Philosophical Studies (2023) 180:3533–3565
https://doi.org/10.1007/s11098-023-02015-4

An event algebra for causal counterfactuals

Tomasz Wysocki[1]

Accepted: 11 July 2023 / Published online: 26 October 2023

Abstract
"If the tower is any taller than 320 ms, it may collapse," Eiffel thinks out loud. Although understanding this counterfactual poses no trouble, the most successful interventionist semantics struggle to model it because the antecedent can come about in infinitely many ways. My aim is to provide a semantics that will make modeling such counterfactuals easy for philosophers, computer scientists, and cognitive scientists who work on causation and causal reasoning. I first propose three desiderata that will guide my theory: it should be general, yet conservative, yet useful. Next, I develop a formalization of events in the form of an algebra. I identify an event with all the ways in which it can be brought about and provide rules for determining the referent of an arbitrary event description. I apply this algebra to counterfactuals expressed using underdeterministic causal models-models that encode non-probabilistic causal indeterminacies. Specifically, I develop semaphore interventions, which represent how the target system may be modified from without in a coordinated fashion. This, in turn, allows me to bring about any event within a single model. Finally, I explain the advantages of this semantics over other interventionist competitors.

Keywords Counterfactuals · Might-counterfactuals · Causation · Underdeterminism · Causal models · Semaphore interventions · Disjunctive events · Event algebra · Interventionist semantics · Strong centering

"If the tower is any taller than 320 ms, it may collapse," Nouguier thinks out loud. Although understanding this counterfactual poses no trouble, the most successful interventionist semantics struggle to model it because the antecedent can come about in infinitely many ways. My aim is to provide a semantics that will make modeling such counterfactuals easy for philosophers, computer scientists, and cognitive scientists who work on causation and causal reasoning.

✉ Tomasz Wysocki
tomwysocki@pitt.edu

1 Georg-Elias-Müller-Institut für Psychologie, Georg-August-Universität Göttingen, Göttingen, Germany

I first propose three desiderata that will guide my theory: it should be general, yet conservative, yet useful (Sect. 1). Next (Sect. 2), I develop a formalization of events in the form of an algebra. I identify an event with all the ways in which it can be brought about and provide rules for determining the referent of an arbitrary event description. I apply this algebra to counterfactuals expressed using underdeterministic causal models (Sect. 3)—models that encode non-probabilistic causal indeterminacies. Specifically, I develop semaphore interventions, which represent how the target system may be modified from without in a coordinated fashion. This, in turn, allows me to bring about any event within a single model. Finally (Sect. 4), I explain the advantages of this semantics over other interventionist competitors.

1 Motivation

The overreaching motivation for my formalism is to provide an interventionist semantics of counterfactuals that will prove useful for philosophers, computer scientists, and cognitive scientists who work on causation. That's why I want the semantics to satisfy three desiderata. *First*, it should handle the broadest class of counterfactuals, including those with disjunctive antecedents that can happen in infinitely many ways. *Second*, it should be conservative, reducing to the standard interventionist semantics (Halpern, 2000; Pearl, 2009, 2011), so that it can be used in current interventionist theories of deterministic (Glymour & Wimberly, 2007; Hitchcock, 2001; Woodward, 2003; Halpern & Pearl, 2005; Weslake, 2015) and underdeterministic (Wysocki ms2) causation. *Third*, causal models used to evaluate counterfactuals should be easy to understand, manipulate, and implement in a machine. They also shouldn't be psychologically implausible, i.e., they should be at least in principle useful for cognitive scientists modeling causal reasoning.

With these desiderata in mind, I develop the theory in two stages, and the devices I introduce in each stage can be used somewhat independently to improve or extend other theories. One device is a set-theoretic formalization of events (Sect. 2), inspired by (and consistent with) Briggs's (2012) and Fine's (2012) theory of truthmakers; the formalization is interventionist-friendly in that it can be used with any type of causal model. The other is the device of semaphore interventions (Sect. 3.4), which can bring about any event, formalized using the algebra, within a single causal model. Before the work begins, though, I'll say more about the desiderata. To do that, I need to introduce—informally for now—some devices I'll formally define later. A causal model is the basic device for representing situations; it's essentially a system of equations that describe counterfactual relationships between atomic events in a situation, where such events are denoted by variable values. An event is conjunctive if it can happen in only one way; it's disjunctive otherwise. If an event can happen in (in)finitely many ways, I'll call it (in)finitely disjunctive. Counterfactuals have events in antecedents and consequents.

