TECHNICAL REPORT
R-142
MAY, 2026

Productive Actual Causation

Kai-Zhan Lee[1] **Elias Bareinboim**[1]

[1]Causal AI Lab, Columbia University, New York, NY, USA

Abstract

Actual causation asks which events caused a specific outcome, a question central to explainable AI, legal liability, and scientific explanation. Formalizing it requires two concepts that can diverge: *dependence*, whether the effect would change had the cause been different, and *production*, transmission of causal influence through intermediates. No definition within the structural causal model (SCM) framework handles both switch variants, where dependence-based accounts fail, and preemption, where production-based accounts fail. We introduce *productive actual causation*, a definition that unifies dependence and production: every intermediate in a causal chain must recursively be a cause, and the alternative value must fail to replicate at least one. This is, to our knowledge, the first SCM definition to correctly classify all canonical cases that separate prior definitions. We show that distinguishing preemption from switching is a modeling question, not a definitional one: intermediate variables trace the production pathway that separates the two cases. For bounded in-degree k and path length L, our memoized algorithm scales quadratically in the number of variables; we illustrate this on random SCMs with up to 20 variables.

1 INTRODUCTION

When a patient dies after receiving a treatment, which factors were *actual causes* of death? When a loan application is denied, which features *actually caused* the rejection? These are questions of *actual causation*: which specific events in a given situation caused a specific outcome, as opposed to type-level causal claims such as "smoking causes cancer" [Halpern, 2016a, Pearl, 2000a]. Questions of actual causation arise throughout science, jurisprudence, and artificial intelligence, with applications in legal liability [Wright, 1985], explainable AI [Miller, 2019], and algorithmic fairness [Plecko and Bareinboim, 2024]. Despite decades of work [Halpern, 2016a, Pearl, 2009, Woodward, 2003], formalizing actual causation within the structural causal model framework remains an open problem: existing definitions give conflicting verdicts on canonical test cases. Without a sound definition, an XAI system may incorrectly attribute an outcome to a feature that merely redirects it, or fail to identify the feature that actually produced it.

Dependence versus production. Actual causation has been analyzed through two fundamentally different lenses [Hall, 2004, Paul and Hall, 2013]: *dependence*, whether the effect would change had the cause been different [Lewis, 1973], and *production*, transmission of causal influence from cause to effect through a chain of intermediates [Salmon, 1984, Dowe, 2000, Hitchcock, 2001]. Hall [2004] argues these are genuinely distinct concepts that any adequate theory must accommodate. We ground this distinction with two canonical examples.

Example 1.1 (Preemption)**.** Suzy and Billy both throw rocks at a bottle. Suzy's rock arrives first and shatters the bottle; had she not thrown, Billy's rock would have shattered it instead. The outcome does not counterfactually depend on Suzy's throw, yet she is intuitively a cause because her rock produced the shattering. □

Example 1.2 (Switch)**.** A switch S routes current through one of two circuits: $C_1 = S$ follows the switch position and $C_2 = \neg S$ opposes it. The light $Y = C_1 \vee C_2$ flashes whenever either circuit is active, so $Y = 1$ regardless of S. Intuitively, the switch is not a cause: it merely reroutes an inevitable outcome. □

Intuitively, Suzy's throw is a cause of the shattering, as her rock struck the bottle, but the switch is not a cause of the light flashing, as it flashes regardless. Preemption exhibits production without dependence: Suzy's rock produced the shattering, yet the outcome does not counterfactually depend

on her throw. Switches exhibit contingent dependence without production: holding one circuit fixed, the light depends on the switch, yet the switch merely selects which circuit is active without producing the outcome through either one more than the other. HP-style definitions formalize dependence via counterfactual conditions but cannot distinguish a variable that reroutes an inevitable outcome from one that produces it; our definition separates these two cases. No existing definition distinguishes the two cases: Beckers and Vennekens [2018] says neither is a cause; Halpern [2015] and Beckers [2021] say both are.

Prior definitions. The dominant definition, Halpern–Pearl (HP) [Halpern and Pearl, 2001, 2005, Halpern, 2015], tests dependence: $X = x$ is an actual cause if holding some variables at their actual values makes the outcome counterfactually depend on X. HP correctly handles preemption and symmetric overdetermination, but wrongly accepts switches as causes: it lacks a production test and cannot distinguish a variable that merely reroutes an inevitable outcome from one that produces it.

Beckers and Vennekens (BV) [Beckers and Vennekens, 2018] add production via an *asymmetry* principle: the alternative value must not produce the outcome. This correctly rejects switches but wrongly rejects preempting causes, because the backup mechanism also produces the outcome, and the original BV definition assumes binary variables. Beckers' later CNESS definition [Beckers, 2021] relaxes asymmetry using the *necessary element of a sufficient set* (NESS) test [Wright, 1985], recovering preemption but reintroducing switches: a switch always belongs to some sufficient set, so the NESS test never rejects it. The nondeterministic generalization [Beckers, 2025] inherits this defect.

Other approaches require information beyond the SCM: normality orderings [Halpern and Hitchcock, 2015, Halpern, 2015], regularity conditions [Andreas and Günther, 2024], and domain knowledge [Denecker et al., 2019]. This paper restricts to the standard SCM framework; all definitions we compare against are SCM-based.

Our approach. We introduce *productive actual causation*, a definition that unifies dependence and production within the SCM framework. The definition combines two mechanisms. *Compositional production* (C2) requires every intermediate in a causal chain to itself be a cause of the outcome. *Asymmetry* (C3) rejects variables whose alternative value satisfies the same dependence and production conditions. In Ex. 1.1, Suzy's throw satisfies production and asymmetry; in Ex. 1.2, both intermediates are descendants of the switch, leaving no eligible witnesses, so counterfactual dependence (C1) fails and the switch is rejected. Specifically, we establish the following contributions:

1. We propose a definition (Def. 3.3) that, to our knowledge, is the first *within the standard SCM framework* to correctly classify switch variants, preemption, and symmetric overdetermination (Sec. 3).
2. We show that distinguishing preemption from switching is a modeling question: any definition that rejects switches must also reject coarse preemption, and recovering the expected verdict requires intermediates that trace the production pathway or an outcome domain that encodes manner (Sec. 3.5). A file-server scenario with three conflict-resolution policies illustrates this analysis (Ex. 3.7).
3. We provide a memoized algorithm (Alg. 1) that is fixed-parameter tractable in the maximum in-degree k and longest path length L: it runs in time polynomial in n and d for fixed k and L (Thm. 4.2), to our knowledge the first such parameterized tractability result for actual causation. Empirical benchmarks on random SCMs validate scalability to $n = 20$ variables (Fig. 3).

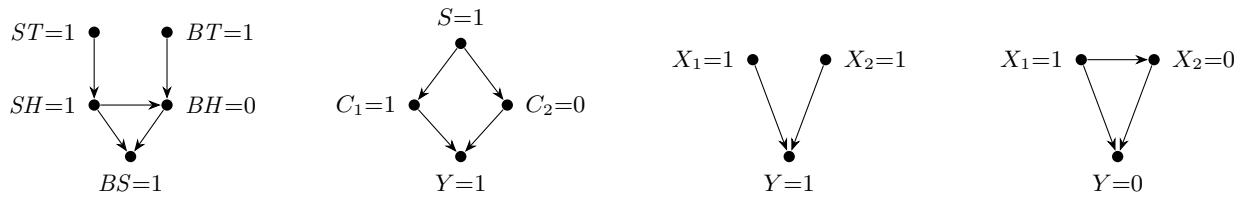
Paper organization. Sec. 2 reviews structural causal models and the Halpern–Pearl definition. Sec. 3 presents the two canonical problems and introduces our definition with correctness proofs. Sec. 4 gives the algorithm with complexity analysis. Sec. 5 concludes. The appendix contains alternative formulations of C3 (App. B), canonical example derivations (App. C), novel examples (App. D), all proofs (App. E), and a discussion of event individuation (App. F).

2 BACKGROUND

2.1 STRUCTURAL CAUSAL MODELS

A *structural causal model* (SCM) [Pearl, 2000b, Bareinboim et al., 2022] is a tuple $\mathcal{M} := \langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{u}) \rangle$. Here $\mathbf{U}$ is a set of exogenous variables, $\mathbf{V}$ is a set of endogenous variables, $\mathcal{F} = \{f_V : V \in \mathbf{V}\}$ is a set of structural functions, and $P(\mathbf{u})$ is a distribution over exogenous variables. Each endogenous variable $V \in \mathbf{V}$ takes values in a finite domain $\mathcal{D}_V$, and each f_V determines the value of V as a function of its endogenous parents $\mathbf{Pa}_V \subseteq \mathbf{V}$ and exogenous parents $\mathbf{U}_V \subseteq \mathbf{U}$. A *causal world* $(\mathcal{M}, \mathbf{u})$ pairs an SCM with a specific assignment $\mathbf{u}$ to the exogenous variables, determining a unique value for each endogenous variable. The *causal graph* $G(\mathcal{M})$ is the directed acyclic graph (DAG) over $\mathbf{V}$ with an edge $V_i \to V_j$ whenever V_i appears as an argument of f_{V_j}.

An *intervention* $do(\mathbf{X} = \mathbf{x})$ replaces the structural functions for variables in $\mathbf{X}$ with constant functions returning $\mathbf{x}$. We write $\mathcal{M}_{X=x'}$ for the submodel obtained by replacing X's structural equation with the constant function returning x'. After the intervention $do(\mathbf{X} = \mathbf{x})$, the *counterfactual* variable $Y_{\mathbf{x}}$ takes the value $Y_{\mathbf{x}}(\mathbf{u})$ in context $\mathbf{u}$; we suppress $\mathbf{u}$ when it is clear from context. For individual variables, $x = X(\mathbf{u})$ and $y = Y(\mathbf{u})$ are the *actual* values of X and Y. An *alternative value* $x' \neq x$ is a competing setting of the



(a) Preemption (b) Switch (c) Overdetermination (d) Shadow variable

Figure 1: Causal graphs for canonical examples. *(a)* Preemption (Ex. 1.1): Suzy's throw preempts Billy's. *(b)* Switch (Ex. 1.2): S routes current through one of two circuits; Y flashes either way. *(c)* Overdetermination (Ex. 3.4): both X_1 and X_2 independently cause Y. *(d)* Shadow variable (Ex. D.1): $X_2 = \neg X_1$, $Y = \neg X_1 \wedge X_2$; X_2 is a dependent intermediate, not a cause of Y.

cause X, and $Y_{x'}(\mathbf{u})$ is the counterfactual it induces. We use $V = v$ for a generic variable and reserve $Y = y$ for when Y is fixed as the effect.

2.2 ACTUAL CAUSATION

Actual causation asks which specific events in a given situation caused a specific outcome [Halpern, 2016a]. The dominant formalization is due to Halpern and Pearl.

Definition 2.1 (HP_m Actual Causation [Halpern, 2015, Def. 2.1])**.** $X = x$ is an actual cause of $Y = y$ in $(\mathcal{M}, \mathbf{u})$ if:

AC1 (Factuality). $X(\mathbf{u}) = x$ and $Y(\mathbf{u}) = y$.

AC2 (Counterfactual dependence). There exist a set $\mathbf{W} \subseteq \mathbf{V} \setminus \{X, Y\}$ and a value $x' \neq x$ such that

$$Y_{x',\mathbf{w}} \neq y,$$

where $\mathbf{w}$ denotes the actual values of $\mathbf{W}$ in $(\mathcal{M}, \mathbf{u})$.

AC3 (Minimality). No strict subset of $\{X\}$ satisfies AC1 and AC2.

The restriction to actual-valued witnesses in AC2 is the key simplification over the original HP definition [Halpern and Pearl, 2005], which allowed arbitrary witness values and required a separate sufficiency condition.

2.3 PRODUCTION-BASED DEFINITIONS

Beckers and Vennekens [2018] propose a production-based alternative grounded in the NESS test [Wright, 1985].

Definition 2.2 (Sufficiency [Beckers, 2021, Def. 4])**.** A set of variable-value pairs L is *sufficient* for $V = v$ w.r.t. $\mathcal{M}$ if the structural equation f_V evaluates to v given the assignments in L, regardless of the values of variables not in L.

Definition 2.3 (Direct NESS [Beckers, 2021, Def. 5])**.** $X = x$ *directly NESS-causes* $Y = y$ in $(\mathcal{M}, \mathbf{u})$ if there exists $\mathbf{W} = \mathbf{w}$ with $(\mathcal{M}, \mathbf{u}) \models X{=}x \wedge \mathbf{W}{=}\mathbf{w}$ such that $\{X{=}x\} \cup \{W_i{=}w_i\}$ is sufficient for $Y = y$ but $\{W_i{=}w_i\}$ alone is not.

Definition 2.4 (NESS [Beckers, 2021, Def. 6])**.** $X = x$ *NESS-causes* $Y = y$ in $(\mathcal{M}, \mathbf{u})$ if there exists a chain $X{=}x = L_1, L_2, \ldots, L_n = Y{=}y$ such that each L_i directly NESS-causes L_{i+1}.

Definition 2.5 (NESS Along a Path [Beckers, 2021, Def. 8])**.** $X = x$ *NESS-causes* $Y = y$ *along a path* p in $(\mathcal{M}, \mathbf{u})$ if the values of the variables in p form a chain of direct NESS causes from $X = x$ to $Y = y$.

Definition 2.6 (BV Actual Causation [Beckers, 2021, Def. 7]; cf. [Beckers and Vennekens, 2018])**.** $X = x$ is a *BV-cause* of $Y = y$ in $(\mathcal{M}, \mathbf{u})$ if:

(i) $X = x$ NESS-causes $Y = y$; and
(ii) there exists $x' \neq x$ such that $X = x'$ does not NESS-cause $Y = y$ in $(\mathcal{M}_{X=x'}, \mathbf{u})$.

Condition (i) requires production: the actual value contributes to the outcome through a chain of sufficient sets. Condition (ii) requires asymmetry: some alternative value fails to produce the outcome.

Definition 2.7 (CNESS Actual Causation [Beckers, 2021, Def. 9])**.** $X = x$ is a *CNESS-cause* of $Y = y$ in $(\mathcal{M}, \mathbf{u})$ if:

(i) $X = x$ NESS-causes $Y = y$ along some path p; and
(ii) there exists $x' \neq x$ such that $X = x'$ does not NESS-cause $Y = y$ along any subpath of p in $(\mathcal{M}_{X=x'}, \mathbf{u})$.

CNESS weakens BV's asymmetry by comparing NESS-causation along specific paths rather than globally.

