar" to $S$. These are, of course, subject to the same counterexamples as the current principle.

# 3 Troubles for Counterfactualist Accounts of Causation

## 3.1 Against Lewis (1973a) and Hall (2007)

So, there can be determinate, non-causal counterfactual dependence, even between distinct (and otherwise suitable) events. What does that mean for counterfactualist reductions of causation? Any account which entails **Sufficiency** should be rejected. At least two such accounts come to mind.

The most famous is Lewis (1973a). It says that $c$ causes $e$ iff there is a chain of actual events $d_1, ..., d_n$ with $d_1 = c$ and $d_n = e$ such that, for all $i = 1, ..., n-1$, $(d_i, d_{i+1})$ is a suitable event pair and $\neg O(d_i) \mathrel{\Box\!\!\rightarrow} \neg O(d_{i+1})$. In particular, then, if $\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e)$ for a suitable pair $(c, e)$ of actual events, $c$ causes $e$, and so Lewis's account entails **Sufficiency**. Our argument against Lewis's account joins the ranks of many previous objections raised against it—notably its failure to handle cases of late preemption and symmetric overdetermination. However, our argument also applies to successor theories which handle these cases.

One of these successors is Hall (2007). According to it, $c$ causes $e$ iff $(c, e)$ is a suitable pair of actual events and there is a "*reduction*" of the actual world in which $c$ counterfactually depends on $e$. It needn't concern us what exactly a reduction is;[28] what matters here is that every world counts as a reduction of *itself* (Hall, 2007, p. 127). Thus we have again, that, where $(c, e)$ is a suitable pair of actual events, $c$'s counterfactually depending on $e$ is sufficient for $c$'s causing $e$—we have, that is, **Sufficiency**. So Hall's (2007) account should be rejected too.[29,30]

[28]Just to give a flavor: roughly, it's a situation in which zero or more parts of the world that are actually in a "non-default" (or "deviant") state adopt their default state instead, while the rest is unchanged.

[29]In Hall's defense, he is aware of these limitations, explicitly bracketing the case of causal loops (p. 114, esp. fn. 6). But this doesn't change the fact that his account isn't a satisfactory analysis of causation.

[30]Earlier I discussed Glynn's (2013) account (fn. 5). It's easy to see that it, too, succumbs to our two counterexamples. The counterfactual situation in GAUSS only has "late" miracles anyway—disappearing the proton and changing the field values exactly at $t$—with everything before and after $t$ evolving according to the actual laws. So we straightforwardly have $\neg O(p_G) \mathrel{\blacksquare\!\!\rightarrow} \neg O(e_G)$.

As for LOOP, the existence of a global time order is a prerequisite of Glynn's account; so, for argument's sake, let's grant that we can identify times across the two strands of spacetime. Moreover, shrink $A$ to a single point, so that it's part of a single time $t$; and let $B$ occur strictly after $t$. In evaluating $A = 0 \mathrel{\blacksquare\!\!\rightarrow} \ldots$ according to Glynn, we then only consider counterfactual worlds with miracles at $t$. But in the closest such world where $A = 0$ we must thus have $B = 0$: by stipulation, there is no miracle after $B$, and hence $A = 0 \wedge B = 1$ would lead to contradiction. So we have $A = 0 \mathrel{\blacksquare\!\!\rightarrow} B = 0$, i.e. $\neg O(L = 1) \mathrel{\blacksquare\!\!\rightarrow} \neg O(R = 1)$. So Glynn's account wrongly entails action at a distance in both GAUSS and LOOP.

## 3.2 Against Model-Theoretic Definitions of Causation

In part due to ongoing issues with preemption and overdetermination, the traditional Lewisian approach has nowadays largely been supplanted by more sophisticated analyses of causation in terms of *structural equations models* (SEMs). An SEM uses variables to represent sets of events and so-called "structural equations" to represent determination relations between these events. Structural equations are expressions of the form $\ulcorner X := f_X(Y_1, ..., Y_n)\urcorner$ (where $X, Y_1, ..., Y_n$ are variables taking real values and $f_X$ is a function from $n$-tuples of real numbers into real numbers). For any argument $Y$ in which $f_X$ is non-constant, the equation is interpreted to say that $Y$'s value (partially) *determines* $X$'s value—hence the asymmetric symbol ":=". Formally, an SEM is a pair $(\mathcal{V}, \mathcal{E})$ of a set of variables $\mathcal{V}$ and a set of structural equations $\mathcal{E}$ in those variables. Given a set $\mathcal{E}$, the set $\mathcal{V}$ bipartitions into a set $\mathcal{V}_{\text{ex}}$ of *exogenous* variables—those which aren't determined by any other variables in the model[31]—and into a set $\mathcal{V}_{\text{en}}$ of *endogenous* variables—those which *are* determined by other variables in the model.

A few more standard definitions are in order. A *solution* of an SEM $(\mathcal{V}, \mathcal{E})$ is an assignment of values to all variables in $\mathcal{V}$ that is consistent with the conjunction of all equations in $\mathcal{E}$. For any $X \in \mathcal{V}$ and value $x$, $\ulcorner(\mathcal{V}, \mathcal{E}) \vDash X = x\urcorner$ then says that, in all possible solutions of $(\mathcal{V}, \mathcal{E})$, $X = x$. Let $\ulcorner(\mathcal{V}, \mathcal{E})(X \leftarrow x)\urcorner$ denote the SEM with variable set $\mathcal{V}\backslash\{X\}$ and the structural equation set $\mathcal{E}(X \leftarrow x)$ resulting from $\mathcal{E}$ by deleting $X$'s structural equation (if any) and replacing every remaining occurrence of $X$ by $x$. When $\mathcal{V}_{\text{ex}} = \mathbf{v}$, we say that $C = c$ *depends on* $E = e$ *in* $\mathcal{M}$ iff $\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e$ and $\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c') \vDash E = e'$ for some $c' \neq c$ and $e' \neq e$. So, intuitively, $C = c$ depends on $E = e$ in $\mathcal{M}$ if manually setting $E$ to some value other than $e$ also changes the value of $C$ in the model. Finally, where $(\mathcal{V}, \mathcal{E})$ is an SEM with $V, W \in \mathcal{V}$, a *directed path from* $V$ *to* $W$ *in* $(\mathcal{V}, \mathcal{E})$ is a sequence of variables $(X_1, ..., X_n)$ such that $X_1 = V$ and $X_n = W$ and, for all $i = 1, ..., n-1$, $f_{X_{i+1}}$ is non-constant in $X_i$—that is, $f_{X_{i+1}}$'s value depends non-trivially on $X_i$'s value for some assignment of values to $\mathcal{V}\backslash\{X_i, X_{i+1}\}$.

Now, prominent SEM accounts of causation—e.g. Hitchcock (2001), Menzies (2004), Halpern and Pearl (2005), and Halpern (2016)—entail that the following is a sufficient condition for causation:[32]

[31] That is, exogenous variables appear on the left-hand side of a structural equation iff the equation's right-hand side is constant.

[32] Two nuances: Halpern and Pearl (2005) officially only provide a definition of *endogenous* variables' being causes. But this seems like a defect of their account: adequate causal models should faithfully capture also exogenous variables' causal relationships to the rest.

Not a defect is Menzies's (2004) slight deviation from **Sufficiency in Adequate Models**: he evaluates dependence-in-a-model by contrasting $c$ specifically with its *default* alternative. As a result, his account will entail only a weakening of **Acyclic Sufficiency** (see below)—one which additionally requires that $c$'s

**Sufficiency in Adequate Models:** Necessarily, if $(c, e)$ is a suitable pair of actual events and variables $X$ and $Z$ represent alterations of $c$ and $e$, respectively: $c$ is a cause of $e$ if there is an *adequate* SEM $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $X, Z \in \mathcal{V}$ such that (i) there's a causal path from $X$ to $Z$ in $\mathcal{M}$ and (ii) $Z = e$ depends on $X = c$ in $\mathcal{M}$.

"Adequate" is crucial here: to ground causal judgments, an SEM has to faithfully represent the world. What does that mean?

First of all, an adequate SEM mustn't assert outright falsehoods. Above we said that structural equations say that the right-hand side partially *determines* the left-hand side. "Determine" in what sense? Hitchcock (2001), Menzies (2004), Halpern and Pearl (2005, p. 847), and Weslake (2015) all have in mind *counterfactual* determination. Here is Hitchcock (2001, p. 280) (see also Hitchcock (2007, p. 500)):

> "[S]tructural equations encode counterfactuals. For example, $[Z := f_Z(X, Y, ..., W)]$ encodes a set of counterfactuals of the following form:
>
> > If it were the case that $X = x$, $Y = y$, ... , $W = w$, then it would be the case that $Z = f_Z(x, y, ..., w)$."

Similarly, Menzies (2004, p. 822):[33]

> "[The equation $SH := ST$] asserts that if Suzy threw a rock, her rock [would] hit the bottle; and if she didn't throw a rock, her rock [wouldn't have] hit the bottle."

Similar quotes are found in Halpern and Pearl (2005, p.847) and Weslake (2015).