As for the first desideratum: why should we care about causal counterfactuals with disjunctive antecedents? Such counterfactuals appear in everyday language and in science, and classic theories of counterfactuals have never eschewed disjunctive antecedents (Lewis, 1973; Stalnaker, 1968); therefore, a theory of causal counterfactuals shouldn't ignore them. Moreover, many psychologists working on causal cognition explore the possibility that the mind represents causal knowledge in structures resembling causal models (e.g., Gopnik et al., 2004; for an overview, see Gerstenberg & Tenenbaum, 2017 or Woodward, 2021). Since new pieces of information are often expressed as counterfactuals with disjunctive antecedents ("trust me: if you come too early *or* too late, they won't let you in"), if psychologists want to model how causal cognition accommodates such information, they will likely need a formalism that allows them to model such counterfactuals. These reasons generalize to infinitely disjunctive antecedents. Any theoretical or practical science that deals with continuous quantities likely uses such counterfactuals. A civil engineer, for instance, may warn that if the live load on the bridge exceeds a certain threshold, it might collapse, where live load is measured in N/kg^2 and therefore is a continuous quantity. Current interventionist approaches are insufficient for analyzing such claims.

Standard accounts of interventionist counterfactuals (Halpern, 2000; Pearl, 2009, 2011) allow only for conjunctive antecedents. More recent proposals—Briggs's, but also Günther's (2017), which deals with disjunctive causes—allow only for finitely disjunctive antecedents. All standard interventionist semantics express events as finite boolean combinations of atomic sentences, which denote atomic events (Sect. 2). Hence, infinitely disjunctive antecedents cannot be referred to. For example, "Alice grows between 5 and 7 feet" (measured as a rational number) would need to be expressed as "Alice grows by 5.1 feet *or* she grows by 5.13 feet *or* ...," making the boolean combination infinite. The event algebra solves this problem without resorting to an infinitary object language.

Now, as I said, the algebra and the account of interventions are somewhat independent, and in principle you could apply the algebra to derive a set-theoretic representation of an event and, subsequently, use this representation to evaluate the target counterfactual with Briggs's (2012) semantics (Sect. 3.4); such an extension would satisfy the second desideratum, since Briggs's semantics is conservative. However, the extension, like Briggs's original, wouldn't satisfy the third desideratum. To evaluate a counterfactual, the theory requires building a separate model for each way in which the antecedent can be brought about. But such families of models are hard to understand—a single situation is represented by many models at once instead of a single model, and the relationships (e.g., dependence) between events from different models become hard to read off. Manipulating and implementing such families of models in a machine is also hard—or straightforwardly impossible, if the antecedent is infinitely disjunctive. It's also implausible that to process the counterfactual "had you come anytime between five and six o'clock, you would have met her," the listener implicitly constructs an infinity of causal models (or even just 60, if time is measured by the minute). If using multiple models concurrently could be avoided, the tool would be more useful to cognitive scientists working on causal reasoning.

There are other interventionist theories of counterfactuals that allow for disjunctive antecedents (Huber, 2013; Vandenburgh, 2022; Hiddleston, 2005). Later (Sect. 4), I will discuss how they differ from my theory; for now, let me just say that they deliver certain undesirable results as well as violate the second and third desideratum:

they aren't conservative and can't be easily implemented on a machine or utilized by cognitive science.

Now, a theory of causal counterfactuals with disjunctive antecedents commits me to positing disjunctive events. Some don't like such events in the context of causation. For instance, even though his semantics allows for disjunctive antecedents, Lewis banishes disjunctive causes by definition: "Fred talks, and his talking causes Ted to laugh. Suppose that besides Fred's talking there is another event, the disjunctive event of Fred's talking-or-walking. Without it, Fred's talking would not have occurred, and neither would Ted's laughing. So this disjunctive event also causes Ted to laugh. That is intuitively wrong. No such event causes Ted's laughing, or anything else [...] because there is no such event. Hence disjunctive events are to be rejected" (Lewis, 1986:266–7). Analogously, Halpern and Pearl (2005:853) disregard disjunctive causes: a disjunctive event causes an effect only in virtue of some of its disjuncts causing this effect (e.g., Fred's talking in his talking-or-walking), and so entertaining disjunctive events is redundant. Philosophical theories aside, disjunctive events also seem to resist imagination. While it's easy to imagine the event of my stooping *and* your stooping, it's not so easy to imagine the event of my stooping *or* your stooping. If there are no disjunctive events, or they can be safely ignored in the context of causation and causal counterfactuals, then my project is unmotivated.