3 PRODUCTIVE ACTUAL CAUSATION

3.1 SWITCHES VS. PREEMPTION

We formalize the preemption and switch examples from Sec. 1 and show how existing definitions succeed on one while failing on the other.

Example 3.1 (Preemption [Halpern, 2016a, Ex. 2.3.3])**.** Recall Ex. 1.1. Suzy throws ($ST = 1$), Billy throws ($BT = 1$), Suzy's rock hits ($SH = 1$), Billy's does not ($BH = 0$), and the bottle shatters ($BS = 1$).

$$ST = U_{ST}, \qquad BT = U_{BT}, \tag{1}$$
$$SH = ST, \qquad BH = BT \wedge \neg SH, \tag{2}$$
$$BS = SH \vee BH. \tag{3}$$

The causal diagram is in Fig. 1a.

Example 3.2 (Switch (cf. Hall, 2007, p. 118))**.** Recall Ex. 1.2. A switch ($S = 1$) selects between two circuits ($C_1 = 1, C_2 = 0$); a light ($Y = 1$) is on if either is active.

$$S = U_S, \quad C_1 = S, \quad C_2 = \neg S, \quad Y = C_1 \vee C_2. \tag{4}$$

The causal diagram is in Fig. 1b.

Suzy's throw *is* a cause of the shattering; the switch is *not* a cause of the light. We apply each existing definition to both examples.

$\mathbf{HP_m}$ (Def. 2.1). In preemption, hold BH at its actual value 0 and set $ST = 0$: then $SH = 0$, $BS = 0 \vee 0 = 0 \neq 1$, so AC2 holds and $ST = 1$ is an HP_m-cause. In the switch, hold C_2 at its actual value 0 and set $S = 0$: then $C_1 = 0$, $Y = 0 \vee 0 = 0 \neq 1$, so AC2 holds and $S = 1$ is also an HP_m-cause. HP_m lacks a production test and cannot detect that both switch positions produce the outcome.

BV (Def. 2.6). In the switch, $S = 0$ NESS-causes $Y = 1$ via C_2 in $\mathcal{M}_{S=0}$, so condition (ii) fails and S is not a BV-cause. In preemption, $ST = 0$ NESS-causes $BS = 1$ via Billy's backup chain ($BH = 1 \to BS = 1$) in $\mathcal{M}_{ST=0}$, so condition (ii) fails and ST is not a BV-cause either.

CNESS (Def. 2.7). In preemption, $ST = 1$ NESS-causes $BS = 1$ along $ST \to SH \to BS$, and $ST = 0$ does not replicate this path in $\mathcal{M}_{ST=0}$ (Billy's chain uses different variables), so ST is a CNESS-cause. In the switch, $S = 1$ NESS-causes $Y = 1$ along $S \to C_1 \to Y$, and $S = 0$ does not replicate this in $\mathcal{M}_{S=0}$ (since $C_1 = 0$), so S is also a CNESS-cause.

No existing definition handles both (Table 1).

3.2 FORMAL DEFINITION

Key idea. A cause produces the outcome in a way the alternative value cannot replicate. Replicating means satisfying both production and dependence, not production alone. In symmetric overdetermination ($Y = X_1 \vee X_2$, both 1), $X_1 = 1$ produces Y directly, but $X_1 = 0$ does not: though Y remains 1 via X_2, the alternative value contributes nothing to the outcome. In the switch, $S = 1$ produces Y through C_1 and $S = 0$ produces Y through C_2; both values produce the outcome, so neither is distinguishable from its counterfactual. The definition below formalizes "produces" as a chain of intermediates each verified as causes, and "fails to replicate" as asymmetry in the interventional model.

Table 1: Comparison on canonical examples (✓ = correct verdict, ✗ = incorrect). All definitions agree on cases not shown (e.g., firing squad, XOR).

	Switch	Preemption	Overdet.
HP (2001, 2005, 2015)†	✗	✓	✓
BV (2018)	✓	✗	✓
CNESS (2021)	✗	✓	✓
Nondeterministic (2025)	✗	✓	✓
Ours	✓	✓	✓

† HP_m recovers overdetermination only via multivariate causes.

Definition 3.3 (Productive Actual Causation)**.** Let $(\mathcal{M}, \mathbf{u})$ be a causal world with $X(\mathbf{u}) = x$ and $Y(\mathbf{u}) = y$. A single variable $X = x$ **causes** $Y = y$ if and only if there exist an alternative value $x' \in \mathcal{D}_X \setminus \{x\}$, a witness set $\mathbf{W} \subseteq \{V \in \mathbf{V} \setminus \{X, Y\} : V_{x'}(\mathbf{u}) = V(\mathbf{u})\}$, and an assignment $\mathbf{w}'$ to $\mathbf{W}$ satisfying:

C1 *Counterfactual dependence.* $Y_{x',\mathbf{w}'}(\mathbf{u}) \neq y$, and $\{X\} \cup \mathbf{W}$ is minimal: no strict subset $\tilde{\mathbf{W}} \subset \{X\} \cup \mathbf{W}$ admits an assignment $\tilde{\mathbf{a}}$ (with $\tilde{\mathbf{a}}(X) = x'$ when $X \in \tilde{\mathbf{W}}$) with $Y_{\tilde{\mathbf{a}}}(\mathbf{u}) \neq y$.

C2 *Production.* There exists a directed path $\pi = [X, P_2, \ldots, P_m, Y]$ in $G(\mathcal{M})$ whose intermediates lie outside $\mathbf{W}$ such that every intermediate P_i, for $2 \leq i \leq m$:

(a) changes value under the intervention, $(P_i)_{x',\mathbf{w}'}(\mathbf{u}) \neq P_i(\mathbf{u})$, and
(b) is itself a cause of $Y = y$ under this definition, with induced counterfactual $(P_i)_{x',\mathbf{w}'}(\mathbf{u})$ as its alternative value, and witness $\mathbf{W}, \mathbf{w}'$.

C3 *Asymmetry.* In the submodel $\mathcal{M}_{X=x'}$, $X = x'$ does *not* jointly satisfy C1–C2 for $Y = y$.

When π consists of the single edge $X \to Y$, the quantifier over intermediates is empty and C2 holds vacuously. The definition addresses single-variable causes; the extension to joint causes $\mathbf{X} = \mathbf{x}$ via standard minimality over sets is left for future work. The witness restriction prevents manufacturing spurious dependence; see Ex. 3.5.

HP's AC1 (factuality) is implicit: the definition takes as input a world where $X(\mathbf{u}) = x$ and $Y(\mathbf{u}) = y$. The witness assignment $\mathbf{w}'$ can take arbitrary values, not only actual values; the restriction is on *which* variables may serve as witnesses ($V_{x'} = V(\mathbf{u})$), not on the values they take. The recursion is well-founded; termination is proved in Sec. 4.

3.3 CANONICAL EXAMPLES

Switches. In Ex. 3.2, consider whether $S = 1$ causes $Y = 1$. Both C_1 and C_2 shift under $do(S = s')$ ($(C_1)_{s'} = s' \neq s$, $(C_2)_{s'} = \neg s' \neq \neg s$), so $\mathbf{W} = \emptyset$. With $s' = 0$ and empty witness: $C_1 = 0, C_2 = 1, Y = 0 \vee 1 = 1 = y$. The outcome does not change, so C1 fails and $S = 1$ is not a cause. More generally, in any two-pathway binary switch $Y = f(P_1, P_2)$ with $P_1 = S$, $P_2 = \neg S$, and $f(1, 0) = f(0, 1) = 1$: both P_1 and P_2 shift under $do(S = s')$, so $\mathbf{W} = \emptyset$ and $Y_{s'} = f(\neg s, s) = 1 = y$; C1 fails. The result extends to non-binary domains: in the ternary switch (Ex. B.1), each $x' \in \{1, 2\}$ activates a circuit that forces $Y = 1$, so no witness can flip the outcome and C1 fails.

Symmetric overdetermination. A second canonical test is whether each of two independent, individually sufficient causes qualifies as an actual cause. For single-variable causes, HP_m cannot find a witness: actual-value witnesses fail to neutralize the backup. Our definition resolves this with non-actual witnesses; HP_m instead recovers the verdict through multivariate causes (App. C).

Example 3.4 (Symmetric Overdetermination [Halpern, 2016a, Ex. 2.3.1])**.** Two independent fires ($X_1 = 1$, $X_2 = 1$) each suffice to burn a building, and the building burns ($Y = 1$). The structural equations are:

$$X_1 = U_{X_1}, \qquad X_2 = U_{X_2}, \qquad Y = X_1 \vee X_2. \tag{5}$$

The causal diagram is in Fig. 1c.

In $Y = X_1 \vee X_2$ with $X_1 = X_2 = 1$, consider whether $X_1 = 1$ causes $Y = 1$. Take witness $\mathbf{w}' = \{X_2 = 0\}$ with $x_1' = 0$: then $Y_{x_1', \mathbf{w}'} = 0 \neq 1$, and the set is minimal. C2 holds via the direct path $X_1 \to Y$ (no intermediates). For C3, we check whether $X_1 = 0$ satisfies C1–C2 for $Y = 1$ in $\mathcal{M}_{X_1=0}$. In $\mathcal{M}_{X_1=0}$: $Y = 0 \vee 1 = 1$. Setting $X_1 = 1$ forces $Y = 1$ regardless of any witness, so C1 fails for $X_1 = 0$ and C3 holds: $X_1 = 1$ *is* a cause. Note that the witness $\{X_2 = 0\}$ takes a non-actual value; HP_m restricts witnesses to actual values, so no single-variable witness reveals the dependence. HP_m recovers the verdict through multivariate causes: $\{X_1, X_2\} = \{1, 1\}$ jointly causes $Y = 1$ with empty witness, since neither singleton suffices (App. C).

Preemption. Our definition distinguishes preemption from switches based on whether the model includes intermediate variables.

The coarse preemption model $Y = X \vee (B \wedge \neg X)$ with $X = 1$ and $B = 1$ reduces to $Y = X \vee \neg X$, a two-pathway switch with $P_1 = X$ and $P_2 = \neg X$. The reduced equations are isomorphic to the switch: X is not a cause. HP classifying X as a cause in this model amounts to failing to distinguish preemption from switching; without intermediates, the preemptor and backup are structurally symmetric. The fine-grained model below recovers the expected verdict. In the Suzy–Billy model of Ex. 3.1, $ST = 1$ is a cause of $BS = 1$ and $BT = 1$ is not.

Suzy is a cause. To test whether $ST = 1$ causes $BS = 1$, note that only BT is unaffected by $do(ST = 0)$: all other endogenous variables shift. Take witness $\mathbf{w}' = \{BT = 0\}$ with $st' = 0$. Then $SH = 0$, $BH = 0 \wedge 1 = 0$, and $BS = 0$, flipping the outcome. The set is minimal: $\{ST\}$ alone does not flip BS ($BH = 1$, $BS = 1$), nor does $\{BT\}$ alone ($SH = 1$, $BS = 1$). For C2, consider the path $ST \to SH \to BS$. The intermediate SH changes from 1 to 0 under the intervention. We verify recursively that $SH = 1$ causes $BS = 1$. The witness $\{BT = 0\}$ flips BS to 0, since BT is unaffected by $do(SH = 0)$, and the set is minimal, so C1 holds. C2 is vacuous because $SH \to BS$ is a direct edge. For C3 of this inner check, in $\mathcal{M}_{SH=0}$ setting $SH = 1$ gives $BH = 0$ and $BS = 1$; no witness can flip BS away from 1, so C1 fails for $SH = 0$ and C3 holds. Thus $SH = 1$ is a cause of $BS = 1$, completing C2 of the outer check.

For C3 of the outer check, in $\mathcal{M}_{ST=0}$: $SH = 0$, $BH = 1$, $BS = 1$. Setting $st' = 1$ with any $\mathbf{w}' \subseteq \{BT\}$ gives $BS = 1$, so C1 fails for $ST = 0$ and C3 holds. Therefore, $ST = 1$ *is* a cause of $BS = 1$.

Billy is not a cause. To test whether $BT = 1$ causes $BS = 1$, take witness $\mathbf{w}' = \{SH = 0\}$ with $bt' = 0$: then $BH = 0 \wedge 1 = 0$ and $BS = 0$, flipping the outcome; the set is minimal. For C2, the only path is $BT \to BH \to BS$. Under the witness intervention, $BH = 0 \wedge \neg 0 = 0$, which equals the actual value $BH = 0$: the intermediate does not change, so C2 fails. Therefore, $BT = 1$ is *not* a cause of $BS = 1$.

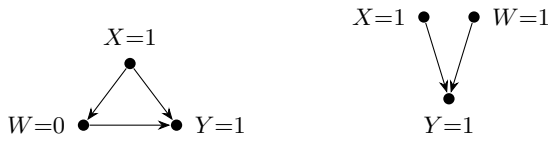
Witness eligibility. The definition restricts witnesses to variables unaffected by the intervention ($V_{x'} = V(\mathbf{u})$). This restriction distinguishes switches from overdetermination: in both, a backup pathway sustains the outcome, but in a switch the backup is mechanically linked to the cause, while in overdetermination it is independent. The following example illustrates why unrestricted witnesses would conflate the two.

Example 3.5 (Reduced Switch)**.** A switch ($X = 1$) controls a relay (W) and a light (Y) directly. The relay opposes the switch; the light is on if either path is active: $W = \neg X$, $Y = X \vee W$. In the actual world, $X = 1$, $W = 0$, and $Y = 1$.

This is a tautological switch: $Y = X \vee \neg X = 1$ for all X.

3.4 COMPARISON WITH PRIOR WORK

Four features distinguish this definition from HP_m: (i) C2 requires every intermediate to be a cause, not just a changed variable; (ii) witnesses take arbitrary values, not only actual



(a) Reduced switch (b) Overdetermination

Figure 2: Witness eligibility contrast. Both models share $Y = X \vee W$ with X=1, Y=1. *(a)* $W = \neg X$: W shifts under $do(X{=}0)$, so W is ineligible as witness; C1 fails and X is not a cause. *(b)* $W = U_W$: W does not shift, so $\{W{=}0\}$ is eligible; X is a cause.

values; (iii) C3 checks asymmetry against both production and dependence; and (iv) witnesses are restricted to variables unaffected by the intervention ($V_{x'} = V(\mathbf{u})$).