Where $\mathbf{X} = \{X_1, ..., X_k\}$, I write $f_Z(X_1, ..., X_k)$ also as $f_Z(\mathbf{X})$ and, where additionally $\mathbf{x} = \{x_1, ..., x_k\}$, I'll write $X_1 = x_1 \wedge ... \wedge X_k = x_k$ as $\mathbf{X} = \mathbf{x}$. (Henceforth let **bold-face** letters denote sets of variables or values.) I'll choose the convention where $f_Z$ is always a function of *all* variables in $\mathcal{V} \setminus \{Z\}$, while generally being *non-constant* only in a select few of those. (Hitchcock (2001, p. 281) chooses a different convention, where $f_Z$ is a function only of those variables in which it is non-constant. Philosophically this makes no difference,

non-occurrence is default. But this won't affect our argument: the *absence* of a proton is plausibly default in the requisite sense, as are both a marble's being white and a marble's being grey. So, for simplicity, I'll ignore this nuance in the following.

Finally, to see that Hitchcock's (2001, p. 287, 290) proposal entails **Sufficiency in Adequate Models**, note that a directed path $\langle X, Y_1, ..., Y_n, Z \rangle$ is "weakly active" in $\mathcal{M}$ if $Z$ depends on $X$ in $\mathcal{M}$ (Hitchcock, 2001, p. 290). (There are merely some terminological differences: his "appropriate" is my "adequate", and his "causal route" is my "directed path".)

[33] Curiously, Menzies uses indicative conditionals here, even though he means them to be "counterfactuals". I've thus substituted the subjunctive form.

but it turns out that Hitchcock's convention would significantly complicate the notation, especially in the proofs of the Appendix.)

The demand that $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ be true is then just the following demand: whenever $[Z := f_Z(\mathcal{V}\backslash\{Z\})] \in \mathcal{E}$, then for all values $\mathbf{v}$ of $\mathcal{V}\backslash\{Z\}$,

$$\mathcal{V}\backslash\{Z\} = \mathbf{v} \;\Box\!\!\rightarrow Z = f_Z(\mathbf{v}).$$

Moreover, an adequate SEM is supposed to be *exhaustive*: it should contain a structural equation for every variable in the model. (Recall that, as we've defined it, exogenous variables are those whose structural equations have constant right-hand side.) Here is Hitchcock:[34]

> "By the same token, [$Z$'s equation] in $\mathcal{E}$ must always include as arguments any variables in $\mathcal{V}$ upon which $Z$ counterfactually depends, given the values of the other variables. If, for some $x, x', y, z, ..., w$, $f_Z(x, y, ..., w) \neq f_Z(x', y, ..., w)$, then the value of $Z$ does depend upon the value of $X$, and [$f_Z$ is non-constant in $X$]. The correct equation for $Z$ can be arrived at by expressing the value of $Z$ as a function of *all* other variables in $\mathcal{V}$"; then $f_Z$ will be constant in "those variables whose values are redundant given every assignment of values to the other variables." (p. 281)

This demand for exhaustiveness is just the converse of the conditional at the start of this paragraph; putting them together, we get the following biconditional: $[Z := f_Z(\mathcal{V}\backslash\{Z\})] \in \mathcal{E}$ *iff*, for all values $\mathbf{v}$ of $\mathcal{V}\backslash\{Z\}$, $\mathcal{V}\backslash\{Z\} = \mathbf{v} \;\Box\!\!\rightarrow Z = f_Z(\mathbf{v})$.

As a final condition on adequacy, the counterfactualist should require that all pairs of variables in $\mathcal{V}$ be *suitable* and that all value assignments to $\mathcal{V}$ are metaphysically possible.[35] If both of these things are the case, say that $\mathcal{V}$ *is suitable*.

**Counterfactual Adequacy:**[36] Necessarily, $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ is adequate only if: $\mathcal{V}$ is

---

[34] Given his convention, his expression "$Z = f_Z(Y, ..., W)$ is not in $\mathcal{E}$" means what, on my convention, "$f_Z$ is non-constant in $X$" means. For the same reason, we translate his "and then eliminating those variables..." as "$f_Z$ will be constant in those variables...."

[35] Specifically, what matters is (the weaker demand) that all value assignments to any $|\mathcal{V}| - 1$-element subset of $\mathcal{V}$ can serve as antecedents in non-vacuous counterfactuals.

[36] It's worth noting that this notion of adequacy immediately satisfies the additional *minimality* condition Hitchcock (2001, p. 280) lays out:

> "Equations in $\mathcal{E}$ must always be written in minimal form: [if $f_Z$ is the right-hand side of $Z$'s structural equation and] for all values $x, x'$ of $X$ and $\mathbf{v}$ of $\mathcal{V}\backslash\{X, Z\}$, $f_Z(x, \mathbf{v}) = f_Z(x', \mathbf{v})$, then the value of $Z$ does not depend upon the value of $X$ at all" (p. 280, notation adjusted).

For if, for all $x, x', \mathbf{v}$, $f_Z(x, \mathbf{v}) = f_Z(x', \mathbf{v})$, then, by the left-to-right direction of **Counterfactual Adequacy**: for all $x, x', z$, if $X = x \wedge \mathcal{V}\backslash\{X, Z\} = \mathbf{v} \;\Box\!\!\rightarrow Z = z$, then $X = x' \wedge \mathcal{V}\backslash\{X, Z\} = \mathbf{v} \;\Box\!\!\rightarrow Z = z$.

suitable and, for all $Z \in \mathcal{V}$, $[Z := f_Z(\mathcal{V} \backslash \{Z\})] \in \mathcal{E}$ *iff* for all values $\mathbf{v}$ of $\mathcal{V} \backslash \{Z\}$,

$$\mathcal{V} \backslash \{Z\} = \mathbf{v} \;\Box\!\!\rightarrow Z = f_Z(\mathbf{v}).$$

In the Appendix I prove that, given **Counterfactual Adequacy**, **Sufficiency in Adequate Models** entails that counterfactual dependence is sufficient for causation in the absence of cycles.[37] That is, **Counterfactual Adequacy** and **Sufficiency in Adequate Models** jointly entail the following principle:[38],[39]

**Acyclic Sufficiency:** Necessarily, if $(c, e)$ is a suitable pair of actual events, $X$ and $Z$ are variables representing alterations of $c$ and $e$, respectively, and there is an adequate, acyclic SEM including $X$ and $Z$: if $e$ wouldn't have occurred if $c$ hadn't occurred, $c$ causes $e$.

**Acyclic Sufficiency** entails that, if there's *any* adequate acyclic model in GAUSS whose variables represent alterations of the proton's presence at $x$ at $t$ and of the electric flux's being positive at $S$ at $t^+$, then the former causes the latter. But there better be such models: otherwise it's hard to see how to avoid the distrastrous conclusion that there are ubiquitous, genuine causal loops in GAUSS. Thus, all prominent SEM accounts of causation are plausibly committed to the false conclusion that the proton's presence at $x$ at $t$ causes the electric flux's being positive at $S$ at $t^+$.

Note that, since **Counterfactual Adequacy** merely posits a necessary condition for adequacy, this argument cannot be avoided by further strengthening adequacy. In particular, no additional demands on the richness of an SEM—how many variables it ought to contains, or how fine-grained their values ought to be—will help.

In contrast to GAUSS, LOOP plausibly *does* contain causal loops. Rich enough SEMs—namely whose variable sets contain at least two events within the loop region—will thus generically be cyclic. Since **Acyclic Sufficiency** only concerns the predictions of *acyclic*

[37] The proof assumes CEM—a natural assumption to make for SEM reductions of causation in the absence of cycles.

[38] Given **Counterfactual Adequacy**, an explicit demand for $(c, e)$'s suitability is technically redundant—I'll still include it hear for easier comparison with the other principles.

[39] **Acyclic Sufficiency** is clearly weaker than full **Sufficiency**. Do **Counterfactual Adequacy** and **Sufficiency in Adequate Models** also entail full **Sufficiency**? A reasonable assumption is that any pair of variables representing alterations of suitable events can be embedded in an adequate SEM. If so, then the only obstacle to full **Sufficiency** is the requirement of acyclicity. If and how one might dispense with it, I don't (yet) know. The answer presumably depends on one's preferred SEM account of causation in the presence of cycles—I shall leave this to future work. For now let it be noted that **Acyclic Sufficiency** does its job for GAUSS. In the case of LOOP where, plausibly, cyclic models are adequate, we can independently confirm that **Counterfactual Adequacy** and **Sufficiency in Adequate Models** misfire.

SEMs, one might hope that, by positing additional richness constraints on adequacy, one might be able to avoid trouble in LOOP.

Alas, the hope is in vain: while I don't have a general theorem extending to the cyclic case, it's easy to see that **Counterfactual Adequacy** and **Sufficiency in Adequate Models** still entail the troubling conclusion for LOOP. For consider the model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{L, R, M, N\}$, where $N$ registers the middle particle's shade shortly after the collision. Now, the only nomically permitted solutions are as follows:

| $L$ | $R$ | $M$ | $N$ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 |

Thus the counterfactual dependence of $R$ on $L$ obtains regardless of the values of $M$ and $N$. In particular, if $L = 0 \mathrel{\Box\!\!\rightarrow} R = 0$, then also

$$L = 0 \wedge M = N = 0 \mathrel{\Box\!\!\rightarrow} R = 0. \tag{2}$$

But since actually $L = R = 1$ and $M = N = 0$, we have, by And-to-If,

$$L = 1 \wedge M = N = 0 \mathrel{\Box\!\!\rightarrow} R = 1. \tag{3}$$

So, by **Counterfactual Adequacy**, conditions 2 and 3 entail that $f_R$ depends on $L$—specifically, $f_R(L, M, N) = L$. In a similar fashion, we'll obtain $f_N(L, R, M) = LRM + \overline{LR}M + \overline{L}R\overline{M} + L\overline{RM}$ and $f_M(L, R, N) = N$. It follows that $\mathcal{M}(L \leftarrow 0) \vDash R = 0$ and $\mathcal{M}(L \leftarrow 1) \vDash R = 1$. Note that neither $M$ nor $N$ are exogenous in $\mathcal{M}$. So, by **Sufficiency in Adequate Models**, $L = 1$ is a cause of $R = 1$ in LOOP.