I don't think disjunctive events can be ignored, however. First, some posit genuinely disjunctive causes. Consider a story by Sartorio (2006), later represented by Günther (2017) in the language of causal models. There is a train coming down the tracks, which diverge and converge before a place where a hostage is tied. The right track is disconnected. If Flipper throws a switch, the train will continue on the left track and run over the hostage. But if Flipper does not throw the switch, Reconnecter will reconnect the right track, and the train will still run over the hostage. Flipper throws the switch; the train runs over the hostage. Sartorio and Günther claim that the disjunctive event "Flipper throws the switch or Reconnecter unblocks the right track" is a cause of the hostage's fate, but the disjuncts "Flipper throws the switch" and "Reconnecter reconnects the right track" are not. If this is right, Lewis and Halpern and Pearl are wrong, and we need to take disjunctive causes seriously.

Furthermore, one of the typical benefits of a formalism is that there is a lot of leeway in interpreting it, and the current one is no exception. Even if your ontology doesn't admit disjunctive events, you still can use the event algebra and the subsequent amendments to the interventionist framework as rules for evaluating such counterfactuals. The rules needn't carry metaphysical implications. Yet, I think that the interventionist framework does in fact yield an intuitive picture of disjunctive events as I use them here. On any interventionist theory, evaluating a counterfactual requires entertaining a situation produced by bringing about the antecedent from without, and from the interventionist perspective, facilitating such interventions is the main role of an account of events. Whence a slogan, imprecise but easy to remember: *the meaning of an event is the method of bringing it about*. According to this interpretation, the event of my stooping *or* your stooping can be brought about by making me stoop, making you stoop, or making us both stoop, and there's nothing more to it.

2 The event algebra

Quite some time ago, Ellis et al. (1977) raised a problem for any semantics of counterfactuals with disjunctive antecedents. You don't want it to treat counterfactuals with logically equivalent antecedents as logically equivalent, for if $p \mathrel{\Box\!\!\rightarrow} q$ is equivalent to $(p \wedge r \vee p \wedge \neg r) \mathrel{\Box\!\!\rightarrow} q$, then (under the principle of the simplification of disjunctive antecedents, which you don't want to mess with) $p \mathrel{\Box\!\!\rightarrow} q$ entails $(p \wedge r) \mathrel{\Box\!\!\rightarrow} q$, and that cannot be. "Had I stooped, the king would have been pleased" shouldn't entail "had I stooped and insulted the king, he would have been pleased." One possible solution to this problem is Fine's (2012) semantics of truthmakers, on which a proposition is associated with a set of truthmakers—possible states that verify it. For instance, "I am stooping" is verified by (the state of) my stooping but not by my stooping and insulting the king, which is why you cannot replace the former with the latter in the antecedent of a counterfactual; the crisis pointed out by Ellis et al. can be pronounced averted. Fine proposed rules for deriving the truthmakers for any proposition, and Briggs (2012) adapted these rules to causal models (explained in Sect. 3.4).[1] For the event algebra, I also have adopted Fine's rules. However, unlike Briggs's adaptation, mine uses more complex algebraic means that can handle infinitely disjunctive events. I also dropped the language of truthmakers. I don't speak of propositions and what makes them true but—which is more appropriate in the interventionist context—of events and the ways they can happen.

A causal model of any kind (deterministic, probabilistic, underdeterministic) comes with variables $\vec{\mathcal{V}}$ and their ranges $\mathcal{R} = \{\mathcal{R}_X\}_{X \in \vec{\mathcal{V}}}$. Values of variables correspond to atomic events. Capital letters denote variables; a capital letter topped with an arrow denotes a set of variables. Variables from $\vec{\mathcal{V}}$ and all its subsets are always ordered in a specific way, called a topological order (to be explained in Sect. 3). There are quite a few rather uncontroversial constraints on the choice of variables and their values Woodward (2016). For instance, variable values should represent events that we are willing to take seriously (e.g., your population-genetics model shouldn't account for the possibility of the organisms evolving bilocation), and an intervention on one variable shouldn't (directly) change the value of other variables (e.g., you don't want a model with both weight and BMI as variables, for intervening on the former would directly change the latter). The second constraint is reminiscent of (though not fully equivalent to) another one, increasingly popular in the literature: if two atomic events could be brought about together, they should be denoted by values of different variables; otherwise, they should belong to one variable (Gallow, ms.; Ross & Woodward, 2022; Wysocki, 2023a). Here's an example that I'll use throughout.