Joint minimality versus AC3. HP's AC3 minimizes only the cause set $\mathbf{X}$: no proper subset of $\mathbf{X}$ satisfies AC2. Our C1 minimizes the joint set $\{X\} \cup \mathbf{W}$: no proper subset admits any assignment that flips the outcome. This stronger requirement prevents a candidate cause from *hitchhiking* on witness variables that already do the causal work. In the gun-loader example (Ex. C.6), HP's original definition accepts $A = 1$ with witness $\{B = 1, C = 0\}$, but the set is not minimal under C1 since $C = 0$ alone flips D. C1 rejects A by enforcing that neither the cause alone nor the witness alone suffices to flip the outcome.

Comparison with BV. Like BV, C3 evaluates in the submodel $\mathcal{M}_{X=x'}$. The key difference is *what* C3 checks: BV requires only that x' not *produce* Y (NESS-causation), whereas C3 requires that x' not satisfy both production (C2) *and* counterfactual dependence (C1). This resolves the tradeoff: in preemption, the backup produces Y but no witness can make Y counterfactually depend on x', so C1 fails for x' and C3 holds. C1's minimality requirement provides a further separation: in the shadow variable (Fig. 1d), BV accepts $X_2 = 0$ as a cause because its asymmetry check passes, but our C1 rejects it since $\{X_2, X_1\}$ is not minimal ($X_1 = 0$ alone flips Y; see App. D).

CNESS regression on switches. As shown in the CNESS walkthrough above, CNESS's path-restricted asymmetry fails to detect that $S = 0$ also produces Y via C_2, an issue not discussed by Beckers [2021].

Proposition 3.6 (Nondeterministic switch misclassification)**.** *The nondeterministic generalization also misclassifies the switch. Structural simplification removes the edge $C_2 \to Y$, since C_2 is not part of a sufficient set for Y along the path through C_1. Under the simplified model, $S = 1$ is counterfactually necessary for $Y = 1$, so S is classified as a cause.*

NESS and counterfactual dependence. For binary variables, direct NESS (Def. 2.3) is equivalent to counterfactual dependence given an actual-value background: $X{=}x$ is necessary within a sufficient set for $f_Y = y$ if and only if some actual-value parent context makes Y counterfactually dependent on X. For $|\mathcal{D}_X| \geq 3$, sufficiency requires $f_Y = y$ for *all* assignments of unspecified parents, strictly stronger than a point counterfactual, and CNESS's path-restricted asymmetry no longer detects multi-valued switches (Ex. B.1).

3.5 THE EXPLANANDUM AND GRANULARITY

The coarse preemption model reduces to $Y = X \vee \neg X$, the same equation as the switch with intermediates removed. Any definition that rejects switches must also reject X here, so distinguishing preemption from switching requires that the model encode what makes them different [Halpern, 2016a, Chapter 2]. The SCM framework offers two routes: intermediate variables that trace the production pathway, or a multi-valued outcome domain that encodes manner directly.

The Suzy–Billy model (Ex. 3.1) takes the first route. The intermediate $BH = BT \wedge \neg SH$ encodes that Billy's rock does not arrive when Suzy's hits first. C2 verifies the intermediate SH as a cause of BS along the primary path, while the backup path through BH fails C2 because BH does not change.[1] The choice of intermediates reflects an asymmetry in how narratives translate to SCMs: a story that emphasizes Suzy's rock arriving first naturally produces $BH = BT \wedge \neg SH$, encoding temporal priority; a story that treats the shattering as inevitable regardless of source produces the coarse $Y = X \vee \neg X$.

Example 3.7 (File Server)**.** Two clients, Alice ($X_1 = 1$) and Bob ($X_2 = 1$), each send a request to a server with outcome $Y \in \{., A, B, AB\}$. The merge policy determines $f(1,1)$: *lock* retains the first writer ($f(1,1) = A$), *overwrite* retains the last ($f(1,1) = B$), and *merge* retains both ($f(1,1) = AB$); the remaining cases are shared ($f(1,0) = A$, $f(0,1) = B$, $f(0,0) = .$). The structural equations are:

$$X_1 = U_{X_1}, \quad X_2 = U_{X_2}, \quad Y = f(X_1, X_2). \tag{6}$$

Each policy parallels a canonical causal structure. Under lock ($Y = A$), only Alice is a cause ($Y_{x'_1=0} = B \neq A$, $Y_{x'_2=0} = A = y$): this is preemption, with Alice's file playing the role of Suzy's rock. Under overwrite ($Y = B$), the roles reverse: Bob's later write supersedes Alice's, so only Bob is a cause. Physical preemption is typically irreversible, so the overwrite pattern is most natural in digital systems; Suzy would not be a cause of the shattering if Billy's rock could reset the bottle. Under merge ($Y = AB$), both are but-for necessary ($Y_{x'_1=0} = B \neq AB$, $Y_{x'_2=0} = A \neq AB$).

[1] Early preemption, where the backup withdraws upon observing the primary cause ($BT = BT_{\text{exo}} \wedge \neg ST$), is structurally a switch for the same reason (Sec. C.8).

Coarsening to $E = \mathbb{1}[Y \neq .]$ maps all three policies to $E = X_1 \vee X_2$, symmetric overdetermination: both are causes. For each cause in the fine-grained cases, but-for dependence suffices with empty witness; the coarsened case requires a non-actual witness as in Ex. 3.4 (see Sec. D.2). The Suzy–Billy model resolves preemption through intermediates that encode hit order; the file server encodes it directly in $\mathcal{D}_Y$, illustrating the two routes and motivating the fragility question below.

One might object that fine-grained individuation leads to *fragility*: if the outcome event encodes every detail of manner and timing, any factor affecting those details becomes a cause [Halpern, 2016a, Chapter 2]. This objection conflates the granularity of the model with the granularity of the query. A natural-language "why" question carries an implicit contrastive foil [Schaffer, 2005, Miller, 2019]: "why did the bottle shatter?" contrasts shattering with not shattering, collapsing manner variants into a single outcome value. In the SCM framework, this contrast is encoded in $\mathcal{D}_Y$, and the definition gives well-defined verdicts relative to the chosen granularity. Fragility arises only from individuating beyond what the question demands; the SCM framework makes this modeling choice explicit and scrutinizable, whereas event-based frameworks leave it implicit (App. F).

4 ALGORITHM AND COMPLEXITY

The recursive structure of Def. 3.3 admits a natural memoized algorithm. Both procedures take a target outcome y as a global parameter, fixed once by the top-level query. CAUSES (Alg. 1) checks the full definition: it enumerates alternative values x', calls PRODUCES to find a witness satisfying C1–C2, then verifies C3 by checking that no alternative value x'' satisfies C1–C2 for x' in the submodel $\mathcal{M}_{X=x'}$. PRODUCES (Alg. 2) checks whether a specific alternative value x' satisfies C1–C2: it searches over directed paths π from X to Y and witness sets $(\mathbf{W}, \mathbf{w}')$, returning a triple $(x', \mathbf{w}', \pi)$ on success or $\perp$ on failure.

Walkthrough. CAUSES iterates over alternative values x' (line 2). For each x', it calls PRODUCES to search for a path π and witness $(\mathbf{W}, \mathbf{w}')$ jointly satisfying C1–C2 (line 3). If PRODUCES succeeds, the C3 check (lines 6–8) tests whether x' can itself satisfy C1–C2 in the submodel $\mathcal{M}_{X=x'}$ for any alternative x''; if no x'' succeeds, C3 holds and the algorithm returns the witness triple $(x', \mathbf{w}', \pi)$ (line 9). The call PRODUCES$(\mathcal{M}_{X=x'}, \mathbf{u}, X, Y, x'')$ re-intervenes on X inside the submodel $\mathcal{M}_{X=x'}$, overriding the fixed value x' with x''; for $|\mathcal{D}_X| = 2$ this is the unique remaining value, and for $|\mathcal{D}_X| \geq 3$ the loop enumerates all alternatives to x' in turn.

PRODUCES iterates over directed paths from X to Y (line 1). For each path, it first verifies C2(b): every intermediate P_i must recursively be a cause of $Y = y$ (line 2). It then constructs the eligible witness set $\mathcal{E}$ (line 4): variables outside π that are unaffected by the intervention ($V_{x'} = V(\mathbf{u})$), combining the definition's witness restriction with C2's requirement that path intermediates lie outside $\mathbf{W}$. The search is restricted to the ancestor frontier $F^*(\pi)$ (line 5; see Prop. 4.1). For each witness assignment, it checks C1 (outcome flip, line 6; minimality, line 7) and C2(a) (all intermediates shift, line 9). The minimality check (line 7) tests every strict subset $\tilde{\mathbf{W}}$ of the intervention set $\{X\} \cup \mathbf{W}$: the assignment $\tilde{\mathbf{a}}$ maps each variable in $\tilde{\mathbf{W}}$ to a value, and $Y_{\tilde{\mathbf{a}}}$ intervenes only on $\tilde{\mathbf{W}}$. When $X \in \tilde{\mathbf{W}}$, the assignment sets $X = x'$; when $X \notin \tilde{\mathbf{W}}$, X remains at its factual value.

Algorithm 1 CAUSES$(\mathcal{M}, \mathbf{u}, X, Y)$: does $X{=}x$ cause $Y{=}y$?

```
1: x ← X(u)
2: for all x' ∈ D_X \ {x} do
3:     r ← PRODUCES(M, u, X, Y, x')
4:     if r = ⊥ then continue                ▷ C1–C2 fail for x'
5:     end if
6:     asym ← true
7:     for all x'' ∈ D_X \ {x'} do           ▷ C3 check in M_{X=x'}
8:         if PRODUCES(M_{X=x'}, u, X, Y, x'') ≠ ⊥ then
9:             asym ← false; break
10:        end if
11:    end for
12:    if asym then return r                 ▷ C1–C3 all hold
13:    end if
14: end for
15: return ⊥
```

Both procedures are memoized on their full argument tuple $(\mathcal{M}, \mathbf{u}, X, Y)$; since $\mathbf{u}$ and Y are fixed throughout each top-level query, the effective key is $(\mathcal{M}, X)$. Termination follows from the DAG structure: C2(b) recurses on intermediates strictly closer to Y, and each C3 call permanently fixes one additional variable, so neither recursion can repeat (App. E).

Witness frontier. Without further optimization, the witness search in PRODUCES is exponential: each of $n - 2$ non-target variables may be absent or take any of d values, giving $(d{+}1)^{n-2}$ configurations per path. The causal graph admits a tighter bound. For a directed path π from X to Y, the values along π depend only on the non-path parents of path nodes.

Proposition 4.1 (Witness frontier)**.** *Let $\pi = [X, P_2, \ldots, P_m, Y]$ be a directed path in $G(\mathcal{M})$. Define the* ancestor frontier *of π as the non-path ancestors of path nodes: $F^*(\pi) = \mathrm{Anc}(\pi) \setminus \pi$. If $(x', \mathbf{w}', \pi)$ jointly satisfy C1–C2, then there exists an eligible witness $\mathbf{W}^* \subseteq F^*(\pi)$ with values $\mathbf{w}^*$ such that $(x', \mathbf{w}^*, \pi)$ also satisfies C1–C2. The witness search can be restricted to eligible subsets of $F^*(\pi)$.*

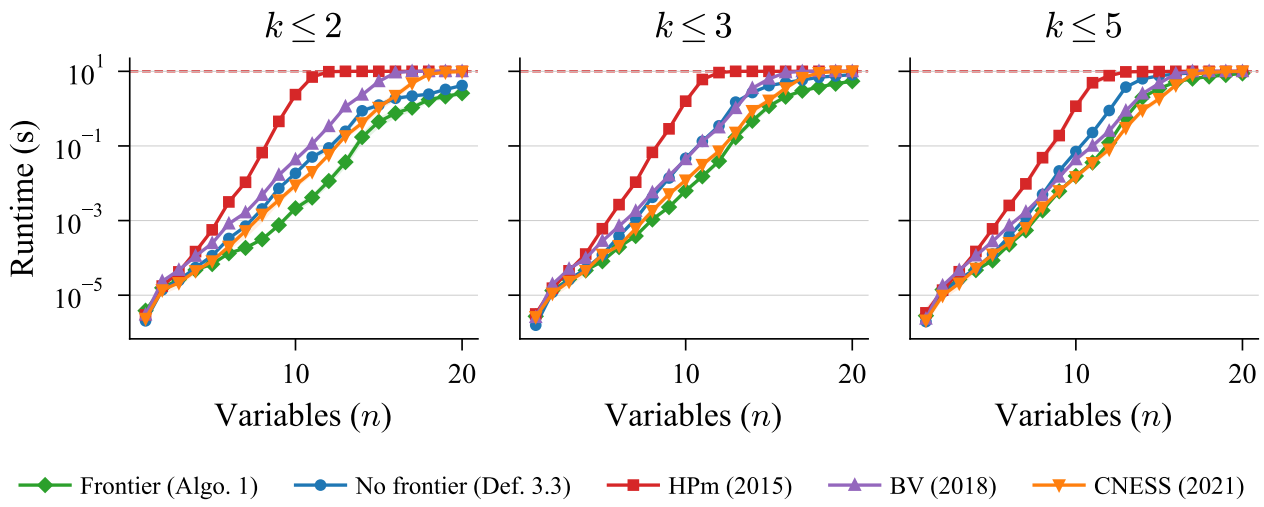


Figure 3: Runtime vs. number of variables n for random binary SCMs (density 0.4, 50 seeds per configuration, 10 s timeout); the dashed red line denotes the timeout. Lines show medians; shaded regions show the interquartile range (IQR) over 50 seeds. *Left:* $k \leq 3$. *Center:* $k \leq 5$. *Right:* Unbounded in-degree. "Def. 3.3 (no frontier)" evaluates Def. 3.3 by enumerating all eligible witnesses without the frontier restriction of Prop. 4.1. The frontier-restricted checker remains below the timeout at $n = 20$ when $k \leq 3$, while HP_m, BV, and CNESS all exceed it beyond $n \approx 12$. As k grows, the frontier advantage narrows.