Now, it's easy to convince oneself that adding additional variables won't substantively change the result. **Counterfactual Adequacy** ensures that whatever variables represent the shades of the left and the right particle just before the collision depend on each other in the resulting model. So, notwithstanding the lack of general theorem for the cyclic case, we see that SEM reductions of causation yield the wrong result in LOOP, no matter how rich we require adequate models to be.[40]

[40] Gallow's (2016) more sophisticated counterfactualist theory of adequacy also fails in cases of synchronic laws. Let $\phi$ be the selection function for your favorite semantics of counterfactual conditionals, mapping proposition-world pairs into sets of worlds. For any world $w$ and any given set of variables $\mathbf{X}$, let the $\mathbf{X}$-*closure of $w$ under $\phi$* be the closure of $\{w\}$ under the set of functions $\{\phi(\mathbf{X}' = \mathbf{x}', \cdot) | \mathbf{X}' \subseteq \mathbf{X}$ and $\mathbf{x}$ is in the range of $\mathbf{X}\}$, i.e., the smallest set $W$ such that: (i) $w \in W$ and (ii) if $w' \in W$ and $\mathbf{x}'$ is in the range of some subset $\mathbf{X}' \subseteq \mathbf{X}$,

### 3.3 No Easy Fix

As we saw in Section 1, earlier arguments against Sufficiency could be answered by retreating to weaker, still substantive, versions of the principle. Is a similar strategy available for defusing the threat of synchronic laws? I'll discuss two attempts at this strategy here, both of which fail.

First, one might hope to exclude counterfactual dependencies due to synchronic laws by only considering dependencies between *non-synchronic events.* Specifically, consider the following weakening of **Sufficiency**:

> **Sufficiency***: Necessarily, if $(c, e)$ is a suitable pair of occurring events such that $e$ wouldn't have occurred if $c$ hadn't occurred, and $e$ occurs after $c$, then $c$ causes $e$.

Unfortunately, **Sufficiency*** still fails, because synchronic counterfactual dependencies can still induce dependencies between non-synchronic events. Consider the following modification of GAUSS:

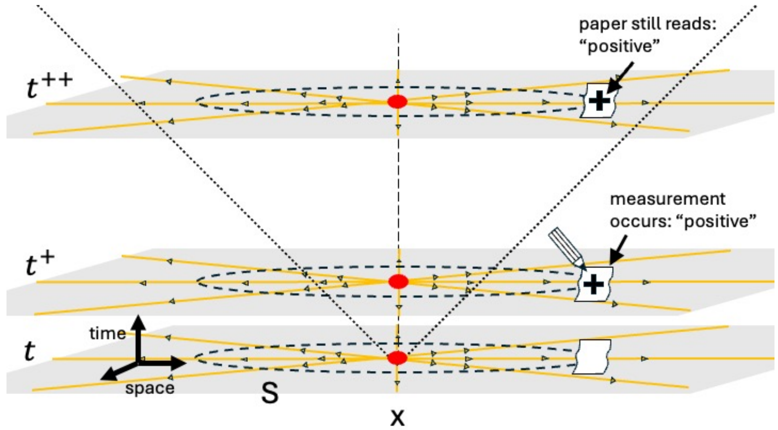
> ***Case.*** **Measurement:** A measurement apparatus is placed somewhere along $S$ at $t^+$, recording the local electric flux through $S$ at $t^+$. It transcribes this result ("positive") on a piece of paper. Remaining at rest, the piece of paper will eventually be located inside of $(t, x)$'s future light-cone, say at time $t^{++}$.

See fig. 3a for a sketch. For concreteness, let's assume that, if the proton had not been located at $x$ at $t$, then the electric field would have been *zero* everywhere on $S$ at $t^+$.[41] So, if the proton hadn't been present at $x$ at $t$, then the flux through $S$ at $t^+$ would have been

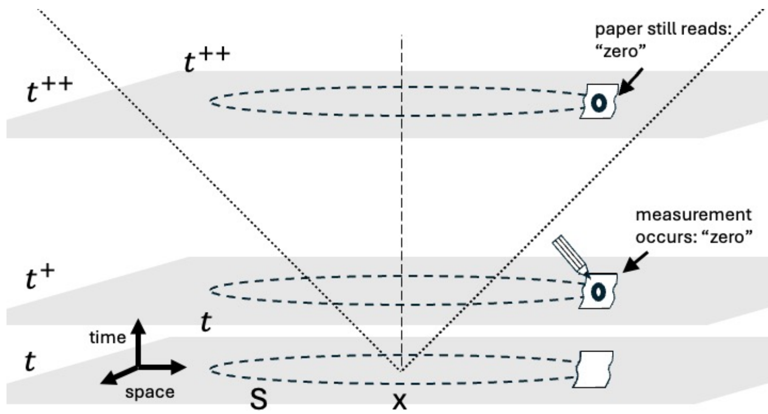
then $\phi(\mathbf{X}' = \mathbf{x}', w') \subseteq W$. Intuitively, the $\mathbf{X}$-*closure of* $w$ *under* $\phi$ is exactly the set of worlds you can reach from $w$ by repeatedly subjunctively supposing $\mathbf{X}' = \mathbf{x}'$—i.e., taking conditionals of the form $\mathbf{X}' = \mathbf{x}' \mathrel{\Box\!\!\rightarrow}$ ...—where $\mathbf{X}'$ is a subset of $\mathbf{X}$. Now, according to Gallow, given a selection function $\phi$, an adequate SEM $\mathcal{M} = (\mathbf{U}, \mathbf{V}, \mathcal{E})$ contains a structural equation $(V := f_V(\mathcal{V}\backslash\{V\})) \in \mathcal{E}$ only if the (ordinary) equation $V = f_V(\mathcal{V}\backslash\{V\})$ is true throughout the actual world's $\mathcal{V}\backslash\{V\}$-closure under $\phi$. The "only if" is strengthened to a biconditional if additionally all variables in $\mathbf{U}$ are mutually counterfactually independent throughout that closure (formally, if no SEM with the same variable set but "strictly more" determination relations (i.e., directed causal paths) satisfies the aforementioned property).

According to this semantics, $\mathcal{M}_G$, if it is to be an adequate SEM for GAUSS, must contain the structural equation $E := P$, where $E$ represents positive electric flux at $S$ at $t^+$ and $P$ the proton's presence at $x$ at $t$. However many subjunctive suppositions of the form $P = i \mathrel{\Box\!\!\rightarrow} \ldots$, for $i = 0, 1$, are nested, Gauss's law would still hold at all times after $t$. Similarly for $\mathcal{M}_L$: however many subjunctive suppositions of the form $L = i \mathrel{\Box\!\!\rightarrow} \ldots$, for $i = 0, 1$, are nested, the dynamics and topological structure downstream from $L$ and $R$ would be unchanged and so we'd still have that $L = R$. So Gallow's theory of adequacy does no better than Hitchcock's when faced with synchronic laws.

[41] Now, recall from our discussion of Gauss's Law that there are zero-flux solutions with (everywhere) *non-vanishing* electric field. What if some such scenario is among the closest $\neg O(p_G)$-worlds? Then it'll not be guaranteed that any *particular* device in **Measurement** would read a different value in the proton's absence. We can address this problem with a mild complication of the case: place a measurement apparatus




(a) A sketch of the situation in **Measurement**.




(b) A sketch of the same situation had the proton not been present at $x$ at $t$.

Figure 3

zero, and hence the paper at $S$ at $t^+$ would have read "zero"; but then also the paper at $S$ at $t^{++}$ would read "zero"—see fig. 3b. By **Sufficiency**$^*$, the proton's presence at $x$ at $t$ thus *causes* the paper's reading "positive" at $t^{++}$. But that's false—the paper is a record of a measurement that took place at $t^+$, *outside* of $(t, x)$'s future light-cone!

(An analogous counterexample can be constructed for LOOP: record the right particle's shade at $t$ on a piece of paper. Assuming the underlying spacetime structure is classical (with an absolute notion of simultaneity), the piece of paper will immediately be located in the left particle's future.[42] **Sufficiency**$^*$ thus entails, wrongly, that the left particle's being

at *every* point along $S$ at $t^+$, with each result transcribed on a separate piece of paper. If the proton had been absent from $x$ at $t$, at least one measurement apparatus would have to give a different reading, and so at least one piece of paper would be different at $t^{++}$.

[42]Otherwise, if the spacetime structure is relativistic, simply wait long enough for the piece of paper to be located in the future light-cone of the left particle at $t$.

white at $t$ *causes* the right particle's being white at $t$.)