A ship traversing the strait could be either swallowed by Charybdis or sunk by Scylla hurling rocks, but the rest Homer got wrong—Odysseus dies. Exactly when Scylla's rock crushed Odysseus' ship, two of Charybdis' heads maul the deck. I'll represent these events with five variables. $S = 1$ if Scylla notices the ship, $S = 0$ if she doesn't; therefore, S's range is $\mathcal{R}_S = \{0, 1\}$. $C = 1$ if Charybdis notices the ship, $C = 0$ if she doesn't; $\mathcal{R}_C = \{0, 1\}$. R denotes how many rocks Scylla throws

[1] Fine's semantics also requires an additional set of situations that falsify a proposition. In the interventionist context, they aren't relevant, and both Briggs's and my theory do without them.

at the ship, $\mathcal{R}_R = \{0, 1, 2\}$. H denotes how many of Charybdis's heads attack the ship, $\mathcal{R}_H = \{0, 1, 2, 3\}$. $D = 1$ if Odysseus dies, $D = 0$ if he survives; $\mathcal{R}_D = \{0, 1\}$. The atomic events of Charybdis noticing the ship and of Odysseus surviving belong to different variables because you can imagine bringing them about together. Because you can't make Scylla throw no and two rocks in the same situation, the events belong to the same variable.

Crucial for my formalism is the notion of an *assignment over* variables $\vec{X}$ ($\vec{X} \subseteq \vec{\mathcal{V}}$), which is a mapping from $\vec{X}$ to their ranges. An arbitrary assignment is denoted by a lowercase letter topped with an arrow, e.g., $\vec{x}$. A concrete assignment is represented by a tuple of values in the topological order, where target variables are explicitly indicated, e.g., assignment $\langle 2_H, 0_R \rangle$ assigns 2 to H and 0 to R. $\vec{x}[\vec{Y}]$ denotes a projection of $\vec{x}$ onto variables $\vec{Y}$, where $\vec{x}$ must assign values to all variables from $\vec{Y}$, e.g., $\langle 1_C, 2_H, 0_R \rangle [C, H] = \langle 1_C, 2_H \rangle$.[2]

One of the main roles of assignments is to represent the ways in which an event can happen, which means that events are sets of assignments. Therefore, although *atomic events* correspond to variable values, formally they are singletons of assignments that assign a value to a single variable, e.g., $\{\langle 1_C \rangle\}$ is the event of Charybdis noticing the ship. *Conjunctive events* are singletons of assignments, e.g., $\{\langle 1_S, 0_D \rangle\}$ is the event of Scylla noticing Odysseus and him surviving; therefore, atomic events count as conjunctive. Genuinely *disjunctive events* are sets of at least two assignments, e.g., $\{\langle 1_H \rangle, \langle 2_H \rangle\}$ is the disjunctive event of one or two heads attacking the ship. There's also the *empty event*, $\{\langle\rangle\}$, which is denoted by $\top$ and counts as conjunctive, and the contradictory event, $\emptyset$, which is denoted by $\bot$ and can't happen or be brought about in any way.

Not any set of assignments corresponds to an event—all and only assignment sets that have names do. An *event name over* $\vec{X}$ (or a *name* in short) is a boolean combination of statements over variables $\vec{X}$. Specifically, a *brick sentence over* $\vec{X}$ is a statement over $\vec{X}$ that cannot be parsed as a boolean combination of shorter statements, and for any assignment over $\vec{X}$, the sentence has a determinate truth-value when the values from the assignment are plugged into the sentence.[3] Any name is either a brick sentence or a boolean combination of such sentences. The syntax of brick sentences depends on the operations afforded by ranges. In almost all cases used in the literature, ranges are subsets of real or natural numbers; in such cases, brick sentences are simply arithmetic equations or inequalities over some variables from $\vec{\mathcal{V}}$. So, $H+R<4$ is a brick sentence, while $C = 1 \wedge R \geq 1$ is a name but not a brick sentence. Brick sentences play the *syntactic* role of atomic sentences, out of which all other sentences are built, but they don't play the *semantic* role, for they needn't denote atomic events (which is why I don't call them atomic). For example, $H+R < 4$ and $C = 8$ are brick sentences, yet they don't denote atomic events. And conversely, non-brick names also can denote atomic events, e.g., $R \geq 1 \wedge R \leq 1$.