Algorithm 2 PRODUCES$(\mathcal{M}, \mathbf{u}, X, Y, x')$: do C1–C2 hold for counterfactual x'?

for all paths $\pi = [X, P_2, \ldots, P_m, Y]$ in $G(\mathcal{M})$ **do**
  if $\exists\, P_i$ with $\neg$ CAUSES$(\mathcal{M}, \mathbf{u}, P_i, Y)$ **then**
    continue ▷ C2(b) fails
  end if
  $\mathcal{E} \leftarrow \{V \in \mathbf{V} \setminus \pi : V_{x'} = V(\mathbf{u})\}$
  for all $\mathbf{W} \subseteq \mathcal{E} \cap F^*(\pi)$, $\mathbf{w}' \in \prod_{W \in \mathbf{W}} \mathcal{D}_W$ **do**
    if $Y_{x',\mathbf{w}'} = y$ **then continue** ▷ no flip
    end if
    if $\exists\, \tilde{\mathbf{W}} \subset \{X\} \cup \mathbf{W}$, assignment $\tilde{\mathbf{a}}$ to $\tilde{\mathbf{W}}$ with $\tilde{\mathbf{a}}(X) = x'$ when $X \in \tilde{\mathbf{W}}$, and $Y_{\tilde{\mathbf{a}}} \neq y$ **then**
      continue ▷ not minimal
    end if
    if $\forall\, i\colon\ (P_i)_{x',\mathbf{w}'} \neq (P_i)_{\mathbf{u}}$ **then**
      return $(x', \mathbf{w}', \pi)$ ▷ C1–C2 hold
    end if
  end for
end for
return $\perp$

Only the ancestors of path nodes can affect path-node values: setting any non-ancestor witness to any value leaves the path unchanged. With maximum in-degree k and longest path L, $|F^*(\pi)| \leq O(k^L)$, so the witness enumeration drops from $(d{+}1)^{n-2}$ to $(d{+}1)^{O(k^L)}$, independent of n.

Tractability. The algorithm is fixed-parameter tractable in (k, L): for fixed constants k and L, it runs in time polynomial in n and d. The frontier limits each witness search to $d^{O(kL)}$ configurations; each variable has at most $O(k^L)$ paths to Y; and memoization ensures each $(\mathcal{M}, X)$ pair is solved once, with at most n variables per model. C3 generates submodels by fixing variables to constants, but with bounded k and L at most $O(k^L)$ ancestors participate, bounding the number of distinct submodels.

Theorem 4.2 (Fixed-parameter tractability)**.** *Let k denote the maximum in-degree, L the longest directed path length in $G(\mathcal{M})$, and d the maximum domain size. Alg. 1 is fixed-parameter tractable in (k, L): for any fixed constants k and L, it checks whether $X = x$ causes $Y = y$ in time polynomial in both n and d. Specifically, the runtime is $O\big(n^2 \cdot d^{O(k^L)}\big)$.*

The bound follows from three ingredients: C3 generates at most $(d{+}1)^{O(k^L)}$ submodels (one per variable fixed to a constant), each variable has $O(k^L)$ directed paths to Y, and each path has $(d{+}1)^{O(k^L)}$ witness configurations (App. E).

For general k and L, the problem is in PSPACE (Lem. E.2); whether it is in Σ_2^P, matching the complexity of HP [Aleksandrowicz et al., 2017, Eiter and Lukasiewicz, 2002], remains open.

Empirical validation. Fig. 3 compares the frontier-restricted checker, Def. 3.3 without the frontier restriction, HP_m [Halpern, 2015], BV [Beckers and Vennekens, 2018], and CNESS [Beckers, 2021] on random binary SCMs with n up to 20. When $k \leq 3$, the frontier checker has median

runtime under one second at $n = 20$, an order-of-magnitude speedup over all baselines, which exceed the 10 s timeout beyond $n \approx 12$. HP_{m} checks whether X belongs to some minimal multivariate cause set, requiring enumeration over all subsets of variables [Eiter and Lukasiewicz, 2002]. As k grows, the frontier approaches the full variable set and the speedup diminishes. Longest path lengths in the benchmark graphs (density 0.4, $n = 20$, 50 seeds) range from 5 to 10 for $k \leq 3$; $Thm.\ 4.2$ guarantees polynomial runtime when k and L are fixed constants, and the benchmark results show that the frontier checker remains fast well beyond this regime.

5 CONCLUSION

We introduced productive actual causation, the first definition in the SCM framework to correctly classify both switch variants and preemption. The key mechanism is compositional: every intermediate in a causal chain must itself be a cause, and the alternative value must fail to replicate this chain. Distinguishing preemption from switching is a modeling question, not a definitional one: intermediates trace the production pathway, and outcome granularity encodes manner. We provide a memoized algorithm that is fixed-parameter tractable in the maximum in-degree k and longest path length L, the first parameterized tractability result for any structural definition of actual causation.

References

Gadi Aleksandrowicz, Hana Chockler, Joseph Y Halpern, and Alexander Ivrii. The computational complexity of structure-based causality. *Journal of Artificial Intelligence Research*, 58:431–451, 2017. doi: 10.1613/jair.5505.

Holger Andreas and Mario Günther. A regularity theory of causation. *Pacific Philosophical Quarterly*, 105(1):2–32, 2024. doi: 10.1111/papq.12447.

Elias Bareinboim, Juan D. Correa, Duligur Ibeling, and Thomas Icard. On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 507–556. Association for Computing Machinery, New York, NY, USA, 1st edition, 2022. doi: 10.1145/3501714.3501743.

Sander Beckers. The counterfactual ness definition of causation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(7):6210–6217, May 2021. ISSN 2159-5399. doi: 10.1609/aaai.v35i7.16772. URL `http://dx.doi.org/10.1609/aaai.v35i7.16772`.

Sander Beckers. Actual causation and nondeterministic causal models. In *Proceedings of the Fourth Conference on Causal Learning and Reasoning (CLeaR 2025)*, volume 275 of *PMLR*, pages 514–532, 2025.

Sander Beckers and Joost Vennekens. A principled approach to defining actual causation. *Synthese*, 195(2):835–862, 2018. doi: 10.1007/s11229-016-1247-1.

Marc Denecker, Bart Bogaerts, and Joost Vennekens. Explaining actual causation in terms of possible causal processes. In *Logics in Artificial Intelligence (JELIA 2019)*, pages 214–230. Springer, 2019. doi: 10.1007/978-3-030-19570-0_14.

Phil Dowe. *Physical Causation*. Cambridge University Press, Cambridge, 2000. doi: 10.1017/CBO9780511570650.

Thomas Eiter and Thomas Lukasiewicz. Complexity results for structure-based causality. *Artificial Intelligence*, 142 (1):53–89, 2002. doi: 10.1016/S0004-3702(02)00271-0.

Ned Hall. Two concepts of causation. In *Causation and Counterfactuals*, pages 225–276. MIT Press, 2004. doi: 10.7551/mitpress/1752.003.0010.

Ned Hall. Structural equations and causation. *Philosophical Studies*, 132(1):109–136, 2007. doi: 10.1007/s11098-006-9052-0.

Joseph Y Halpern. A modification of the Halpern-Pearl definition of causality. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, pages 3022–3033, 2015.

Joseph Y Halpern. *Actual Causality*. MIT Press, 2016a. doi: 10.7551/mitpress/10809.001.0001.

Joseph Y Halpern. Appropriate causal models and the stability of causation. *The Review of Symbolic Logic*, 9(1): 76–102, 2016b. doi: 10.1017/S1755020315000246.

Joseph Y Halpern and Christopher Hitchcock. Graded causation and defaults. *The British Journal for the Philosophy of Science*, 66(2):413–457, 2015. doi: 10.1093/bjps/axt050.

Joseph Y Halpern and Judea Pearl. Causes and explanations: A structural-model approach. Part I: Causes. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI 2001)*, pages 194–202, 2001.

Joseph Y Halpern and Judea Pearl. Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science*, 56(4):843–887, 2005. doi: 10.1093/bjps/axi147.

H. L. A. Hart and Tony Honoré. *Causation in the Law*. Oxford University Press, 2nd edition, 1985. doi: 10.1093/acprof:oso/9780198254744.001.0001.

Christopher Hitchcock. The intransitivity of causation revealed in equations and graphs. *The Journal of Philosophy*, 98(6):273–299, 2001. doi: 10.2307/2678432.

Christopher Hitchcock and Joshua Knobe. Cause and norm. *Journal of Philosophy*, 106(11):587–612, 2009. doi: 10.5840/jphil20091061128.

David Lewis. Causation. *The Journal of Philosophy*, 70 (17):556–567, 1973. doi: 10.2307/2025310.

Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019. doi: 10.1016/j.artint.2018.07.007.

L. A. Paul and Ned Hall. *Causation: A User's Guide*. Oxford University Press, 2013. doi: 10.1093/acprof:oso/9780199673445.001.0001.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000a. ISBN 0521773628.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000b.

Judea Pearl. *Causality*. Cambridge University Press, 2009. doi: 10.1017/CBO9780511803161.

Drago Plecko and Elias Bareinboim. Causal fairness analysis: A causal toolkit for fair machine learning. *Foundations and Trends in Machine Learning*, 17(3):304–589, 2024. doi: 10.1561/2200000106.

Wesley C. Salmon. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton, NJ, 1984. doi: 10.1515/9780691221489.

Jonathan Schaffer. Contrastive causation. *The Philosophical Review*, 114(3):327–358, 2005. doi: 10.1215/00318108-114-3-327.

Brad Weslake. A partial theory of actual causation. Unpublished manuscript, 2015.

James Woodward. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, 2003. doi: 10.1093/0195155270.001.0001.

Richard W Wright. Causation in tort law. *California Law Review*, 73(6):1735–1828, 1985. doi: 10.2307/3480373.

Tomasz Wysocki. Conjoined cases. *Synthese*, 201(6), 2023. doi: 10.1007/s11229-023-04101-w.

Stephen Yablo. Mental causation. *The Philosophical Review*, 101(2):245–280, 1992. doi: 10.2307/2185535.

Productive Actual Causation (Supplementary Material)

Kai-Zhan Lee[1] **Elias Bareinboim**[1]

[1]Causal AI Lab, Columbia University, New York, NY, USA

Contents

A PRIOR DEFINITIONS

All prior definitions (Defs. 2.1 to 2.7) appear in Sec. 2. CNESS weakens BV's asymmetry by comparing NESS-causation along specific paths rather than globally. This allows CNESS to handle preemption, since the backup path uses different variables than the actual path, but it also weakens the condition enough to misclassify switches (Sec. 3.4).

B ALTERNATIVE FORMULATIONS OF C3

C3 checks that the alternative value x' does not satisfy C1–C2 in $\mathcal{M}_{X=x'}$. The C1 sub-check within this evaluation may use any alternative value $x'' \neq x'$. Two natural alternatives restrict or strengthen this check; we explain why both are inadequate.

Three formulations. Let $X = x$ be the candidate cause, $Y = y$ the effect, and $x' \neq x$ the alternative value selected by C1.

(A) Unrestricted (adopted in Def. 3.3). $X = x'$ does not satisfy C1–C2 for $Y = y$ in $(\mathcal{M}_{X=x'}, \mathbf{u})$. The C1 sub-check for x' may use any $x'' \neq x'$.

(B) Restricted to original value. Same as (A), but the C1 sub-check for x' must use $x'' = x$ (the original actual value).

(C) Full recursive. $X = x'$ does not *cause* $Y = y$ in $(\mathcal{M}_{X=x'}, \mathbf{u})$, i.e., does not satisfy C1–C3. C3 invokes itself recursively.

For binary variables ($|\mathcal{D}_X| = 2$), all three coincide: the only counterfactual for x' is x itself, so (A) and (B) are identical, and the fixpoint analysis shows (C) agrees as well. For $|\mathcal{D}_X| \geq 3$, the formulations diverge.

Example B.1 (Ternary switch)**.** A selector ($X \in \{0, 1, 2\}$) routes power through one of three circuits (C_1, C_2, or a direct channel), with exogenous gates $G_1 = G_2 = 1$. Here $\mathbb{1}_{X=k}$ equals 1 when $X = k$ and 0 otherwise.

$$C_1 = \mathbb{1}_{X=1} \cdot G_1, \tag{7}$$
$$C_2 = \mathbb{1}_{X=2} \cdot G_2, \tag{8}$$
$$Y = \mathbb{1}_{X=0} \vee C_1 \vee C_2. \tag{9}$$

With $X = 0$, the actual values are $C_1 = 0$, $C_2 = 0$, and $Y = 1$. This is a three-way switch: $X = 0$ produces Y through the direct channel, $X = 1$ would produce Y through C_1, and $X = 2$ through C_2. No value of X is a cause of Y.

Under the witness restriction (Def. 3.3), C1 fails for every x': with $x' = 1$, $C_1 = 1$ forces $Y = 1$ regardless of C_2; with $x' = 2$, $C_2 = 1$ forces $Y = 1$ regardless of C_1 (Sec. 3). The following analysis relaxes the restriction to $\mathbf{W} \subseteq \mathbf{V} \setminus \{X, Y\}$ to show that formulation (A) independently rejects the switch through C3.

(A) correctly rejects. Consider whether $X = 0$ causes $Y = 1$ with $x' = 1$. For C1, witness $\mathbf{w} = \{C_1 = 0\}$ gives $Y_{x',c_1} = 0 \vee 0 \vee 0 = 0 \neq 1$, and the set is minimal. C2 holds via the direct path $X \to Y$ (no intermediates). For C3, we check whether $X = 1$ satisfies C1–C2 in $\mathcal{M}_{X=1}$. In $\mathcal{M}_{X=1}$: $C_1 = 1$, $C_2 = 0$, $Y = 1$, so factuality holds. C1 holds with $x'' = 2$ and $\mathbf{w}' = \{C_2 = 0\}$: $Y_{x'',c_2} = 0 \vee 0 \vee 0 = 0$. C2 holds via path $X \to C_1 \to Y$ with C_1 changing from 1 to 0 and being a cause of Y. Since $X = 1$ satisfies C1–C2, C3 fails. The same argument applies to $x' = 2$ with $x'' = 1$ in the C3 sub-check. $X = 0$ is correctly rejected.

(B) incorrectly accepts. The C3 sub-check for $x' = 1$ is restricted to $x'' = 0$. In $\mathcal{M}_{X=1}$, C1 with $x'' = 0$ gives $Y_{x'',\mathbf{w}'} = \mathbb{1}_{0=0} \vee (C_1)_{\mathbf{w}'} \vee (C_2)_{\mathbf{w}'} \geq 1$ for any $\mathbf{w}'$. The direct channel $\mathbb{1}_{X=0}$ is unblockable by witness variables, so Y cannot be flipped and C1 fails. Since $X = 1$ does not satisfy C1–C2, C3 holds, and $X = 0$ is incorrectly accepted as a cause.

The issue is that returning X to its original value 0 reactivates the direct channel $\mathbb{1}_{X=0}$, preventing any witness from flipping Y. Formulation (A) avoids this by testing $x'' = 2$, which deactivates the direct channel and reveals that $X = 1$ can produce Y through C_1.