Now, another feature of synchronic laws is that they tend to generate *mutual* counterfactual dependence. This is clearest in LOOP, where not only $R = 1$ counterfactually depends on $L = 1$, but also (by an exactly symmetric argument) $L = 1$ counterfactually depends on $R = 1$. Likewise, we may grant that, in GAUSS, if the electric flux through $S$ was 0, no proton would be present at $x$ at $t$. So, as a second stab at the strategy, one might propose the following:[43]

> **Sufficiency**$^\dagger$: Necessarily, if $(c, e)$ is a suitable pair of occurring events such that $e$ wouldn't have occurred if $c$ hadn't occurred, *and* it's not the case that $c$ wouldn't have occurred if $e$ hadn't, then $c$ causes $e$.

But the proposal succumbs to the same counterexamples. For Sufficiency$^\dagger$ to avoid the wrong result in GAUSS, the following counterfactual would have to hold:

> **(?G$^\dagger$)** If the paper didn't read "positive" at $t^{++}$, then the proton wouldn't be present at $x$ at $t$.

But $t$ is much earlier than $t^{++}$, and the counterfactual dependence tradition broadly adopts Lewis's distinction between *standard* and *backtracking* contexts for subjunctive conditionals, cautioning us to evaluate the conditional in **Sufficiency** in the standard context. But, clearly, **(?G$^\dagger$)** is false in the standard context. To bring this out intuitively: suppose $t^{++}$ occurs a week after $t$. Then, at $t^{++}$, you'd express **(?**$G^\dagger$**)** with the following subjunctive:

> **(??G$^\dagger$)** If the paper didn't read "positive" today, then the proton would be absent at $x$ a week earlier.

A true reading of **(??G$^\dagger$)** requires explicit backward reasoning: if the paper didn't read "positive" today, that would have to be because the proton was already absent a week earlier. This parallels familiar backtracking cases: if I didn't return the book today, that would have to be because Susy and I agreed on a later date to begin with.

On a miracles account—again something the counterfactual dependence tradition widely adopts (and should adopt)—the standard context instead only supports the following: if the paper had not read "positive" today, the far past would be unchanged—instead, a small miracle in the immediate past would have deleted the ink, smeared it beyond recognition, incinerated the paper, or something of that sort.

So, neither synchronic laws' synchronicity, nor their tendency to induce mutual counterfactual dependence, can be exploited for easy fixes of **Sufficiency**. Instead, I say, the problem must be addressed closer to its root.

[43] Thanks to **[redacted]** here.

## 4 Conclusion

The preceding discussion suggests that there are two rather different sources of counterfactual dependence: first, counterfactual dependence due to *synchronic* laws (e.g. Mirror Law, Gauss); second, counterfactual dependence due to *diachronic* (or "dynamical") laws.

Examples of diachronic laws are Newton's Second Law ($\mathbf{F} = m \cdot \mathbf{a}$) and two of Maxwell's laws (Faraday's Law and Ampére's Law, cf. fn. 12).[44] They explain how systems evolve over time: given the state of a system at one moment, a diachronic law generates its future (or, more generally, a probability distribution over possible futures).[45] In contrast to the constraints imposed by synchronic laws, this temporal evolution is plausibly *causal*. The foregoing discussion thus suggests that a successful counterfactualist reduction of causation must isolate the diachronic component of counterfactual dependence. How to achieve this is now an open question.

Regardless of the approach taken, I hope to have demonstrated that standing still is not an option. Any *definition* of causation must confront its consequences in worlds with synchronic laws—something which no existing proposal does adequately.

# Appendix

## A Proof: Counterfactualist SEM Accounts & Sufficiency

Throughout this Appendix, I am concerned with acyclic SEMs. I also assume the validity of *Conditional Excluded Middle*—$\vdash (A \mathrel{\Box\!\!\rightarrow} B) \lor (A \mathrel{\Box\!\!\rightarrow} \neg B)$. This is the natural setting for acyclic SEMs, since all value assignments to exogenous variables have unique solutions.

Now recall the counterfactualist notion of SEM adequacy:

> **Counterfactual Adequacy:** Necessarily, $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ is adequate only if: $\mathcal{V}$ is suitable and, for all $Z \in \mathcal{V}$, $[Z := f_Z(\mathcal{V} \backslash \{Z\})] \in \mathcal{E}$ *iff* for all values $\mathbf{v}$ of $\mathcal{V} \backslash \{Z\}$,
>
> $$\mathcal{V} \backslash \{Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z(\mathbf{v}).$$

Throughout the Appendix, "adequate" is assumed to satisfy **Counterfactual Adequacy**,

[44]Other examples include the heat equation, the Navier-Stokes equations of fluid mechanics, and the Schrödinger equation.

[45]*Mathematically*, diachronic laws will tend to take the form of partial differential equations involving time derivatives.

which I'll abbreviate as **CA**.

## A.1 Paths in Adequate Models

In a slight abuse of notation, where $\mathcal{E}$ is a set of structural equations, let $(X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_n) \in \mathcal{E}$ denote the fact that, for each $i = 1, \ldots, n-1$, $f_{X_{i+1}}$ is non-constant in $X_i$ (for some assignment of values to $\mathcal{V} \backslash \{X_{i+1}, X_i\}$). Recall that, when $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ and $(X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_n) \in \mathcal{E}$, I say that $\mathcal{M}$ *contains a directed path from* $X_1$ *to* $X_n$.

We now show that, if there's an *immediate* directed path from one variable to another in an adequate acyclic model, then adding an extra variable to the model either preserves that path or extends it by one variable, unless it creates cycles.

> ***Lemma A1.*** **Path Extension.** For any adequate acyclic model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$, any $C, E \in \mathcal{V}$, and any variable $X$: if $(C \rightarrow E) \in \mathcal{E}$ and $\mathcal{M}' = (\mathcal{V} \cup \{X\}, \mathcal{E}')$ is adequate and acyclic, then either $(C \rightarrow E) \in \mathcal{E}'$ or $(C \rightarrow X \rightarrow E) \in \mathcal{E}'$.

*Proof of Lemma A1*: Let $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ be adequate with $(C \rightarrow E) \in \mathcal{E}$. By **CA**, there are values $c^*, c^{**}, e^*, e^{**}, \mathbf{v}^*$ with $c^* \neq c^{**}$ and $e^* \neq e^{**}$ such that

$$C = c^* \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v}^* \mathrel{\Box\!\!\rightarrow} E = e^*, \text{ and} \tag{4}$$

$$C = c^{**} \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v}^* \mathrel{\Box\!\!\rightarrow} E = e^{**}. \tag{5}$$

Let $\mathcal{M}' = (\mathcal{V} \cup \{X\}, \mathcal{E}')$ be adequate. By CEM, we have two cases—intuitively, they correspond to either $X$ screening off all dependence of $E$ on $C$ or failing to do so:

**Case 1 (No Screening).** There are values $c^\dagger, c^\ddagger$, $e^\dagger$, $e^\ddagger$, $\mathbf{v}$, with $c^\dagger \neq c^\ddagger$ and $e^\dagger \neq e^\ddagger$, and a value $x$ of $X$ such that

$$C = c^\dagger \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v} \wedge X = x \mathrel{\Box\!\!\rightarrow} E = e^\dagger, \text{ and}$$
$$C = c^\ddagger \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v} \wedge X = x \mathrel{\Box\!\!\rightarrow} E = e^\ddagger.$$

By **CA**, it follows that $(C \rightarrow E) \in \mathcal{E}'$.

**Case 2 (Screening).** For every value $x$ of $X$ and $\mathbf{v}$ of $\mathcal{V} \backslash \{C, E\}$, there is a value $e_{x,\mathbf{v}}$ of $E$ such that, for every value $\gamma$ of $C$,

$$C = \gamma \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v} \wedge X = x \mathrel{\Box\!\!\rightarrow} E = e_{x,\mathbf{v}}. \tag{6}$$

We first prove that $(X \rightarrow E) \in \mathcal{E}$. The rule *Conjunction Shift*—$(A \mathrel{\Box\!\!\rightarrow} (B \wedge C)) \vdash ((A \wedge$

$B) \boxright C)$—is valid in any standard logic for counterfactuals (including Stalnaker (1968) and Lewis (1973b)).[46]Given condition 4, CEM, and Conjunction Shift, there is a value $x'$ of $X$ such that

$$C = c^* \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \wedge X = x' \boxright E = e^*. \tag{7}$$

By condition 6,

$$C = c^{**} \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \wedge X = x' \boxright E = e^*. \tag{8}$$

*Negation Transfer*—$(A \boxright C), (A \wedge B \boxright \neg C) \vdash (A \boxright \neg B)$—is also valid in any standard logic for counterfactuals.[47] From 5, 8, and Negation Transfer,

$$C = c^{**} \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \boxright E = e^{**} \wedge X \neq x',$$

and so, by CEM and Conjunction Shift, there is a value $x'' \neq x'$ such that

$$C = c^{**} \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \wedge X = x'' \boxright E = e^{**}. \tag{9}$$

From 9 and 6,

$$C = c^* \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \wedge X = x'' \boxright E = e^{**}. \tag{10}$$

Finally, by **CA**, 7, and 10, $(X \to E) \in \mathcal{E}$.