To anyone accustomed to syntax paralleling semantics in logical systems, brick sentences will appear weird; yet, play a critical role in the algebra. If I wanted to build

[2] If $\vec{Y}$ is a single variable, Y, I'll abuse the notation and typically let $\vec{x}[Y]$ denote Y's value instead of the assignment over one value, e.g., $\langle 1_C, 2_H, 0_R \rangle [H] = 2$ instead of $\langle 1_C, 2_H, 0_R \rangle [H] = \langle 2_H \rangle$. A projection onto an empty list of variables produces the empty assignment, $\vec{x}[\emptyset] = \langle\rangle$.

[3] This requirement rules out expressions where, for example, you divide by 0.

names of disjunctive events out of names of atomic events in the orthodox fashion—atomic events are denoted by atomic names in the form of $X = x$, complex events are denoted by boolean combinations of atomic names—I would run into trouble trying to name infinitely disjunctive events. The event of Alice drinking between 1 and 3 potions poses no problems, as it's denoted by $P = 1 \vee P = 2 \vee P = 3$, where P stands for the number of potions drunk. The event of Alice growing between 5 and 7 feet can't be treated the same way, for there would have to be infinitely many disjuncts—$\aleph_0$ if the range is rational and $\mathfrak{c}$ if real. One way around it is to use an object language that allows for infinite boolean combinations. Yet, such a language employs an extremely elaborate formal apparatus (Bell, 2016), which would make the entire theory utterly impractical, violating the third desideratum and (abstracting from any desiderata) rendering the theory largely inaccessible. Whence brick sentences, which demand only primary-school syntactic means.

For any name φ, $[\![\varphi]\!]$ is the unique event denoted by this name, and the event algebra provides rules for identifying this event. Because only assignment sets that have names constitute events, the algebra also carves out events from the family of all possible assignment sets.[4] To introduce the algebra, I first need the operation of *spatial concatenation*, which I denote by $\star$. Spatial concatenation combines two assignments into one, which agrees with either assignment on all values: $\left(\vec{u}_{\vec{X}} \star \vec{w}_{\vec{Y}}\right)[\vec{X}] = \vec{u}_{\vec{X}}$, $\left(\vec{u}_{\vec{X}} \star \vec{w}_{\vec{Y}}\right)[\vec{Y}] = \vec{w}_{\vec{Y}}$, where $\vec{u}_{\vec{X}}$ is over $\vec{X}$, $\vec{w}_{\vec{Y}}$ is over $\vec{Y}$, and $\vec{u}_{\vec{X}} \star \vec{w}_{\vec{Y}}$ is over $\vec{X} \cup \vec{Y}$. The operation is defined only for assignments that agree on the values of shared variables. So, $\langle 3_H, 2_R \rangle \star \langle 2_R, 0_D \rangle = \langle 3_H, 2_R, 0_D \rangle$, $\langle 3_H, 2_R \rangle \star \langle 2_R \rangle = \langle 3_H, 2_R \rangle$, and $\langle 3_H, 2_R \rangle \star \langle 3_R, 0_D \rangle$ is undefined. Spatial concatenation is associative and commutative, and the empty assignment $\langle\rangle$ is its one (the neutral element).

With spatial concatenation, I can define *junction* (the verb: to *join*)— an operation that takes as arguments two assignment sets and returns an assignment set that contains all spatial combinations (whenever they are defined) of the assignments from the first set with the assignments from the second set:

$$U \bowtie W = \bigcup_{\vec{u} \in U} \bigcup_{\vec{w} \in W} \{\vec{u} \star \vec{w}\}. \tag{1}$$

For instance, $\{\langle 3_H, 2_R \rangle\} \bowtie \{\langle 2_R, 0_D \rangle, \langle 2_R \rangle, \langle 3_R, 0_D \rangle\} = \{\langle 3_H, 2_R, 0_D \rangle, \langle 3_H, 2_R \rangle\}$. Junction is associative and commutative, $[\![\top]\!] = \{\langle\rangle\}$ is its one, and $[\![\bot]\!] = \emptyset$ is its zero (the annihilating element). Both $\top$ and $\bot$ are names over no variables.