(C) is ill-defined for $|\mathcal{D}_X| \geq 3$. Under formulation (C), checking whether $X = 0$ causes $Y = 1$ triggers a cycle: C3 for $X = 0$ asks whether $X = 1$ causes Y in $\mathcal{M}_{X=1}$, whose C3 asks whether $X = 2$ causes Y in $\mathcal{M}_{X=2}$, whose C3 asks whether $X = 0$ causes Y in $\mathcal{M}_{X=0}$. If all three values are causes, each value's C3 fails, a contradiction. If no value is a cause, each value's C3 holds, making every value a cause, also a contradiction. No consistent truth assignment exists under standard two-valued semantics; a fixpoint semantics could be defined but would require additional theoretical machinery beyond the scope of this work.

Conclusion. Formulation (A) is the correct generalization: it preserves switch-rejection for all domain sizes while maintaining well-founded recursion, since only C2 recurses and it does so on path length. Formulation (B) is too permissive for $|\mathcal{D}_X| \geq 3$, and (C) is ill-defined beyond binary variables. For binary domains, all three coincide.

B.1 STABILITY UNDER CONSERVATIVE EXTENSIONS

Stability of non-causation. A *conservative extension* of an SCM $\mathcal{M}$ is a model $\mathcal{M}'$ that adds new endogenous variables while preserving all counterfactuals between variables in $\mathbf{V}(\mathcal{M})$ [Halpern, 2016b, Definition 4.2]. Under conservative extensions, HP causal verdicts can alternate infinitely between "cause" and "non-cause" [Halpern, 2016b, Theorem 6.1]; normality assumptions are required to prevent this. Our definition prevents this alternation without additional assumptions.

Proposition B.2 (Non-causation stability)**.** *Let $\mathcal{M}'$ be a conservative extension of $\mathcal{M}$, and let X be binary. If $X \to Y$ is an edge in both $G(\mathcal{M})$ and $G(\mathcal{M}')$ and there exists $x' \neq x$ such that $X = x'$ satisfies factuality and C1 in $(\mathcal{M}_{X=x'}, \mathbf{u})$, then $X = x$ is not a cause of $Y = y$ in $(\mathcal{M}', \mathbf{u})$.*

Corollary B.3. *The alternation of Halpern [2016b, Theorem 6.1] does not occur under Def. 3.3: the verdict stabilizes at "non-cause" for all $n \geq 2$.*

C CANONICAL EXAMPLES

We work through canonical examples from the actual causation literature, verifying the verdicts summarized in Table 1. The first three examples compare all four definitions; the remaining examples focus on our definition. For HP, we evaluate the original formulation [Halpern and Pearl, 2005], which allows arbitrary witness values in AC2, not only actual values; the modified formulation [Halpern, 2015] restricts witnesses to actual values and yields different verdicts on symmetric overdetermination and the shadow variable. Main-text derivations for Productive Actual Causation appear in Sec. 3; we include them here for completeness.

C.1 SWITCH

The structural equations are $C_1 = S$, $C_2 = \neg S$, and $Y = C_1 \vee C_2$. With $S = 1$, the actual values are $C_1 = 1$, $C_2 = 0$, and $Y = 1$.

HP (incorrectly accepts). Set $\mathbf{W} = \{C_2\}$ at its actual value 0 and $s' = 0$: then $C_1 = 0$ and $Y_{s',c_2} = 0 \vee 0 = 0 \neq 1$, so AC2 holds. Thus $S = 1$ is an HP-cause of $Y = 1$.

BV (correctly rejects). $S = 1$ NESS-causes $Y = 1$: $S = 1$ directly NESS-causes $C_1 = 1$ (since $\{S{=}1\}$ is sufficient for $C_1 = 1$ but $\varnothing$ is not), and $C_1 = 1$ directly NESS-causes $Y = 1$ (since $\{C_1{=}1\}$ is sufficient for $Y = 1$ but $\varnothing$ is not). Asymmetry check: in $\mathcal{M}_{S=0}$, $C_1 = 0$, $C_2 = 1$, $Y = 1$. $S = 0$ directly NESS-causes $C_2 = 1$, and $C_2 = 1$ directly NESS-causes $Y = 1$, so $S = 0$ NESS-causes $Y = 1$. Condition (ii) of Def. 2.6 fails; $S = 1$ is not a BV-cause.

CNESS (incorrectly accepts). $S = 1$ NESS-causes $Y = 1$ along path $p = (S, C_1, Y)$. In $\mathcal{M}_{S=0}$: for $S = 0$ to NESS-cause $Y = 1$ along any subpath of p, we would need $C_1 = 0$ to directly NESS-cause $Y = 1$. Since $\{C_1{=}0\}$ is not sufficient for $Y = 1$ (as $f_Y(0, C_2) = C_2$, which is not always 1), no such chain exists. Condition (ii) of Def. 2.7 holds, so $S = 1$ is a CNESS-cause.

Ours (correctly rejects). $S = 1$ is not a cause; see derivation in Sec. 3.

C.2 SUZY–BILLY PREEMPTION

The structural equations are $SH = ST$, $BH = BT \wedge \neg SH$, and $BS = SH \vee BH$. In the actual world, $ST = 1$, $BT = 1$, $SH = 1$, $BH = 0$, and $BS = 1$.

HP (correctly accepts). Set $\mathbf{W} = \{BH\}$ at its actual value 0 and $st' = 0$: then $SH = 0$, BH is held at 0, and $BS = 0 \vee 0 = 0 \neq 1$. AC2 holds. Thus $ST = 1$ is an HP-cause.

BV (incorrectly rejects). $ST = 1$ NESS-causes $BS = 1$ via the chain $ST{=}1 \to SH{=}1 \to BS{=}1$. Asymmetry check: in $\mathcal{M}_{ST=0}$, $SH = 0$, $BH = 1 \wedge \neg 0 = 1$, $BS = 0 \vee 1 = 1$. $ST = 0$ directly NESS-causes $SH = 0$; $SH = 0$ directly NESS-causes $BH = 1$ (since $\{SH{=}0, BT{=}1\}$ is sufficient for $BH = 1$, but $\{BT{=}1\}$ alone is not); $BH = 1$ directly NESS-causes $BS = 1$. So $ST = 0$ NESS-causes $BS = 1$; condition (ii) fails and $ST = 1$ is not a BV-cause.

CNESS (correctly accepts). $ST = 1$ NESS-causes $BS = 1$ along path $p = (ST, SH, BS)$. In $\mathcal{M}_{ST=0}$: $SH = 0$. For $ST = 0$ to NESS-cause $BS = 1$ along any subpath of p, we would need $SH = 0$ to directly NESS-cause $BS = 1$. Since $\{SH{=}0\}$ is not sufficient for $BS = 1$ (as $f_{BS}(0, BH) = BH$), the chain breaks. Condition (ii) holds; $ST = 1$ is a CNESS-cause.

Ours (correctly accepts). $ST = 1$ is a cause; see derivation in Sec. 3.

C.3 SYMMETRIC OVERDETERMINATION

The structural equation is $Y = X_1 \vee X_2$, with actual values $X_1 = 1$, $X_2 = 1$, and $Y = 1$.

HP (correctly accepts). Under the original formulation, $\mathbf{W}$ may take non-actual values. Set $\mathbf{W} = \{X_2\}$ to 0 (non-actual) and $x_1' = 0$: then $Y_{x_1',x_2} = 0 \vee 0 = 0 \neq 1$, so AC2(a) holds. For AC2(b), $Z = \varnothing$ and $Y_{X_1=1,X_2=0} = 1 \vee 0 = 1 = y$. Under the modified formulation [Halpern, 2015], witnesses are restricted to actual values; since X_2's actual value is 1, no singleton witness can break the redundancy. However, HP_m recovers the verdict through multivariate causes: $\{X_1, X_2\} = \{1, 1\}$ jointly causes $Y = 1$ with empty witness, and neither singleton suffices alone, so the set is minimal (AC3). Without the multivariate route, normality assumptions would be required.

BV (correctly accepts). $X_1 = 1$ directly NESS-causes $Y = 1$: $\{X_1{=}1\}$ is sufficient for $Y = 1$ (since $f_Y(1, X_2) = 1$ for all X_2) and $\varnothing$ is not. Asymmetry: in $\mathcal{M}_{X_1=0}$, $Y = 0 \vee 1 = 1$. $\{X_1{=}0\}$ is not sufficient for $Y = 1$ (since $f_Y(0, X_2) = X_2$, which is not always 1), so $X_1 = 0$ cannot directly NESS-cause $Y = 1$ and cannot start any NESS chain to $Y = 1$. Condition (ii) holds; $X_1 = 1$ is a BV-cause.

CNESS (correctly accepts). Same reasoning as BV: $X_1 = 0$ cannot NESS-cause $Y = 1$ along any path, so condition (ii) holds.

Ours (correctly accepts). $X_1 = 1$ is a cause; see derivation in Sec. 3.

C.4 FIRING SQUAD

Example C.1 (Firing Squad [Halpern, 2016a, Ex. 2.3.2])**.** A captain gives an order ($C = 1$). Rifleman A and rifleman B each fire if the captain orders. The prisoner dies (D) if either rifleman fires.

$$A = C, \tag{10}$$
$$B = C, \tag{11}$$
$$D = A \vee B. \tag{12}$$

In the actual world, the captain orders ($C = 1$), both riflemen fire ($A = B = 1$), and the prisoner dies ($D = 1$). The question is whether the captain's order caused the prisoner's death.

Analysis. To test whether $C = 1$ causes $D = 1$, take witness $\mathbf{w} = \varnothing$ with $c' = 0$: then $A = 0$, $B = 0$, $D = 0$, flipping the outcome with an empty, hence minimal, witness. For C2, the path $C \to A \to D$ has intermediate A changing from 1 to 0. We verify that $A = 1$ is a cause of $D = 1$. For C1, $D_{a',b'} = 0$ with witness $\{B = 0\}$ and $a' = 0$; this is minimal since the empty witness fails when $B = C = 1$. For C2, the direct path $A \to D$ is valid. For C3, in $\mathcal{M}_{A=0}$: $D = 0 \vee 1 = 1$. C1 for $A = 0$ requires $a'' = 1$ and witness $\mathbf{w}'$ with $D_{a'',\mathbf{w}'} \neq 1$; but $D = 1 \vee B = 1$ regardless of $\mathbf{w}'$, so C1 fails and C3 holds. For C3, in $\mathcal{M}_{C=0}$: $A = 0$, $B = 0$, $D = 0 \neq 1$, so the outcome does not obtain and C3 holds. Therefore, $C = 1$ *is* a cause of $D = 1$. Each rifleman is also a cause, by symmetric overdetermination.

C.5 XOR

Example C.2 (Three-Switch Lamp)**.** Three switches (S_1, S_2, S_3) control a lamp (L) via exclusive-or.

$$L = S_1 \oplus S_2 \oplus S_3. \tag{13}$$

In the actual world, $S_1 = 1$, $S_2 = 0$, $S_3 = 0$, and $L = 1$. The question is which switches caused the lamp to be on.

Analysis. $S_1 = 1$ causes $L = 1$: with empty witness and $s_1' = 0$, $L = 0 \oplus 0 \oplus 0 = 0$, flipping the outcome. The direct path $S_1 \to L$ satisfies C2, and in $\mathcal{M}_{S_1=0}$, $L = 0 \neq 1$, so the outcome does not obtain and C3 holds.

$S_2 = 0$ causes $L = 1$: with empty witness and $s_2' = 1$, $L = 1 \oplus 1 \oplus 0 = 0$, flipping the outcome. The direct path $S_2 \to L$ satisfies C2, and in $\mathcal{M}_{S_2=1}$, $L = 0 \neq 1$, so factuality fails and C3 holds. Likewise $S_3 = 0$ is a cause. All three switches are causes of L, including those set to 0. XOR provides no redundancy: every input is counterfactually necessary given the other inputs, so the structure admits no backup pathway.

C.6 TRUMPING PREEMPTION

Example C.3 (Colonel and Sergeant [Halpern, 2016a, Ex. 3.1])**.** A colonel ($C = 1$) and a sergeant ($S = 1$) both order soldiers to advance (A). Soldiers obey the highest-ranking order.

$$A = C. \tag{14}$$

In the actual world, $C = 1$, $S = 1$, and $A = 1$. The question is whether $C = 1$, $S = 1$, or both are causes of $A = 1$.

Analysis. $C = 1$ causes $A = 1$: with empty witness and $c' = 0$, $A = 0$, flipping the outcome. The direct path $C \to A$ satisfies C2 since it has no intermediates, and in $\mathcal{M}_{C=0}$, $A = 0 \neq 1$, so C3 holds.

$S = 1$ does *not* cause $A = 1$: the structural equation $A = C$ does not reference S, so no edge from S to A exists in $G(\mathcal{M})$. Without a directed path from S to A, C2 fails immediately. Therefore, $C = 1$ *is* a cause of $A = 1$, while $S = 1$ is *not*. The trumping relationship must be encoded in the structural equations. If modeled as $A = C \vee S$, which is symmetric overdetermination, both would be causes.

C.7 DOUBLE PREVENTION

Example C.4 (Fighter and Bomber)**.** A fighter ($F = 1$) shoots down an interceptor (I). The interceptor would have shot down a bomber (B). The bomber destroys a target (T).

$$I = \neg F, \tag{15}$$

$$B = \neg I, \tag{16}$$

$$T = B. \tag{17}$$

In the actual world, $F = 1$, $I = 0$, $B = 1$, and $T = 1$. The question is whether the fighter's action ($F = 1$) caused the target's destruction ($T = 1$).

Analysis. To test whether $F = 1$ causes $T = 1$, take empty witness with $f' = 0$: then $I = 1$, $B = 0$, $T = 0$, flipping the outcome; the witness is minimal. For C2, the path $F \to I \to B \to T$ has intermediates I and B both changing under $do(F = 0)$. We verify recursively that $I = 0$ is a cause of $T = 1$. For C1, with empty witness and $i' = 1$: $B = 0$ and $T = 0$. For C2, the path $I \to B \to T$ has the bomber shifting from reaching the target ($B = 1$) to being intercepted ($B = 0$). The bomber reaching the target ($B = 1$) is a cause of the target's destruction ($T = 1$) via the direct path $B \to T$. C3 holds because the target survives in $\mathcal{M}_{B=0}$ ($T = 0 \neq 1$), so the outcome does not obtain. For C3, in $\mathcal{M}_{I=1}$: $T = 0 \neq 1$, so the outcome does not obtain and C3 holds. We also verify that the bomber reaching the target ($B = 1$) is a cause of the target's destruction ($T = 1$). The direct path $B \to T$ satisfies C2. For C1, $T_{b'} = 0 \neq 1$ with an empty witness. In $\mathcal{M}_{B=0}$ the target survives ($T = 0 \neq 1$), so the outcome does not obtain and C3 holds. For C3, in $\mathcal{M}_{F=0}$: $T = 0 \neq 1$, so the outcome does not obtain and C3 holds. Therefore, $F = 1$ *is* a cause of $T = 1$. The recursive C2 check traces through the prevention chain $F \to I \to B \to T$, verifying that each intermediate along the double-negation chain satisfies the full causation definition.