Second, we prove that either $(C \to X) \in \mathcal{E}$ or $(E \to X) \in \mathcal{E}$. Let $x$ be any value of $X$. Since $e \neq e'$, either $e_{x,\mathbf{v}^*} \neq e^*$ or $e_{x,\mathbf{v}^*} \neq e^{**}$. Suppose $e_{x,\mathbf{v}^*} \neq e^*$. Then, from conditions 4 and 6, Negation Transfer, and Agglomeration,

$$C = c^* \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \boxright (X \neq x \wedge E = e^*),$$

Thus, by CEM, there is an $x^* \neq x$ such that

$$C = c^* \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \boxright (X = x^* \wedge E = e^*). \tag{11}$$

Suppose, for contradiction, that

$$C = c^{**} \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \boxright X = x^*.$$

[46]Given Agglomeration (see below), Conjunction Shift is equivalent to *Cautious Monotonicity*, $(A \boxright B), (A \boxright C)) \vdash ((A \wedge B) \boxright C)$.

[47]Given CEM, Negation Transfer is the contrapositive of *Rational Monotonicity*, $((A \boxright B) \wedge \neg(A \boxright \neg C)) \supset ((A \wedge C) \boxright B)$.

Then, by 5,

$$C = c^{**} \wedge \mathcal{V}\backslash\{C,E\} = \mathbf{v}^* \;\Box\!\!\rightarrow (X = x^* \wedge E = e^{**}). \tag{12}$$

By Conjunction Shift, 11 and 12 give us, respectively,

$$C = c^* \wedge \mathcal{V}\backslash\{C,E\} = \mathbf{v}^* \wedge X = x^* \;\Box\!\!\rightarrow E = e^*), \text{ and}$$
$$C = c^{**} \wedge \mathcal{V}\backslash\{C,E\} = \mathbf{v}^* \wedge X = x^* \;\Box\!\!\rightarrow E = e^{**},$$

in contradiction with condition 6. So, there is a $x^{**} \neq x^*$ such that

$$C = c^{**} \wedge \mathcal{V}\backslash\{C,E\} = \mathbf{v}^* \;\Box\!\!\rightarrow X = x^{**}. \tag{13}$$

By Conjunction Shift, 11 entails

$$C = c^* \wedge \mathcal{V}\backslash\{C,E\} = \mathbf{v}^* \wedge E = e^* \;\Box\!\!\rightarrow X = x^*, \tag{14}$$

and 13 and 5 together entail

$$C = c^{**} \wedge \mathcal{V}\backslash\{C,E\} = \mathbf{v}^* \wedge E = e^{**} \;\Box\!\!\rightarrow X = x^{**}. \tag{15}$$

14 and 15 together entail that either $(C \to X) \in \mathcal{E}$ or $(E \to X) \in \mathcal{E}$.[48]

So, we have that $(X \to E) \in \mathcal{E}$ and either $(C \to X) \in \mathcal{E}$ or $(E \to X) \in \mathcal{E}$. In the first case, we have $(C \to X \to E) \in \mathcal{E}$, as desired. In the second case, we have $(E \to X) \in \mathcal{E}$ and $(X \to E) \in \mathcal{E}$, in contradiction with $\mathcal{M}'$'s acyclicity. So, $(C \to X \to E) \in \mathcal{E}$. ■

We say that there is a directed path from a *set* $\mathbf{X}$ to a variable $Y$ iff there is a directed path from some member of $\mathbf{X}$ to $Y$. Lemma A1 entails the following theorem:

[48] For suppose that $(C \to X) \notin \mathcal{E}$; that is, for all $\mathbf{v}, e$, there is a $x_{\mathbf{v},e}$ such that, for all $\gamma$,

$$C = \gamma \wedge \mathcal{V}\backslash\{C,E\} = \mathbf{v} \wedge E = e \;\Box\!\!\rightarrow X = x_{\mathbf{v},e}.$$

Then 15 entails

$$C = c^* \wedge \mathcal{V}\backslash\{C,E\} = \mathbf{v}^* \wedge E = e^{**} \;\Box\!\!\rightarrow X = x^{**},$$

which together with 14 entails $(E \to X) \in \mathcal{E}$.

Analogously, suppose that $(E \to X) \notin \mathcal{E}$; that is, for all $c, \mathbf{v}$, there is a $x[c,\mathbf{v}]$ such that, for all $\varepsilon$,

$$C = c \wedge \mathcal{V}\backslash\{C,E\} = \mathbf{v} \wedge E = \varepsilon \;\Box\!\!\rightarrow X = x[c,\mathbf{v}].$$

Then 14 entails

$$C = c^* \wedge \mathcal{V}\backslash\{C,E\} = \mathbf{v}^* \wedge E = e^{**} \;\Box\!\!\rightarrow X = x^*,$$

which together with 15 entails $(C \to X) \in \mathcal{E}$.

***Theorem A1.*** Let $\mathbf{C} \subseteq \mathcal{V}$ and $E \in \mathcal{V}$ be such that $E \not\in \mathbf{C}$. If there are values $\mathbf{c}, \mathbf{c}', e$ such that $\mathbf{C} = \mathbf{c} \; \Box\!\!\rightarrow E = e$ and $\mathbf{C} = \mathbf{c}' \; \Box\!\!\rightarrow E \neq e$, then in any adequate acyclic model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathbf{C} \cup \{E\} \subseteq \mathcal{V}$, there is a directed path from $\mathbf{C}$ to $E$.

*Proof of Theorem A1*: Let $\mathbf{C}$ and $E$ be such that $E \not\in \mathbf{C}$. We proceed by induction on the number of variables $n$ in $\mathcal{V}$. Our base case is $n = |\mathbf{C}| + 1$. We want to show that, for any adequate acyclic model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \mathbf{C} \cup \{E\}$, there is a directed path from $\mathbf{C}$ to $E$. We prove this base case, in turn, by (sub-)induction on the number $k \leq n$ of variables in $|\mathbf{C}|$.

- Sub-induction on $k$: The base case $k = 1$ for our sub-induction is immediate: given CEM, $C = c' \; \Box\!\!\rightarrow E \neq e$ entails that there is an $e' \neq e$ with $C = c' \; \Box\!\!\rightarrow E = e'$, and given **CA**, it then follows from $C = c \; \Box\!\!\rightarrow E = e$ and $C = c' \; \Box\!\!\rightarrow E = e'$ that there's a path from $\mathbf{C} = \{C\}$ to $E$ in $\mathcal{M}$. For the induction step, $k \mapsto k + 1 \leq n$, assume (as the induction hypothesis) that, for any $\mathbf{C}, E$ with $E \not\in \mathbf{C}$ and $|\mathbf{C}| = k$, and any adequate, acyclic model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \mathbf{C} \cup \{E\}$ such that there are value assignments $\mathbf{c}, \mathbf{c}', e$ with $\mathbf{C} = \mathbf{c} \; \Box\!\!\rightarrow E = e$ and $\mathbf{C} = \mathbf{c}' \; \Box\!\!\rightarrow E \neq e$, there is a directed path from $\mathbf{C}$ to $E$. Let now $\mathcal{M}' = (\mathcal{V}', \mathcal{E}')$ with $\mathcal{V}' = \mathbf{C}' \cup \{E'\}$ and $|\mathbf{C}'| = k + 1$ such that there are value assignments $\mathbf{c}, \mathbf{c}', e$ with $\mathbf{C}' = \mathbf{c}' \; \Box\!\!\rightarrow E' = e'$ and $\mathbf{C}' = \mathbf{c}' \; \Box\!\!\rightarrow E' \neq e$. Pick any $V \in \mathcal{V}' \backslash (\mathbf{C}' \cup \{E'\})$. It follows from the definition of suitability that, whenever a variable set $\mathbf{X}$ is suitable, any subset of $\mathbf{X}$ is suitable. Hence, by **Corollary A1**, if any model with variable set $\mathbf{X}$ is adequate, then for any $X \in \mathbf{X}$, there is an adequate model with variable set $\mathbf{X} \backslash \{X\}$. In particular, there is an adequate model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \mathcal{V}' \backslash \{V\}$. Hence, by the induction hypothesis, $(X_1 \rightarrow X_2 \rightarrow ... \rightarrow X_{n-1} \rightarrow X_n) \in \mathcal{E}$ for some $\{X_i\}_{i=1}^n \subseteq \mathcal{V}$ with $X_1 \in \mathbf{C}$ and $X_n = E$. **Lemma A1** entails that, for all $i = 1, ..., n$, either $(X_i \rightarrow X_{i+1}) \in \mathcal{E}'$ or $(X_i \rightarrow V \rightarrow X_{i+1}) \in \mathcal{E}'$. Hence, for every $i = 1, ..., n$ there is a path from $X_i$ to $X_{i+1}$ in $\mathcal{M}'$, and so there is a path from $X_1$ to $E$, and hence from $\mathbf{C}$ to $E$, in $\mathcal{M}'$.

This proves the base case.