I can lay out the algebra now. For any brick sentence $\beta_{\vec{X}}$ that is over variables $\vec{X}$, and any names φ and ψ,

$$[\![\top]\!] = \{\langle\rangle\} \tag{2}$$

$$[\![\bot]\!] = \emptyset \tag{3}$$

$$[\![\beta_{\vec{X}}]\!] = \left\{\vec{x} : \ \vec{x} \text{ satisfies } \beta_{\vec{X}}\right\} \quad \text{where } \vec{x} \text{ is over } \vec{X} \tag{4}$$

$$[\![\neg\beta_{\vec{X}}]\!] = \left\{\vec{x} : \ \vec{x} \text{ doesn't satisfy } \beta_{\vec{X}}\right\} \quad \text{where } \vec{x} \text{ is over } \vec{X} \tag{5}$$

[4] Events are similar to teams (Barbero & Sandu, 2021) in that they are sets of assignments; however, unlike teams, events can include assignments that aren't over the same variables.

$$[\![\neg\neg\varphi]\!] = [\![\varphi]\!] \tag{6}$$

$$[\![\varphi \wedge \psi]\!] = [\![\varphi]\!] \bowtie [\![\psi]\!] \tag{7}$$

$$[\![\neg(\varphi \wedge \psi)]\!] = [\![\neg\varphi \vee \neg\psi]\!] = [\![\neg\varphi]\!] \cup [\![\neg\psi]\!] \cup [\![\neg\varphi]\!] \bowtie [\![\neg\psi]\!] \tag{8}$$

$$[\![\varphi \vee \psi]\!] = [\![\varphi]\!] \cup [\![\psi]\!] \cup [\![\varphi \wedge \psi]\!] = [\![\varphi]\!] \cup [\![\psi]\!] \cup [\![\varphi]\!] \bowtie [\![\psi]\!] \tag{9}$$

$$[\![\neg(\varphi \vee \psi)]\!] = [\![\neg\varphi \wedge \neg\psi]\!] = [\![\neg\varphi]\!] \bowtie [\![\neg\psi]\!] \tag{10}$$

$$[\![\varphi \rightarrow \psi]\!] = [\![\neg\varphi \vee \psi]\!] = [\![\neg\varphi]\!] \cup [\![\psi]\!] \cup [\![\neg\varphi]\!] \bowtie [\![\psi]\!] \tag{11}$$

$$[\![\neg(\varphi \rightarrow \psi)]\!] = [\![\varphi \wedge \neg\psi]\!] = [\![\varphi]\!] \bowtie [\![\neg\psi]\!] \tag{12}$$

$$[\![\varphi \equiv \psi]\!] = [\![\varphi \wedge \psi]\!] \cup [\![\neg\varphi \wedge \neg\psi]\!] = [\![\varphi]\!] \bowtie [\![\psi]\!] \cup [\![\neg\varphi]\!] \bowtie [\![\neg\psi]\!] \tag{13}$$

$$[\![\neg(\varphi \equiv \psi)]\!] = [\![\varphi \wedge \neg\psi]\!] \cup [\![\neg\varphi \wedge \psi]\!] = [\![\varphi]\!] \bowtie [\![\neg\psi]\!] \cup [\![\neg\varphi]\!] \bowtie [\![\psi]\!] \tag{14}$$

Rules (7)-(14) parse any name with a binary connective as either a junction or a union of events. The event denoted by a conjunction is the junction of the events denoted by the conjuncts (7, 10, 12); it is so because any way in which one *and* another event can happen at once is a combination (read: spatial concatenation) of some ways in which either event can happen. The event denoted by a disjunction is the union of three events: the two events denoted by the disjuncts and the event denoted by the conjunction of the disjuncts (8, 9, 11, 13, 14); it is so because any way in which one *or* another event can happen is a way in which the first event, or the second event, or both at once can happen. Second, the rules push negation down until two negations cancel each other out (6) or until it reaches a name without any logical connectives (5). Negation is normally handled by the set complement.[5] However, there's no sensible interpretation of the complement of an event because events can contain assignments over different values.