C.8 EARLY PREEMPTION

In *late* preemption (Ex. 3.1), the backup process is physically blocked: $BH = BT \wedge \neg SH$. In *early* preemption, the backup agent voluntarily withdraws upon observing the primary cause, giving $BT = BT_{\text{exo}} \wedge \neg ST$. Unlike late preemption, this model is structurally a switch: the edge $ST \to BT$ means ST selects which pathway produces BS. All endogenous variables shift under $do(ST = 0)$ (via $SH = ST$ and $BT = BT_{\text{exo}} \wedge \neg ST$), so the witness restriction gives $\mathbf{W} = \emptyset$. With empty witness, $BS_{st'=0} = 0 \vee 1 = 1 = y$: the backup activates, the outcome does not change, C1 fails, and the definition rejects $ST = 1$. The equation $BT = BT_{\text{exo}} \wedge \neg ST$ treats Billy's withdrawal as a mechanical consequence of Suzy's throw, encoding the symmetric dependency that characterizes switches. The coarse equation treats withdrawal as a mechanical consequence, which is exactly what the legal doctrine of *novus actus interveniens* (NAI) denies [Hart and Honoré, 1985]: a voluntary intervening act breaks the causal chain precisely because the agent acts as an autonomous source, not a passive relay.

Recovery via outcome individuation. The coarse model fails to represent early preemption faithfully because it discards the identity of the agent whose throw caused the shattering. Fine-graining BS to record which agent's throw shattered the bottle (as discussed in Sec. 3.5) restores this information. Let $BS \in \{\text{Suzy}, \text{Billy}, \perp\}$ with actual value $BS = \text{Suzy}$. Under $do(ST = 0)$: $SH = 0$, $BT = 1$, $BH = 1$, so $BS = \text{Billy} \neq \text{Suzy}$; C1 holds with $\mathbf{W} = \emptyset$. C2 holds via path $ST \to SH \to BS$: SH changes and directly causes $BS = \text{Suzy}$. For C3, in $\mathcal{M}_{ST=0}$ setting $ST = 1$ gives $BS = \text{Suzy}$; no witness can flip BS away from Suzy when $ST = 1$ forces $SH = 1$, so C1 fails for $ST = 0$ and C3 holds. Therefore $ST = 1$ *is* a cause of $BS = \text{Suzy}$.

Recovery via novus actus interveniens. Alternatively, declare BT an actor node (a variable whose value is determined by voluntary choice). To evaluate causation under the NAI extension, sever all incoming edges to BT, fix it to its actual value $BT(\mathbf{u}) = 0$, and apply Definition 3.3 in the resulting submodel; actor nodes are thereby eligible witnesses even when edges $X \to A$ exist in the original graph. Taking $\mathbf{W} = \{BT = 0\}$ with $st' = 0$: $SH = 0$, $BH = 0$, $BS = 0 \neq 1$; C1 holds. Actor nodes must be declared by the modeler, but the NAI doctrine already requires identifying which act constitutes the intervening voluntary agency.

Symmetric timing. If both rocks arrive simultaneously with no temporal asymmetry encoded, the model reduces to $SH = ST$, $BH = BT$, and $BS = SH \vee BH$ with $SH = 1$, $BH = 1$, $BS = 1$. This is symmetric overdetermination (Sec. 3), and both throws are causes. The asymmetry introduced by encoding temporal precedence, e.g. $BH = BT \wedge \neg SH$ in late preemption, is essential to distinguish preemption from overdetermination.

C.9 BOULDER (NON-TRANSITIVITY)

Example C.5 (Boulder [Hitchcock and Knobe, 2009])**.** A boulder (B) rolls toward a hiker. The hiker sees the boulder and ducks (D). The boulder passes overhead and the hiker survives (S).

$$D = B, \tag{18}$$

$$S = \neg B \vee D. \tag{19}$$

In the actual world, $B = 1$, $D = 1$, and $S = 1$. The question is whether the boulder ($B = 1$) caused the hiker's survival ($S = 1$).

Analysis. The causal graph has edges $B \to D$, $B \to S$, and $D \to S$. Every endogenous variable shifts under $do(B = 0)$ (D changes from 1 to 0), so the witness must be empty. With $\mathbf{W} = \emptyset$ and $b' = 0$: $D_{b'} = 0$ and $S_{b'} = \neg 0 \vee 0 = 1 = s$. The outcome does not change, so C1 fails and $B = 1$ is *not* a cause of $S = 1$.

The boulder triggers both a threat (the direct $\neg B$ term in S) and its own prevention ($B \to D \to S$); the two effects cancel, and no intervention on B alone changes the outcome.

Non-transitivity. The boulder illustrates that productive actual causation is not transitive. $B = 1$ is a cause of $D = 1$: flipping B to 0 gives $D = 0$, and C3 holds because $D = 0 \neq 1$ in $\mathcal{M}_{B=0}$. $D = 1$ is a cause of $S = 1$: flipping D to 0 gives $S = \neg 1 \vee 0 = 0 \neq 1$, with witness $\mathbf{W} = \emptyset$; C3 holds because $S = \neg 1 \vee 0 = 0 \neq 1$ in $\mathcal{M}_{D=0}$. Yet $B = 1$ is not a cause of $S = 1$, as shown above. Non-transitivity is standard for structural definitions of actual causation and widely considered desirable [Hall, 2004]; the boulder is one of the canonical motivating examples.

C.10 PRISONER (GUN LOADER)

Example C.6 (Gun Loader [Halpern, 2016a, Ex. 2.8.1])**.** A loads B's gun ($A = 1$), B does not shoot ($B = 0$), C independently loads and fires his own gun ($C = 1$), and the prisoner dies ($D = 1$).

$$D = (A \wedge B) \vee C. \tag{20}$$

In the actual world, $A = 1$, $B = 0$, $C = 1$, and $D = 1$. A loaded a gun that was never fired. The question is whether $A = 1$ caused $D = 1$.

Verdict. $C = 1$ is a cause of $D = 1$; $A = 1$ is not.

$C = 1$ is a cause. C1: with empty witness, $D_{C=0} = (1 \wedge 0) \vee 0 = 0 \neq 1$; the singleton $\{C\}$ is minimal. C2: vacuous via the direct path $C \to D$. C3: in $\mathcal{M}_{C=0}$, $D = A \wedge B = 0 \neq 1$, so factuality fails for $C = 0$ and C3 holds.

$A = 1$ is not a cause. With empty witness: $D_{A=0} = (0 \wedge 0) \vee 1 = 1$; no flip. With $\mathbf{W} = \{C\}$, $\mathbf{w} = \{C = 0\}$: $D_{A=0,C=0} = 0$, a flip. But $\{A\} \cup \{C\}$ is not minimal: $\{C = 0\}$ alone gives $D_{C=0} = (1 \wedge 0) \vee 0 = 0$, so C alone flips D. C1 fails for every witness set, and $A = 1$ is correctly rejected.

Role of C1 minimality. The original HP definition [Halpern and Pearl, 2005] incorrectly accepts $A = 1$ using the witness $B = 1$, $C = 0$, which changes B from its actual value; this motivated the updated and modified formulations [Halpern, 2016a, 2015], both of which correctly reject A. Under our definition, the rejection follows from C1 minimality over $\{X\} \cup \mathbf{W}$: any witness set containing C fails minimality because C alone suffices to flip D. This illustrates C1's joint minimality requirement: a candidate cause is rejected when its effect on the outcome is redundant given the witness.

C.11 THREE-SWITCH LAMP (MAJORITY MATCH)

Example C.7 (Three-Switch Lamp [Weslake, 2015]; [Halpern, 2015, Ex. 3.5])**.** Three switches $A, B, C \in \{-1, 0, 1\}$ control a lamp $L \in \{0, 1\}$:

$$L = \mathbf{1}[(A = B) \vee (B = C) \vee (A = C)]. \tag{21}$$

The lamp turns on whenever at least two switches match. In the actual world, $A = 1$, $B = -1$, $C = -1$, and $L = 1$ (since $B = C$). The question is whether $A = 1$ caused $L = 1$.

$A = 1$ **is not a cause.** Since A has no descendants other than L, both B and C are eligible witnesses for any a'. First, $\{A\}$ alone (empty witness) never flips L: for any $a' \in \{0, -1\}$, the clause $(B = C) = (-1 = -1)$ remains true, so $L = 1$. C1 therefore requires a nonempty witness $\mathbf{W}$ drawn from $\{B, C\}$. But every eligible variable can flip L on its own. For B: setting $B = 0$ gives $L = \mathbf{1}[(1 = 0) \vee (0 = -1) \vee (1 = -1)] = 0$, so $\{B\}$ admits an assignment flipping L. For C: setting $C = 0$ gives $L = \mathbf{1}[(1 = -1) \vee (-1 = 0) \vee (1 = 0)] = 0$, so $\{C\}$ admits an assignment flipping L. Every witness set $\mathbf{W}$ containing B or C therefore has a strict subset that flips L, making $\{A\} \cup \mathbf{W}$ non-minimal. Since C1 fails for every witness, $A = 1$ is *not* a cause.

$B = -1$ **is a cause.** With empty witness and $b' = 0$: $L = \mathbf{1}[(1 = 0) \vee (0 = -1) \vee (1 = -1)] = 0 \neq 1$; the singleton is minimal. C2: vacuous (direct edge $B \to L$). For C3, in $\mathcal{M}_{B=0}$: $L = \mathbf{1}[(1 = 0) \vee (0 = -1) \vee (1 = -1)] = 0 \neq 1$. Factuality fails ($L \neq 1$), so $B = 0$ cannot cause $L = 1$ and C3 holds. Therefore, $B = -1$ *is* a cause.

Comparison. The original HP definition incorrectly accepts $A = 1$ using the non-actual witness $B = 1$ (changing B from its actual value -1). HP_m blocks this by restricting witnesses to actual values. Our definition blocks it because C1 fails for every eligible witness: the ternary domain allows each witness variable to flip L on its own, so adding A to any flipping set is never minimal.

C.12 SHORT CIRCUIT

Example C.8 (Short Circuit [Hall, 2007, §5.3])**.** Neurons A through F form an inhibitory network:

$$B = C, \qquad D = C, \tag{22}$$

$$F = D \cdot (1 - B), \qquad E = A \cdot (1 - F). \tag{23}$$

Neuron C fires, activating both D and B. Because B inhibits F, we get $F = 1 \cdot (1-1) = 0$, and E fires ($E = 1 \cdot (1-0) = 1$). In the actual world, $A = 1$, $C = 1$, $B = 1$, $D = 1$, $F = 0$, and $E = 1$. C initiates a threat to E (via $D \to F \to E$) but simultaneously cancels it (via $B \to F$). The question is whether $C = 1$ caused $E = 1$.

$C = 1$ **is not a cause.** Under $do(C = 0)$: $B_{c'} = 0 \neq 1 = B(\mathbf{u})$ and $D_{c'} = 0 \neq 1 = D(\mathbf{u})$, so B and D are *ineligible* witnesses. The remaining endogenous variables are F and A: $F_{c'} = 0 \cdot (1 - 0) = 0 = F(\mathbf{u})$ and $A_{c'} = 1 = A(\mathbf{u})$, so both are eligible. We enumerate all witness subsets of $\{A, F\}$:

- $\mathbf{W} = \varnothing$: $E_{c'} = 1 \cdot (1 - 0) = 1 = e$; no flip.
- $\mathbf{W} = \{F\}$, $\mathbf{w} = \{F = 1\}$: $E = 1 \cdot (1 - 1) = 0 \neq 1$; but $\{F = 1\}$ alone gives $E = 0$, so F alone flips E and the set is not minimal.
- $\mathbf{W} = \{A\}$, $\mathbf{w} = \{A = 0\}$: $E = 0 \cdot (1 - 0) = 0 \neq 1$; but $\{A = 0\}$ alone gives $E = 0$, so A alone flips E and the set is not minimal.
- $\mathbf{W} = \{A, F\}$: any flipping assignment reduces to A or F alone; not minimal.

C1 fails for every eligible witness set. $C = 1$ is *not* a cause of $E = 1$.

Comparison. The original HP definition incorrectly accepts $C = 1$ by freezing D at its actual value 1 while setting $C = 0$, manufacturing a scenario where D fires despite C not firing. HP_m restricts witnesses to actual values but does not require that they remain stable under the intervention; since $D = 1$ is the actual value, D is a valid HP_m witness and HP_m also incorrectly accepts $C = 1$. Our definition adds the eligibility constraint $V_{x'} = V(\mathbf{u})$: since $D_{c'=0} = 0 \neq 1 = D(\mathbf{u})$, D is ineligible, and without D as a witness, no subset of eligible variables satisfies C1.

C.13 OMISSION

Example C.9 (Plant and Gardener)**.** A gardener neglects to water a plant ($W = 0$). Without water, the plant dies ($D = 1$).

$$D = \neg W. \tag{24}$$

In the actual world, $W = 0$ and $D = 1$. The question is whether the gardener's failure to water caused the plant's death.

Analysis. With empty witness and $w' = 1$: $D_{w'} = \neg 1 = 0 \neq 1$, flipping the outcome; the singleton is minimal. C2 holds via the direct path $W \to D$. For C3, in $\mathcal{M}_{W=1}$: $D = 0 \neq 1$, so factuality fails and C3 holds. Therefore, $W = 0$ *is* a cause of $D = 1$.

All purely structural definitions accept omissions as causes because the counterfactual dependence is clear. Whether omissions are "real" causes is debated in philosophy [Halpern, 2016a]; in legal and moral contexts, the verdict often depends on duty or negligence, which can be captured by a normality ordering layered on top of the structural definition (Sec. C.14).

C.14 NORMALITY AND PRAGMATIC JUDGMENTS

Structural definitions identify all variables that structurally produce the outcome. Some causal intuitions, however, depend on which variable settings are *normal* or *expected* rather than on structural features. Two canonical cases illustrate this.