For the (main) induction step, $n \mapsto n + 1$, suppose (as the induction hypothesis) that in any adequate acyclic model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n \geq |\mathbf{C}| + 1$, if $\mathbf{C} \cup \{E\} \subseteq \mathcal{V}$ with $E \not\in \mathbf{C}$, then $\mathcal{M}$ contains a directed path from $\mathbf{C}$ to $E$. Let $\mathcal{M}' = (\mathcal{V}', \mathcal{E}')$ with $|\mathcal{V}'| = n + 1$ and $\mathbf{C} \cup \{E\} \subseteq \mathcal{V}'$ be an adequate acyclic model. Choose any $V \in \mathcal{V}' \backslash (\mathbf{C} \cup \{E\})$. Then, by the same reasoning as before, there is an adequate model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \mathcal{V}' \backslash \{V\}$. By the induction hypothesis, $(X_1 \rightarrow X_2 \rightarrow ... \rightarrow X_{n-1} \rightarrow X_n) \in \mathcal{E}$ for some $\{X_i\}_{i=1}^n \subseteq \mathcal{V}$ with $X_1 \in \mathbf{C}$ and $X_n = E$. By **Lemma A1**, for all $i = 1, ..., n$, either $(X_i \rightarrow X_{i+1}) \in \mathcal{E}'$ or

$(X_i \rightarrow V \rightarrow X_{i+1}) \in \mathcal{E}'$. Hence, for every $i = 1, ..., n$ there is a path from $X_i$ to $X_{i+1}$ in $\mathcal{M}'$, and so there is a path from $X_1$ to $E$, and hence from $\mathbf{C}$ to $E$, in $\mathcal{M}'$. ■

## A.2 Counterfactual Dependence and Dependence in Acyclic Models

For any SEM $\mathcal{M} = (\mathcal{V}, \mathcal{E})$, $\mathcal{V}^{\mathcal{E}}_{\text{ex}}$ and $\mathcal{V}^{\mathcal{E}}_{\text{en}}$ denote $\mathcal{M}$'s exogenous variables and $\mathcal{M}$'s endogenous variables, respectively. Where the model is obvious from context, I'll omit the superscripted $\mathcal{E}$.

For any SEM $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $X \in \mathcal{V}_{\text{en}}$, let $\mathcal{M}_{\rangle X \langle} = (\mathcal{V} \backslash \{X\}, \mathcal{E}_{\rangle X \langle})$ be *the result of collapsing $X$ in $\mathcal{M}$*: where $[X := f_X(\mathcal{V} \backslash \{X\})] \in \mathcal{E}$, $\mathcal{E}_{\rangle X \langle}$ is the result of deleting $[X := f_X(\mathcal{V} \backslash \{X\})]$ from $\mathcal{E}$ and replacing all other instances of $X$ in $\mathcal{E}$ by $f_X(\mathcal{V} \backslash \{X\})$. A standard result about acyclic SEMs is that collapsing preserves entailment relations. It also turns out to preserve adequacy.

> ***Lemma A2.1.*** **Endogenous Collapse:** Let $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ be acyclic with $X \in \mathcal{V}_{\text{en}}$. Then $\mathcal{M}_{\rangle X \langle}$ is acyclic, adequate, and, for all $\mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V} \backslash \{X\}$ and values $\mathbf{y}, \mathbf{z}$ of $\mathbf{Y}$ and $\mathbf{Z}$,
>
> $$(\mathcal{M}(\mathbf{Y} \leftarrow \mathbf{y}) \vDash \mathbf{Z} = \mathbf{z}) \leftrightarrow (\mathcal{M}_{\rangle X \langle}(\mathbf{Y} \leftarrow \mathbf{y}) \vDash \mathbf{Z} = \mathbf{z}).$$

*Proof of Lemma A*2.1: Both $\mathcal{M}_{\rangle X \langle}$'s acyclicity and the biconditional are standard results. It remains to prove $\mathcal{M}_{\rangle X \langle}$'s adequacy.

Left-to-right direction: Suppose $\mathcal{M}$ is adequate. Let $Z \neq X$ with $[Z := f_Z^{\rangle X \langle}(\mathcal{V} \backslash \{X, Z\})] \in \mathcal{E}_{\rangle X \langle}$. Then $[Z := f_Z(\mathcal{V} \backslash \{Z\})] \in \mathcal{E}$ with $f_Z$ either (i) constant in $X$ or (ii) non-constant in $X$. Suppose (i). Then $f_Z(\mathcal{V} \backslash \{Z\}) = f_Z^{\rangle X \langle}(\mathcal{V} \backslash \{X, Z\})$. Since $\mathcal{M}$ is adequate, it follows by **CA** that, for all values $\mathbf{v}$ of $\mathcal{V} \backslash \{Z\}$,

$$\mathcal{V} \backslash \{Z\} = \mathbf{v} \;\Box\!\!\rightarrow Z = f_Z(\mathbf{v}),$$

and hence, for all values $x$ of $X$ and $\mathbf{v}'$ of $\mathcal{V} \backslash \{X, Z\}$,

$$X = x \wedge \mathcal{V} \backslash \{X, Z\} = \mathbf{v}' \;\Box\!\!\rightarrow Z = f_Z^{\rangle X \langle}(\mathbf{v}').$$

By the disjunction rule—$(A \;\Box\!\!\rightarrow C) \wedge (B \;\Box\!\!\rightarrow C) \vdash ((A \vee B) \;\Box\!\!\rightarrow C)$—it follows that, for all values $\mathbf{v}'$ of $\mathcal{V} \backslash \{X, Z\}$,

$$\mathcal{V} \backslash \{X, Z\} = \mathbf{v}' \;\Box\!\!\rightarrow Z = f_Z^{\rangle X \langle}(\mathbf{v}').$$

Suppose instead (ii). Since $\mathcal{M}$ is adequate, it follows by **CA** that, for all values $\mathbf{v}$ of

$\mathcal{V}\backslash\{X\}$,

$$\mathcal{V}\backslash\{X\} = \mathbf{v} \boxright X = f_X(\mathbf{v}). \tag{16}$$

Since $\mathcal{M}$ is acyclic, 16 implies that, for all values $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$ and $z$ of $Z$,

$$\mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \boxright X = f_X(z, \mathbf{v}'). \tag{17}$$

By $\mathcal{M}$'s adequacy, we also have

$$X = f_X(z, \mathbf{v}') \wedge \mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \boxright Z = f_Z(f_X(z, \mathbf{v}'), \mathbf{v}'). \tag{18}$$

From 17, 18, and *Cautious Transitivity*—$((A \boxright B) \wedge ((A \wedge B) \boxright C)) \vdash (A \boxright C)$—we obtain, for all values $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$, and $z$ of $Z$,

$$\mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \boxright Z = f_Z(f_X(z, \mathbf{v}'), \mathbf{v}'). \tag{19}$$

Since $f_Z(f_X(z, \mathbf{v}'), \mathbf{v}')$ is constant in $z$, $f_Z(f_X(z, \mathbf{v}'), \mathbf{v}') = f_Z^{\rangle X\langle}(\mathbf{v}')$. Hence, from 19, for all $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$,

$$\mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \boxright Z = f_Z^{\rangle X\langle}(\mathbf{v}').$$

Right-to-left direction: Conversely, suppose that, for some function $g$ and for all $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$,

$$\mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \boxright Z = g(\mathbf{v}'). \tag{20}$$

Let $[X := f_X(\mathcal{V}\backslash\{X\})] \in \mathcal{E}$. Since $\mathcal{M}$ is adequate, it follows by **CA** that, for all $z$of $Z$ and $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$,

$$Z = z \wedge \mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \boxright X = f_X(z, \mathbf{v}'). \tag{21}$$

Either (i) $f_X(\mathcal{V})$ is constant in $Z$ or (ii) $f_X(\mathcal{V})$ is non-constant in $Z$. Suppose (i). Then, by the Disjunction Rule and 21,

$$\mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \boxright X = f_X(z, \mathbf{v}'). \tag{22}$$

From 20, 22, Aggregation, and Conjunction Shift, for all $z$ of $Z$ and $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$,

$$X = f_X(z, \mathbf{v}') \wedge \mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \boxright Z = g(f_X(z, \mathbf{v}'), \mathbf{v}').$$

Since $\mathcal{M}$ is adequate, it follows by **CA** that $[Z := h(X, \mathcal{V}\backslash\{X, Z\})] \in \mathcal{E}$ with $h|_{X \in \text{range}(f_X)} = g$. Since $[X := f_X(Z, \mathcal{V}\backslash\{X\})] \in \mathcal{E}$ with $f_X$ constant in $Z$, we thus have $[Z := h^{\rangle X\langle}(\mathcal{V}\backslash\{X, Z\})] \in \mathcal{E}_{\rangle X\langle}$. Since $h|_{X \in \text{range}(f_X)} = g$, we have $h^{\rangle X\langle}(\mathcal{V}\backslash\{X, Z\}) = g$, and thus $[Z := g(\mathcal{V}\backslash\{X, Z\})] \in$

$\mathcal{E}_{\rangle X \langle}$.

Suppose instead (ii). Since $\mathcal{M}$ is acyclic, $f_Z$ is constant in $X$. Thus, since $\mathcal{M}$ is adequate, we have by **CA** and the Disjunction Rule that, for all $\mathbf{v}'$ of $\mathcal{V} \backslash \{X, Z\}$ and $x$ of $X$,

$$\mathcal{V} \backslash \{X, Z\} = \mathbf{v}' \mathrel{\Box\!\!\rightarrow} Z = f_Z(x, \mathbf{v}'). \tag{23}$$

From 20, 23, and $\mathcal{V}$'s suitability (cf. fn. 35), $f_Z(x, \mathbf{v}') = g(\mathbf{v}')$ for all $\mathbf{v}'$ of $\mathcal{V} \backslash \{X, Z\}$. Thus, $[Z := g(\mathcal{V} \backslash \{X, Z\})] \in \mathcal{E}$; since $g$ is independent of $X$, $[Z := g(\mathcal{V} \backslash \{X, Z\})] \in \mathcal{E}_{\rangle X \langle}$. ■

> ***Lemma A2.2.*** **Exogenous Substitution.** Let $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ be adequate and acyclic. Then, for any $U \in \mathcal{V}_{\text{ex}}$: $\mathcal{M}(U \leftarrow u)$ is acyclic; if moreover $U = u$, then $\mathcal{M}(U \leftarrow u)$ is adequate.

*Proof of Lemma A2.2*: Since substitution can only *remove* edges, acyclicity is immediate. It remains to prove adequacy.

For any function $f_Z$, let $f_Z^u$ denote the result of replacing all instances of $U$ in $f_Z$ by $u$. Let $U = u$. We want to show that $[Z := f_Z^u(\mathcal{V} \backslash \{U, Z\})] \in \mathcal{E}(U \leftarrow u)$ iff, for all $\mathbf{v}$ of $\mathcal{V} \backslash \{U\}$,

$$\mathcal{V} \backslash \{U, Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z^u(\mathbf{v}).$$

Right-to-left direction: Suppose that, for all $\mathbf{v}$ of $\mathcal{V} \backslash \{U, Z\}$,

$$\mathcal{V} \backslash \{U, Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z^u(\mathbf{v}). \tag{24}$$

Suppose, for contradiction, that for some $\mathbf{v}^\dagger$ and some $z^\dagger \neq f_Z^u(\mathbf{v}^\dagger)$,

$$U = u \wedge \mathcal{V} \backslash \{U, Z\} = \mathbf{v}^\dagger \mathrel{\Box\!\!\rightarrow} Z = z^\dagger. \tag{25}$$

By 24, 25, Negation Transfer, and Conjunction Shift,

$$\mathcal{V} \backslash \{U, Z\} = \mathbf{v}^\dagger \wedge Z = z^\dagger \mathrel{\Box\!\!\rightarrow} U \neq u. \tag{26}$$

Since $U = u$, we have, by And-to-If,

$$\mathcal{V} \backslash \{U, Z\} = \mathbf{v}^* \wedge Z = z^* \mathrel{\Box\!\!\rightarrow} U = u. \tag{27}$$

where $\mathcal{V} \backslash \{U\} = \mathbf{v}^*$ and $Z = z$. By **Theorem 1**, Conditions 26 and 27 entail that there's a directed path from $\mathcal{V} \backslash \{U\}$ to $U$, in contradiction with the assumption that $U \in \mathcal{V}_{\text{ex}}$.

Therefore, we have, for all values $\mathbf{x}$ of $\mathbf{X}$ and $\mathbf{w}$ of $\mathcal{V}\backslash(\mathbf{X} \cup \{U\})$,

$$U = u \wedge \mathcal{V}\backslash\{U, Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z^u(\mathbf{v}). \tag{28}$$

Since $\mathcal{M}$ is adequate, by **CA** condition 28 entails that $[Z := f_Z(\mathcal{V}\backslash\{Z\})] \in \mathcal{E}$ with $f_Z(u, \mathbf{v}) = f_Z^u(\mathbf{v})$ for all $\mathbf{v}$ of $\mathcal{V}\backslash\{U, Z\}$. Hence $[Z := f_Z^u(\mathcal{V}\backslash\{U, Z\})] \in \mathcal{E}(U \leftarrow u)$.

Left-to-right direction: Suppose $[Z := f_Z^u(\mathcal{V}\backslash\{U, Z\})] \in \mathcal{E}(U \leftarrow u)$. Then $[Z := f_Z(\mathcal{V}\backslash\{Z\})] \in \mathcal{E}$ with $f_Z(u, \mathcal{V}\backslash\{U, Z\}) = f_Z^u(\mathcal{V}\backslash\{U, Z\})$. Since $\mathcal{M}$ is adequate, it follows by **CA** that, for all values $\mathbf{v}$ of $\mathcal{V}\backslash\{Z\}$,

$$\mathcal{V}\backslash\{Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z(\mathbf{v}),$$

and hence, for all $\mathbf{v}' \in \mathcal{V}\backslash\{U, Z\}$

$$U = u \wedge \mathcal{V}\backslash\{U, Z\} = \mathbf{v}' \mathrel{\Box\!\!\rightarrow} Z = f_Z^u(\mathbf{v}'). \tag{29}$$

Suppose, for contradiction, that for some $\mathbf{v}^\dagger$ and $\mathbf{y}^\dagger$,

$$\mathcal{V}\backslash\{U, Z\} = \mathbf{v}^\dagger \mathrel{\Box\!\!\rightarrow} Z \neq f_Z^u(\mathbf{v}^\dagger). \tag{30}$$

By Negation Transfer, conditions 29 and 30 entail

$$\mathcal{V}\backslash\{U, Z\} = \mathbf{v}^\dagger \mathrel{\Box\!\!\rightarrow} U \neq u. \tag{31}$$

Since $U = u$, by And-to-If,

$$\mathcal{V}\backslash\{U, Z\} = \mathbf{x}^* \mathrel{\Box\!\!\rightarrow} U = u, \tag{32}$$

where $\mathcal{V}\backslash\{U, Z\} = \mathbf{x}^*$. By **Theorem A1**, conditions 31 and 32 now entail that there is a directed path from $\mathcal{V}\backslash\{U, Z\}$ to $U$ in $\mathcal{M}$, in contradiction with $U \in \mathcal{V}_{\text{ex}}$. So, for all values $\mathbf{v}$ of $\mathcal{V}\backslash\{U, Z\}$,

$$\mathcal{V}\backslash\{U, Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z^u(\mathbf{v}). \blacksquare$$

*Theorem A2.* **Counterfactual Dependence and Dependence in a Model:** Let $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $C, E \in \mathcal{V}$ be adequate and acyclic. Then, if $\mathcal{V}_{\text{ex}}\backslash\{C\} = \mathbf{v}$ and $C = c \mathrel{\Box\!\!\rightarrow} E = e$,

$$\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e.$$

*Proof of Theorem A2*: If $C$ and $E$ are the same variable, then $C = c \boxright E = e$ entails that $c = e$ and so $\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e$ immediately. So assume, in the following, that $C$ and $E$ are different variables.

We proceed by induction on the number $n$ of variables in the model. For the base case, let $n = 2$ and let $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{C, E\}$ be adequate. Suppose $E \in \mathcal{V}_{\text{ex}}$. Then since $C = c \boxright E = e$, we have (by **CA**) that, $C = c' \boxright E = e$ for all values $c'$ of $C$. Thus, by *Modus Ponens*—$A, A \boxright B \vdash B$, valid in any standard counterfactual logic—$E = e$. Hence $\mathcal{V}_{\text{ex}}\backslash\{C\} = \{E\}$ and $\mathbf{v} = e$, and so $\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e$. Now suppose $E \notin \mathcal{V}_{\text{ex}}$. Then $\mathcal{V}_{\text{ex}}\backslash\{C\} = \varnothing$ and $[E := f_E(C)] \in \mathcal{E}$, with $f_E$ non-constant in $C$. Since $C = c \boxright E = e$, we have $f_E(c) = e$, and thus $\mathcal{M}(C \leftarrow c) \vDash E = e$.

For the induction step, assume that for all adequate acyclic models $\mathcal{M}' = (\mathcal{V}', \mathcal{E}')$ with $C, E \in \mathcal{V}'$ and $|\mathcal{V}'| = n \geq 2$: for any $\mathbf{v}', c'$, and $e'$, if $\mathcal{V}'_{\text{ex}}\backslash\{C\} = \mathbf{v}'$ and $C = c' \boxright E = e'$,

$$\mathcal{M}'(\mathcal{V}'_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}', C \leftarrow c') \vDash E = e'.$$

Let now $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $C, E \in \mathcal{V}$ and $|\mathcal{V}| = n + 1$ be adequate and acyclic. Moreover, suppose $\mathcal{V}_{\text{ex}}\backslash\{C\} = \mathbf{v}$, and $C = c \boxright E = e$. Since $n + 1 \geq 3$, there is an $X \in \mathcal{V}\backslash\{C, E\}$. Either $X \in \mathcal{V}_{\text{en}}$ or $X \in \mathcal{V}_{\text{ex}}$.

Suppose $X \in \mathcal{V}_{\text{en}}$. By **Lemma A2.1**, $\mathcal{M}_{\rangle X \langle}$ is acyclic and adequate, and

$$(\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e) \leftrightarrow (\mathcal{M}_{\rangle X \langle}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e).$$

But $\mathcal{M}_{\rangle X \langle}$ has variable set $\mathcal{V}_{\rangle X \langle}$ with $|\mathcal{V}_{\rangle X \langle}| = n$. Thus, by the inductive hypothesis,

$$\mathcal{M}_{\rangle X \langle}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e,$$

and therefore

$$\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e.$$

Suppose instead $X \in \mathcal{V}_{\text{ex}}$. Then

$$\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) = (\mathcal{M}(X \leftarrow \mathbf{v}))\,(\mathcal{V}_{\text{ex}}\backslash\{C, X\} \leftarrow \mathbf{v}, C \leftarrow c).$$

But $\mathcal{M}(X \leftarrow \mathbf{v})$ has variable set $\mathcal{V}\backslash\{X\}$ with $|\mathcal{V}\backslash\{X\}| = n$ and exogenous variable set $\mathcal{V}_{\text{ex}}\backslash\{X\}$. Moreover, by **Lemma A2.2**, $\mathcal{M}(X \leftarrow \mathbf{v})$ is acyclic and adequate. Thus it follows by the inductive hypothesis that

$$(\mathcal{M}(X \leftarrow \mathbf{v}))\,(\mathcal{V}_{\text{ex}}\backslash\{C, X\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e,$$

and therefore

$$\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c) \models E = e. \blacksquare$$

When $\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c) \models E = e$ and $\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c') \models E = e'$ for some $c' \neq c$ and $e' \neq e$, say that *E depends on C in $\mathcal{M}$ relative to $\mathcal{V}_{ex}$*$= \mathbf{v}$; if additionally $\mathcal{V}_{\text{ex}} = \mathbf{v}$, say that *E depends on C in $\mathcal{M}$* (simpliciter).