Example C.10 (Bodyguard)**.** An assassin does not poison the victim ($A = 0$). A bodyguard adds an antidote ($B = 1$). The victim (V) survives if either there is no poison or there is an antidote.

$$V = \neg A \vee B. \tag{25}$$

In the actual world, $A = 0$, $B = 1$, and $V = 1$.

Under our definition, $B = 1$ is a cause of $V = 1$: witness $\{A = 1\}$ with $b' = 0$ gives $V_{b',a'} = 0$; C2 holds via the direct path $B \to V$. For C3, in $\mathcal{M}_{B=0}$: $V = \neg A \vee 0 = \neg A$, so $V = 1$ (since $A = 0$). C1 for $B = 0$ requires $b' = 1$ and a witness $\mathbf{w}'$ with $V_{b',\mathbf{w}'} \neq 1$; but $V_{b',\mathbf{w}'} = \neg A \vee 1 = 1$ for any $\mathbf{w}'$, so C1 fails and C3 holds.

Writing $P = \neg A$, the equation becomes $V = P \vee B$ with $P = 1$, $B = 1$: symmetric overdetermination (Sec. 3). The intuition that the bodyguard "prevented nothing" relies on the *normality* of $A = 0$: the assassin not poisoning is the expected state, so $\neg A = 1$ is a "default" pathway and B is redundant. This is a pragmatic judgment, not a structural one.

Example C.11 (Match and Oxygen)**.** A match is struck ($X_1 = 1$) in the presence of oxygen ($X_2 = 1$). A fire (Y) starts if both are present.

$$Y = X_1 \wedge X_2. \tag{26}$$

In the actual world, $X_1 = 1$, $X_2 = 1$, and $Y = 1$.

Both $X_1 = 1$ and $X_2 = 1$ are causes of $Y = 1$: each is but-for necessary, and C3 holds since $Y = 0$ in the respective counterfactual models. The intuition that the match "caused" the fire while oxygen was merely a "background condition" again reflects normality: oxygen is the default state, the match is the abnormal event.

Normality as a separate layer. All purely structural definitions (HP, BV, CNESS, and ours) accept $B = 1$ as a cause in the bodyguard model and both X_1 and X_2 in the match model. Distinguishing "salient" causes from "background" causes requires a normality ordering over variable settings [Halpern, 2016a, Chapter 3], which ranks some worlds as more normal than others and restricts witnesses to abnormal settings. This normality layer is compatible with the structural definition: it can be applied on top of any base definition, including ours. We do not incorporate normality here because the switches–preemption trade-off is a purely structural problem. Mixing structural and pragmatic conditions obscures whether each verdict follows from the structural definition or from the normality ordering. Some accounts treat normality as essential rather than supplementary, arguing that causal judgments are inherently graded by typicality [Hitchcock and Knobe, 2009, Halpern and Hitchcock, 2015]. We do not dispute this but maintain that one should evaluate the structural definition on its own merits before layering normality on top. Halpern [2015] use normality to address switches [Halpern, 2015, Example 3.9], but this approach faces difficulties with conjoined scenarios [Wysocki, 2023]; our definition resolves switches structurally, without normality assumptions.

D NOVEL EXAMPLES

D.1 SHADOW VARIABLE

Example D.1 (Shadow Variable)**.** A controller ($X_1 = 1$) sets a mode. A relay (X_2) inverts the controller's signal. A device (Y) activates only when the controller is off and the relay fires.

$$X_2 = \neg X_1, \tag{27}$$
$$Y = \neg X_1 \wedge X_2. \tag{28}$$

In the actual world, $X_1 = 1$, $X_2 = 0$, and $Y = 0$. Since $X_2 = \neg X_1$, the equation reduces to $Y = \neg X_1 \wedge \neg X_1 = \neg X_1$. The outcome is entirely determined by X_1; X_2 only replicates X_1's value and adds no independent effect. Intuitively, $X_2 = 0$ should not be a cause of $Y = 0$.

HP, BV, and CNESS all incorrectly accept $X_2 = 0$ as a cause. HP finds a valid witness ($\mathbf{w} = \{X_1 = 0\}$ satisfies AC2), but HP's minimality condition (AC3) applies to the cause set X, not to $\{X\} \cup \mathbf{W}$, so it does not detect that X_1 alone suffices. BV and CNESS accept because $X_2 = 0$ appears in a minimal sufficient set for Y's equation; their sufficiency tests check which value assignments satisfy the equation rather than which interventions change the outcome. Our C1 catches the distinction: because $X_2 = \neg X_1$, any intervention that flips Y via X_2 also requires changing X_1, and X_1 alone suffices.

Four-definition comparison. The structural equations are $X_2 = \neg X_1$ and $Y = \neg X_1 \wedge X_2$, with actual values $X_1 = 1$, $X_2 = 0$, and $Y = 0$. We test whether $X_2 = 0$ is a cause of $Y = 0$.

HP (incorrectly accepts). Under the original formulation, set $\mathbf{W} = \{X_1\}$ to 0 (non-actual) and $x_2' = 1$: then $Y_{x_2', x_1} = \neg 0 \wedge 1 = 1 \neq 0$, so AC2(a) holds. For AC2(b), $Z = \varnothing$ and $Y_{X_2=0, X_1=0} = \neg 0 \wedge 0 = 0 = y$. Under the modified formulation, $\mathbf{W} = \{X_1\}$ must use actual value 1: $Y_{x_2'=1, X_1=1} = \neg 1 \wedge 1 = 0 = y$, so the outcome does not flip and AC2 fails. The modified definition correctly rejects $X_2 = 0$.

BV (incorrectly accepts). $X_2 = 0$ directly NESS-causes $Y = 0$: $\{X_2{=}0\}$ is sufficient for $Y = 0$ (since $f_Y(X_1, 0) = \neg X_1 \wedge 0 = 0$ for all X_1) and $\varnothing$ is not (since $f_Y(0, 1) = 1$). Asymmetry: in $\mathcal{M}_{X_2=1}$, $X_1 = 1$, $Y = \neg 1 \wedge 1 = 0$. $\{X_2{=}1\}$ is not sufficient for $Y = 0$ (since $f_Y(0, 1) = 1$). The only other candidate set $\{X_2{=}1, X_1{=}1\}$ makes X_2 unnecessary because $\{X_1{=}1\}$ alone is sufficient for $Y = 0$ (as $f_Y(1, X_2) = 0$). Since X_2 has no outgoing edges except to Y, no indirect NESS chain exists. So $X_2 = 1$ does not NESS-cause $Y = 0$; condition (ii) holds and $X_2 = 0$ is a BV-cause.

CNESS (incorrectly accepts). $X_2 = 0$ NESS-causes $Y = 0$ along the direct path $p = (X_2, Y)$. In $\mathcal{M}_{X_2=1}$, $X_2 = 1$ does not NESS-cause $Y = 0$ along any subpath of p (shown above). Condition (ii) holds; $X_2 = 0$ is a CNESS-cause.

Ours (correctly rejects; see Ex. D.1). C1 fails: the only candidate is $x_2' = 1$ with witness $\{X_1 = 0\}$, giving $Y = \neg 0 \wedge 1 = 1 \neq 0$. But $\{X_2, X_1\}$ is not minimal: $X_1 = 0$ alone gives $X_2 = 1$ and $Y = 1 \neq 0$. With empty witness, $Y_{x_2'=1} = \neg 1 \wedge 1 = 0 = y$, so Y does not flip. $X_2 = 0$ is not a cause.

D.2 FILE SERVER

The structural equation is $Y = f(X_1, X_2)$ with $f(1, 0) = A$, $f(0, 1) = B$, and $f(0, 0) = .$. Alice (X_1) and Bob (X_2) each send or withhold a request; the conflict resolution policy determines $f(1, 1)$. We consider four cases: lock ($f(1, 1) = A$), overwrite ($f(1, 1) = B$), merge ($f(1, 1) = AB$), and the coarsened binary outcome $E = \mathbb{1}[Y \neq .]$. In all cases, both clients send ($X_1 = X_2 = 1$).

Case (a): Lock mode ($f(1, 1) = A$, $Y = A$). The server retains the first writer's content when both send. All four definitions agree: $X_1 = 1$ is a cause of $Y = A$; $X_2 = 1$ is not.

HP. For $X_1 = 1$: with empty witness and $x_1' = 0$, $Y = B \neq A$; AC2 holds. For $X_2 = 1$: with empty witness and $x_2' = 0$, $Y = A = y$ (does not flip). With $\mathbf{W} = \{X_1\}$ at non-actual value 0, $Y = . \neq A$, so AC2(a) holds. However, AC2(b) requires $Y_{X_2=1, X_1=0} = A$; since $f(0, 1) = B \neq A$, AC2(b) fails. No valid witness exists; $X_2 = 1$ is not an HP-cause.

BV. $\{X_1 = 1\}$ is sufficient for $Y = A$ (since $f(1, X_2) = A$ for all X_2) and $\varnothing$ is not; $X_1 = 1$ directly NESS-causes $Y = A$. Asymmetry: in $\mathcal{M}_{X_1=0}$, $Y = B \neq A$, so $X_1 = 0$ cannot NESS-cause $Y = A$; condition (ii) holds. For $X_2 = 1$: $\{X_2 = 1\}$

is not sufficient for $Y = A$ because $f(0, 1) = B$, and in $\{X_1 = 1, X_2 = 1\}$, X_2 is redundant since $\{X_1 = 1\}$ alone suffices. $X_2 = 1$ does not NESS-cause $Y = A$.

CNESS. Same reasoning as BV: $X_2 = 1$ cannot appear in any NESS chain to $Y = A$.

Ours (see Sec. 3.5). For $X_1 = 1$: empty witness, $x_1' = 0$: $Y = B \neq A$; direct path $X_1 \to Y$ satisfies C2. In $\mathcal{M}_{X_1=0}$, $Y = B \neq A$, so factuality fails for $X_1 = 0$ and C3 holds. For $X_2 = 1$: empty witness, $x_2' = 0$: $Y = A = y$ (does not flip). With witness $\{X_1 = 0\}$: $Y = . \neq A$, but $\{X_1\}$ alone suffices since $f(0, 1) = B \neq A$, so $\{X_2, X_1\}$ is not minimal and C1 fails.

Case (b): Overwrite mode ($f(1, 1) = B$, $Y = B$). The later writer overwrites the earlier writer's content. By symmetry with case (a), all four definitions agree: $X_2 = 1$ is a cause of $Y = B$; $X_1 = 1$ is not. Bob posted after Alice; this timing is encoded in the conflict resolution $f(1, 1) = B$.

Case (c): Merge mode ($f(1, 1) = AB$, $Y = AB$). Both files are retained when both send. Each client is but-for necessary: $f(0, 1) = B \neq AB$ and $f(1, 0) = A \neq AB$. All four definitions agree: both $X_1 = 1$ and $X_2 = 1$ are causes of $Y = AB$.

Case (d): Coarsened binary ($E = \mathbb{1}[Y \neq .]$, $E = 1$). The outcome is coarsened to "some file appeared." The equation reduces to $E = X_1 \vee X_2$: symmetric overdetermination. All four definitions agree: both $X_1 = 1$ and $X_2 = 1$ are causes of $E = 1$.

Discussion. The same physical scenario yields four different verdict patterns depending on the server's conflict resolution and the outcome granularity. Lock mode identifies the first writer as the sole cause; overwrite mode identifies the last writer. Merge preserves each client's distinct contribution, making both causes. Coarsening to a binary event loses all manner information, reducing the scenario to symmetric overdetermination. All four definitions agree in every case. The contribution is representational: modeling manner as distinct outcome values lets every definition leverage information that a binary coarsening discards.

D.3 REDUCED SWITCH

The reduced switch (Ex. 3.5) is stated in the main text. The witness restriction rejects $X = 1$: W shifts under $do(X = 0)$ ($W_{x'} = 1 \neq 0 = W(\mathbf{u})$), so W cannot appear in $\mathbf{W}$. With $\mathbf{W} = \emptyset$: $Y_{x'=0} = 0 \vee 1 = 1 = y$, so C1 fails. No eligible witness exists, and the definition correctly rejects $X = 1$.

D.4 MULTIPLEXER SWITCH

Example D.2 (Multiplexer Switch)**.** A multiplexer (Y) selects between two inputs (Z, Q) based on a control signal (X). All edges are direct: $X \to Y$, $Z \to Y$, $Q \to Y$.

$$Y = (X \wedge Z) \vee (\neg X \wedge Q). \tag{29}$$

In the actual world, $X = 1$, $Z = 1$, $Q = 1$, and $Y = 1$. This is a tautological switch: $Y = 1$ regardless of X.

Analysis. For C1, take witness $\mathbf{w} = \{Q = 0\}$ with $x' = 0$: then $Y = (0 \wedge 1) \vee (1 \wedge 0) = 0 \neq 1$. The set $\{X, Q\}$ is minimal: $X = 0$ alone gives $Y = (0 \wedge 1) \vee (1 \wedge 1) = 1$, and $Q = 0$ alone gives $Y = (1 \wedge 1) \vee (0 \wedge 0) = 1$. C1 holds.

C2 holds via the direct path $X \to Y$ (no intermediates).

For C3, check whether $X = 0$ satisfies C1 and C2 for $Y = 1$ in $\mathcal{M}_{X=0}$. In $\mathcal{M}_{X=0}$: $Y = (0 \wedge Z) \vee (1 \wedge Q) = Q = 1$. For C1, take $x' = 1$ with witness $\mathbf{w}' = \{Z = 0\}$: $Y = (1 \wedge 0) \vee (0 \wedge 1) = 0 \neq 1$. The set $\{X, Z\}$ is minimal by symmetric reasoning. C2 holds via the direct path. Since $X = 0$ satisfies C1 and C2, C3 fails: $X = 1$ is *not* a cause of $Y = 1$.

C3 catches the switch because the alternative value $X = 0$ symmetrically satisfies C1 and C2 via the complementary input Z: both values produce Y through parallel pathways, making the production symmetric.

E PROOFS

Lemma E.1 (Termination)**.** *Alg. 1 terminates for any finite SCM.*

Proof. CAUSES recurses only through C2(b), which calls CAUSES$(\mathcal{M}, \mathbf{u}, P_i, Y)$ for intermediates P_i on a directed path from X to Y. Each such P_i satisfies $\lambda(P_i) < \lambda(X)$, where $\lambda(V)$ is the longest directed path length from V to Y. Since λ strictly decreases, the recursion terminates in at most $|\mathbf{V}|$ steps.