## A.3 SEM Accounts Entail Sufficiency (in the Absence of Cycles)

Recall:

> **Acyclic Sufficiency:** Necessarily, if $(c, e)$ is a suitable pair of actual events, $X$ and $Z$ are variables representing alterations of $c$ and $e$, respectively, and there is an adequate, acyclic SEM including $X$ and $Z$: then, if $e$ wouldn't have occurred if $c$ hadn't occurred, $c$ causes $e$.

We can now prove the following theorem. Where "adequate" definitionally satisfies **CA**:

> ***Theorem A3.*** **Sufficiency in Adequate Models** entails **Acyclic Sufficiency**.

*Proof of Theorem A3*: Let $(c, e)$ be a suitable pair of actual events such that $\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e)$. Let $X$ and $Z$ be variables representing alterations of $c$ and $e$, respectively, such that there is an adequate, acyclic SEM $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $X$, $Z \in \mathcal{V}$. To show that **Sufficiency in Adequate Models** entails **Acyclic Sufficiency**, we now only need to show that it implies that $c$ causes $e$.

Since $\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e)$, by CEM there is a unique closest possible $\neg O(c)$-world in which $e$ doesn't occur. Where values $X = x$ and $Z = z$ represent $c$'s and $e$'s actual occurrence, respectively, let value $x' \neq x$ represent the non-occurrence of $c$, and $z' \neq z$ the respective non-occurrence of $e$. We then have $X = x' \mathrel{\Box\!\!\rightarrow} Z = z'$. By And-to-If, we moreover have $X = x \mathrel{\Box\!\!\rightarrow} Z = z$.

Let $\mathbf{v}$ be such that $\mathcal{V}_{\text{ex}} = \mathbf{v}$. By **Theorem A2**, $X = x \mathrel{\Box\!\!\rightarrow} Z = z$ and $X = x' \mathrel{\Box\!\!\rightarrow} Z = z'$ entail, respectively, that $\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{X\} \leftarrow \mathbf{v}, X \leftarrow x) \models Z = z$ and $\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{X\} \leftarrow \mathbf{v}, X \leftarrow x') \models Z = z'$. Thus $Z$ depends on $X$ in $\mathcal{M}$. Moreover, by **Theorem A1**, there is a one directed path from $X$ to $Z$ in $\mathcal{M}$. So, by **Sufficiency in Adequate Models**, $c$ causes $e$. $\blacksquare$

## References

Al-Khalili, Jim (1999). *Black Holes, Wormholes and Time Machines*. 1st edition. Bristol, UK ; Philadelphia, PA: Taylor & Francis.

Albert, David Z. (2015). *After Physics*. Cambridge, Massachusetts London, England: Harvard University Press.

Beckers, Sander and Joost Vennekens (2017). "The Transitivity and Asymmetry of Actual Causation". In: *Ergo, an Open Access Journal of Philosophy* 4.

Beebee, Helen (2004). "Causing and Nothingness". In: *Causation and Counterfactuals*. Ed. by L. A. Paul, E. J. Hall, and J. Collins. MIT Press, pp. 291–308.

Bennett, Jonathan (1984). "Counterfactuals and Temporal Direction". In: *The Philosophical Review* 93.1, pp. 57–91.

— (June 2003). *A Philosophical Guide to Conditionals*. Oxford, New York: Oxford University Press.

Bernstein, Sara (2014). "Omissions as Possibilities". In: *Philosophical Studies* 167.1, pp. 1–23.

Carter, Brandon (1968). "Global Structure of the Kerr Family of Gravitational Fields". In: *Physical Review* 174.5, pp. 1559–1571.

Chalmers, David J. (1996). *The Conscious Mind: In Search of a Fundamental Theory (2nd Edition)*. Oxford University Press.

— (2002). "Does Conceivability Entail Possibility". In: *Conceivability and Possibility*. Ed. by Tamar Szabo Gendler and John Hawthorne. Oxford University Press, pp. 145–200.

Dorr, Cian (2016). "Against Counterfactual Miracles". In: *The Philosophical Review* 125.2, pp. 241–286.

Effingham, Nikk (2020). *Time Travel: Probability and Impossibility*. New York: Oxford University Press.

Elga, Adam (2000a). "Self-Locating Belief and the Sleeping Beauty Problem". In: *Analysis* 60.2, pp. 143–147.

— (2000b). "Statistical Mechanics and the Asymmetry of Counterfactual Dependence". In: *Philosophy of Science* 68.3, pp. 313–324.

Gallow, J. Dmitri (2016). "A Theory of Structural Determination". In: *Philosophical Studies* 173.1, pp. 159–186.

Glynn, Luke (2013). "Of Miracles and Interventions". In: *Erkenntnis* 78.1, pp. 43–64.

Gödel, Kurt (1949). "An Example of a New Type of Cosmological Solutions of Einstein's Field Equations of Gravitation". In: *Reviews of Modern Physics* 21.3, pp. 447–450.

Griffiths, David J. (1981). *Introduction to Electrodynamics*.

Hall, Ned (2004). "Two Concepts of Causation". In: *Causation and Counterfactuals*. Ed. by John Collins, Ned Hall, and Laurie Paul. MIT Press, pp. 225–276.

— (2007). "Structural Equations and Causation". In: *Philosophical Studies*.

Halpern, Joseph (2016). *Actual Causality*. MIT Press.

Halpern, Joseph and Judea Pearl (2005). "Causes and Explanations: A Structural-Model Approach. Part I: Causes". In: *The British Journal for the Philosophy of Science* 56.4, pp. 843–887.

Hawthorne, John (2007). "Craziness and Metasemantics". In: *The Philosophical Review* 116.3, pp. 427–440.

Hitchcock, Christopher (2001). "The Intransitivity of Causation Revealed in Equations and Graphs". In: *The Journal of Philosophy* 98.6, pp. 273–299.

— (2007). "Prevention, Preemption, and the Principle of Sufficient Reason". In: *The Philosophical Review* 116.4, pp. 495–532.

Landau, E.M. and L.D. Lifschitz (1994). *The Classical Theory of Fields*.

Lewis, David (1973a). "Causation". In: *Journal of Philosophy* 70.17, pp. 556–567.

— (1973b). *Counterfactuals*. Malden, Mass.: Blackwell.

— (1976). "The Paradoxes of Time Travel". In: *American Philosophical Quarterly* 13.2, pp. 145–152.

— (1979). "Counterfactual Dependence and Time's Arrow". In: *Noûs* 13.4, pp. 455–476.

— (1986a). "Events". In: *Philosophical Papers Vol. II*. Ed. by David Lewis. Oxford University Press, pp. 241–269.

— (1986b). "Postscripts to 'Causation'". In: *Philosophical Papers Vol. Ii*. Ed. by David Lewis. Oxford University Press.

Loewer, Barry (2007). "Counterfactuals and the Second Law". In: *Causation, Physics and the Constitution of Reality: Russell?S Republic Revisited*. Ed. by Huw Price and Richard Corry. Oxford University Press.

Luzon, Bar (2024). "Explaining Content". PhD thesis. New York University.

Menzies, Peter (2004). "Causal Models, Token Causation, and Processes". In: *Philosophy of Science* 71.5, pp. 820–832.

Norton, John D., Oliver Pooley, and James Read (2023). "The Hole Argument". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta and Uri Nodelman. Summer 2023. Metaphysics Research Lab, Stanford University.

Putnam, Hilary (1973). "Meaning and Reference". In: *Journal of Philosophy* 70.19, pp. 699–711.

Schaffer, Jonathan (2005). "Contrastive Causation". In: *Philosophical Review* 114.3, pp. 327–358.

Stalnaker, Robert (1968). “A Theory of Conditionals”. In: *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*. Ed. by Nicholas Rescher. Blackwell, pp. 98–112.

Thomson, Judith Jarvis (1971). “A Defense of Abortion”. In: *Philosophy and Public Affairs* 1.1, pp. 47–66.

van Stockum, W. J. (1938). “IX.—The Gravitational Field of a Distribution of Particles Rotating about an Axis of Symmetry”. In: *Proceedings of the Royal Society of Edinburgh* 57, pp. 135–154.

Vihvelin, Kadri (1995). “Causes, Effects and Counterfactual Dependence”. In: *Australasian Journal of Philosophy* 73.4, pp. 560–573.

Weslake, Brad (2015). “A Partial Theory of Actual Causation”.

— (2017). “Difference-Making, Closure and Exclusion”. In: *Making a Difference: Essays on the Philosophy of Causation*. Ed. by Helen Beebee, Christopher Hitchcock, and Huw Price. Oxford University Press, pp. 215–231.

Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford Studies in Philosophy of Science. Oxford, New York: Oxford University Press.