C3 calls PRODUCES$(\mathcal{M}_{X=x'}, \mathbf{u}, X, Y, x'')$ in the submodel $\mathcal{M}_{X=x'}$. PRODUCES invokes CAUSES for intermediates via C2(b), which may in turn trigger further C3 evaluations for different variables in deeper submodels. Crucially, each C3 call fixes one additional variable, so the fixing set $\mathrm{Fix}(\mathcal{M})$ grows strictly. Within any single model, C2(b) is well-founded by λ. The lexicographic order $\big(|\mathbf{V}| - |\mathrm{Fix}(\mathcal{M})|,\ \lambda(X)\big)$ strictly decreases at every recursive call: C3 decreases the first component, C2(b) decreases the second. Since both components are bounded by $|\mathbf{V}|$, the recursion terminates.

For the complexity bound (used in Thm. 4.2), the number of distinct submodels visited by C3 is at most $(d{+}1)^{|\mathbf{V}|}$ in general, or $(d{+}1)^{O(k^L)}$ for bounded in-degree k and path length L, since only ancestors of Y are fixed. □

Lemma E.2 (PSPACE membership)**.** *Checking whether $X = x$ causes $Y = y$ under Def. 3.3 is in* PSPACE*.*

Proof. Alg. 1 can be implemented to recompute (rather than store) each submodel's results on demand. There are two sources of recursion depth. Within any single model, C2(b) recurses on intermediates with strictly decreasing λ, contributing at most $|\mathbf{V}|$ stack frames. Across models, each C3 call fixes one additional variable; since the fixing set can grow at most $|\mathbf{V}|$ times, there are at most $|\mathbf{V}|$ levels of model nesting. At each nesting level, C2(b) may add up to $|\mathbf{V}|$ frames. The total stack depth is therefore $O(|\mathbf{V}|^2)$. Each stack frame stores which variables are fixed and their values, requiring $O(|\mathbf{V}|)$ space. The total space is $O(|\mathbf{V}|^3)$, which is polynomial. □

Proof of Prop. 4.1. Suppose $(x', \mathbf{w}', \pi)$ jointly satisfy C1–C2 for the directed path $\pi = [X, P_2, \ldots, P_m, Y]$. We show that restricting the witness to eligible variables in the ancestor frontier $F^*(\pi) = \mathrm{Anc}(\pi) \setminus \pi$ preserves C1–C2.

Since f_{P_i} depends only on $\mathbf{Pa}_{P_i}$, and $\mathbf{Pa}_{P_i} \subseteq \pi \cup F^*(\pi)$ for every path node P_i, the values along π under any intervention $do(x', \mathbf{w})$ are determined by x' and the induced values of the non-path ancestors.

Construct $\mathbf{W}^*$ as follows: for each eligible variable $Z \in F^*(\pi)$ with $Z_{x'} = Z(\mathbf{u})$, include Z in $\mathbf{W}^*$ with value $w^*_Z = Z_{x',\mathbf{w}'}$ (the value Z takes under the full original intervention). We show by a single topological induction over $\{X\} \cup \pi \cup F^*(\pi)$ that every variable in this set takes the same value under $do(x', \mathbf{w}^*)$ as under $do(x', \mathbf{w}')$.

Process variables in topological order (ancestors before descendants). For each variable V:

- $V = X$: set to x' under both interventions.
- $V \in \mathbf{W}^*$ (eligible, in $F^*(\pi)$): $V_{x',\mathbf{w}^*} = w^*_V = V_{x',\mathbf{w}'}$ by construction.
- V on π or $V \in F^*(\pi)$ ineligible: V is not directly intervened on under either $do(x', \mathbf{w}')$ or $do(x', \mathbf{w}^*)$ (ineligible variables cannot appear in any witness set, and path nodes lie outside $\mathbf{W}'$ by C2). Hence $V_{x',\mathbf{w}} = f_V\big((\mathbf{Pa}_V)_{x',\mathbf{w}}\big)$ under both interventions. Every parent of V precedes V in topological order and belongs to $\{X\} \cup \pi \cup F^*(\pi)$, so by the inductive hypothesis all parents take the same values. Hence $V_{x',\mathbf{w}^*} = V_{x',\mathbf{w}'}$.

Applying the induction to Y gives $Y_{x',\mathbf{w}^*} = Y_{x',\mathbf{w}'} \neq y$, so C1 holds for $\mathbf{W}^*$. For C2(a), each intermediate satisfies $(P_i)_{x',\mathbf{w}^*} = (P_i)_{x',\mathbf{w}'} \neq (P_i)_{\mathbf{u}}$, and C2(b) is unchanged since it depends only on the model and context. Minimality: let $\mathbf{W}^{**} \subseteq \mathbf{W}^*$ be a minimal subset such that $\{X\} \cup \mathbf{W}^{**}$ still satisfies $Y_{x',\mathbf{w}^{**}} \neq y$. Then $(x', \mathbf{W}^{**}, \mathbf{w}^{**}, \pi)$ satisfies C1–C2, and the witness search can be restricted to eligible subsets of $F^*(\pi)$. □

Proof of Thm. 4.2. We bound the runtime of Alg. 1 when the maximum in-degree is k and the longest directed path in $G(\mathcal{M})$ has length L. Let $n = |\mathbf{V}|$ and $d = \max_V |\mathcal{D}_V|$.

Step 1: Ancestor count. Each variable has at most k parents. The ancestors of Y reachable within L edges satisfy $|\mathrm{Anc}(Y)| \leq k + k^2 + \cdots + k^L = O(k^L)$.

Step 2: Model count. Each C3 evaluation replaces X's equation with a constant in the submodel $\mathcal{M}_{X=x'}$, and only ancestors of Y are candidates for such replacements. Each ancestor can be fixed to one of at most d values or left unfixed,

giving at most $(d{+}1)^{|\mathrm{Anc}(Y)|} = (d{+}1)^{O(k^L)}$ distinct submodels. Memoization on $(\mathcal{M}, X)$ ensures each (model, variable) pair is evaluated at most once. The total number of $(\mathcal{M}, X)$ pairs is at most $n \cdot (d{+}1)^{O(k^L)}$.

Step 3: Path count per model. In a DAG with maximum in-degree k and longest path length L, the number of directed paths from any variable to Y is at most k^L (at each step along the path, there are at most k choices for the preceding parent).

Step 4: Witness enumeration per path. By Prop. 4.1, the witness search for a path π is restricted to eligible variables in $F^*(\pi) = \mathrm{Anc}(\pi) \setminus \pi$. Since $\mathrm{Anc}(\pi) \subseteq \mathrm{Anc}(Y)$, Step 1 gives $|F^*(\pi)| \leq O(k^L)$. Each ancestor-frontier variable is either absent from $\mathbf{W}$ or takes one of d values, giving at most $(d{+}1)^{O(k^L)}$ witness configurations per path. Checking each configuration requires evaluating counterfactuals ($O(n)$) and verifying minimality over $O(kL)$ subsets at $O(d \cdot n)$ each, giving $O(kL \cdot d \cdot n)$ per configuration; this is $O(n)$ when k and L are bounded constants.

Step 5: Cost of C2(b) per path. Each intermediate P_i on a path requires a recursive CAUSES call within the same model. By memoization, each (M, P_i) pair is computed at most once; the cost is absorbed into the per-model budget.

Combined bound. For each of the $n \cdot (d{+}1)^{O(k^L)}$ (model, variable) pairs, the algorithm enumerates at most k^L paths, and for each path evaluates at most $(d{+}1)^{O(k^L)}$ witness configurations at cost $O(n)$ each. Also, for each of at most d alternative values x', the C3 check calls PRODUCES in the submodel $\mathcal{M}_{X=x'}$ for each of at most d values x''. The total runtime is:

$$O\big(n \cdot (d{+}1)^{O(k^L)} \cdot d \cdot k^L \cdot (d{+}1)^{O(k^L)} \cdot n\big) = O\big(n^2 \cdot d^{O(k^L)} \cdot k^L\big).$$

Since k and L are bounded constants, $k^L = O(1)$, giving $O(n^2 \cdot d^{O(k^L)})$, polynomial in both n and d. □

Proof of Prop. B.2. Since $\mathcal{M}'$ is a conservative extension of $\mathcal{M}$, the submodel $\mathcal{M}'_{X=x'}$ is a conservative extension of $\mathcal{M}_{X=x'}$: every counterfactual between variables in $\mathbf{V}(\mathcal{M})$ under intervention $do(X = x')$ is preserved. By hypothesis $X \to Y$ is an edge in $G(\mathcal{M}')$, and fixing $X = x'$ retains outgoing edges from X, so $X \to Y$ is an edge in $G(\mathcal{M}'_{X=x'})$, providing a direct path with no intermediates; C2 holds vacuously for $X = x'$. (Conservative extensions need not preserve edges in general, so this hypothesis is required; it holds throughout Halpern's construction, where $A \to B$ is direct at every stage.) The C1 witness satisfies $\mathbf{W} \subseteq \{V \in \mathbf{V}(\mathcal{M}) \setminus \{X, Y\} : V_{x'} = V(\mathbf{u})\}$. By hypothesis $X = x'$ satisfies C1 for $Y = y$ in $\mathcal{M}_{X=x'}$; this C1 is witnessed entirely by statements over $\mathbf{V}(\mathcal{M})$ (the witness set $\mathbf{W}$, the eligibility conditions, the outcome flip, and minimality over subsets of $\{X\} \cup \mathbf{W}$). Conservative extensions preserve every such counterfactual [Halpern, 2016b, Lemma 4.3], so $X = x'$ satisfies C1 in $\mathcal{M}'_{X=x'}$ as well. Therefore $X = x'$ satisfies C1 and C2 in $(\mathcal{M}'_{X=x'}, \mathbf{u})$, so C3 fails for the foil x'. Since X is binary, x' is the only alternative to x, so no foil makes $X = x$ a cause, and $X = x$ is not a cause of $Y = y$ in $(\mathcal{M}', \mathbf{u})$. □

Corollary B.3 follows immediately. In Halpern's construction, B's equation references A directly at every stage, so $A \to B$ is a direct edge in every $\mathcal{M}_n$, satisfying the edge-persistence hypothesis of Prop. B.2; the extension $\mathcal{M}_2$ introduces a mechanism via Y_1 by which $A = 0$ satisfies C1 in $\mathcal{M}_{A=0}$. Prop. B.2 ensures this persists in all subsequent extensions. HP's alternation exploits the per-witness condition AC2(b), which can be circumvented by new variables; our C3, which independently checks whether the alternative value can produce Y, is not affected by this.

Proof of Prop. 3.6. The switch has graph $S \to C_1$, $S \to C_2$, $C_1 \to Y$, $C_2 \to Y$ with deterministic equations $C_1 = S$, $C_2 = \neg S$, $Y = C_1 \vee C_2$. Actual state: $\vec{v} = (S{=}1, C_1{=}1, C_2{=}0, Y{=}1)$.

The definition requires a *structural simplification* $(\mathcal{M}_2, \mathbf{u}, \vec{v})$ in which $Y = 1$ *counterfactually depends* on $S = 1$ via the diamond modality: there exists $s' \neq 1$ such that $\langle S \leftarrow s' \rangle Y \neq 1$.

Construct $\mathcal{M}_2$ by removing the edge $C_2 \to Y$. The graph condition holds: C_2 has no outgoing edges in $G(\mathcal{M}_2)$, so $(C_2, Y) \notin \mathrm{Anc}(G(\mathcal{M}_2))$. The generalized function gives $f_Y(c_1) = \bigcup_{c_2} \{c_1 \vee c_2\}$: $f_Y(1) = \{1\}$ and $f_Y(0) = \{0, 1\}$. Thus $\mathcal{M}_2$ is a valid structural simplification.

In $(\mathcal{M}_2, \mathbf{u}, \vec{v})$: $S = 1$ and $Y = 1$ (since $f_Y(1) = \{1\}$), satisfying CD1. Under intervention $S \leftarrow 0$: $C_1 = 0$ and $f_Y(0) = \{0, 1\}$. Since $0 \in f_Y(0)$, there exists a nondeterministic resolution with $Y = 0 \neq 1$, so $\langle S \leftarrow 0 \rangle Y \neq 1$ holds, satisfying CD2.

Therefore $S = 1$ is classified as an actual cause of $Y = 1$. □

F EVENT INDIVIDUATION

The main text (Sec. 3.5) argues that fragility is a modeling error, not a definitional one: the SCM framework makes individuation explicit through variable and domain selection, and the implicit contrastive foil of a "why" query determines the appropriate granularity. This appendix addresses related objections from the philosophical literature.

Proportionality. Yablo [1992] argues that causes should be *proportional* to their effects: neither too specific nor too general. A fire is caused by the match being struck, not by the match being struck at angle $37.2°$. In the SCM framework, proportionality is a constraint on the model, not the definition: the domain $\mathcal{D}_X$ of the cause variable determines the level of specificity. If the model includes the striking angle, the definition correctly identifies it as a cause of the fine-grained outcome; if not, it identifies the coarser event. The modeler's domain choice encodes the proportionality judgment.

Model relativity. Halpern [2016a, Chapter 4] emphasizes that all SCM-based definitions are model-relative: causal verdicts can change when variables are added, removed, or re-ranged. Hall [2007] objects that relativizing causation to a model is "plainly silly." We take the opposing view: model relativity is a feature, not a bug. Disagreements about actual causation reduce to disagreements about which model better represents the relevant aspects of the world [Halpern, 2016a, p. 26]. Our definition makes the primary degree of freedom explicit: the choice of explanandum and intermediate variables. Other definitions hide analogous choices in witness selection heuristics or normality orderings [Halpern, 2016a, Chapter 3].

Stability under refinement. Adding intermediate variables can change causal verdicts. In our framework, this is by design: coarse preemption is structurally a switch (Sec. 3.5), and adding intermediates that trace the production pathway recovers the preemption verdict. Halpern [2016b, Theorem 6.1] shows that HP causal verdicts can alternate under conservative extensions; normality assumptions are required to stabilize them. Our definition prevents this alternation without normality (Prop. B.2), but whether verdicts are fully preserved under conservative extensions remains open.

Contrastive causation. Schaffer [2005] proposes that causation is quaternary: "c rather than c^* causes e rather than e^*." This dissolves fragility by specifying the contrast class explicitly. Our framework is compatible: the SCM query "does $X = x$ cause $Y = y$?" implicitly contrasts x with each alternative $x' \in \mathcal{D}_X \setminus \{x\}$, as formalized in C1, and y with the values Y takes under intervention. The domain $\mathcal{D}_Y$ encodes the contrast class on the effect side, and $\mathcal{D}_X$ on the cause side. Schaffer's quaternary framing maps naturally onto the SCM formulation without modification.