The rules work as expected. The disjunctive event "Scylla fails to notice Odysseus or Charybdis fails" is parsed as

$$\begin{aligned}[\![S = 0 \vee C = 0]\!] &= [\![S = 0]\!] \cup [\![C = 0]\!] \cup [\![S = 0]\!] \bowtie [\![C = 0]\!] \\ &= \{\langle 0_S \rangle\} \cup \{\langle 0_C \rangle\} \cup \{\langle 0_S \rangle\} \bowtie \{\langle 0_C \rangle\}\end{aligned}$$

[5] E.g., the set of worlds where $\neg p$ holds is the complement of the set of worlds where p holds.

$$= \{\langle 0_S \rangle\} \cup \{\langle 0_C \rangle\} \cup \{\langle 0_S, 0_C \rangle\} = \{\langle 0_S \rangle, \langle 0_C \rangle, \langle 0_S, 0_C \rangle\}.$$

This is correct, for the event can happen in three ways: Scylla fails, Charybdis fails, or they both fail. The contradictory event "Scylla throws no rocks and two rocks" is parsed as

$$[\![R = 0 \wedge R = 2]\!] = [\![R = 0]\!] \bowtie [\![R = 2]\!] = \{\langle 0_R \rangle\} \bowtie \{\langle 2_R \rangle\} = \emptyset.$$

Scylla throws no rocks or two rocks: $[\![R = 0 \vee R = 2]\!] = \{0_R, 2_R\}$. As many rocks as heads attack the ship: $[\![R = H]\!] = \{\langle 0_R, 0_H \rangle, \langle 1_R, 1_H \rangle, \langle 2_R, 2_H \rangle\}$. As many rocks as heads attack the ship, and only one head attacks it: $[\![R = H \wedge H = 1]\!] = \{\langle 0_R, 0_H \rangle, \langle 1_R, 1_H \rangle, \langle 2_R, 2_H \rangle\} \bowtie \{\langle 1_H \rangle\} = \{\langle 1_R, 1_H \rangle\}$, exactly one rock and one head attack the ship. It's not the case that Scylla hurls two rocks or Odysseus dies: $[\![\neg(R = 2 \vee D = 1)]\!] = [\![\neg R = 2]\!] \bowtie [\![\neg D = 1]\!] = \{\langle 0_R, 0_D \rangle, \langle 1_R, 0_D \rangle\}$; she throws no rocks and he survives or she throws one rock and he survives.

As advertised, the algebra becomes especially useful when ranges are infinite. Let $T = t$ stand for the event of the tower being t meters tall, and $\mathcal{R}_T = [0, 500]$. "The tower is taller than two hundred but not taller than three hundred meters" is parsed as

$$\begin{aligned} &[\![T{>}200 \wedge \neg T{>}300]\!] = [\![T{>}200]\!] \bowtie [\![\neg T{>}300]\!] \\ &\quad = (200, 500]_T \bowtie \overline{(300, 500]}_T = (200, 500]_T \bowtie (0, 300]_T = (200, 300]_T. \end{aligned}$$

Mass M or mass N are more than 1 kg: where $\mathcal{R}_M = \mathcal{R}_N = [0, +\infty)$, $[\![M{>}1 \vee N{>}1]\!] = (1, +\infty)_M \cup (1, +\infty)_N \cup (1, +\infty)_M \bowtie (1, +\infty)_N$.[6]

Every non-contradictory event φ over at least one variable can be expressed in the *canonical form*—as a junction of the maximum number of juncts J_i over non-overlapping variables

$$[\![\varphi]\!] = \bowtie_{i=1}^{n} J_i, \tag{15}$$

where J_i are events themselves, assignments from J_i don't share variables with assignments from J_j $(i \neq j)$, and $J_i \neq \{\langle\rangle\}$. An event can have only one canonical form (Sect. 6.2). For instance, the event of Scylla throwing some rocks and Charybdis attacking with some heads unfolds as $[\![R{\geq}1 \wedge H{\geq}1]\!] = \{\langle 1_R, 1_H \rangle, \langle 1_R, 2_H \rangle, \langle 1_R, 3_H \rangle, \ldots, \langle 2_R, 3_H \rangle\}$, but its canonical form is $\{\langle 1_R \rangle, \langle 2_R \rangle\} \bowtie \{\langle 1_H \rangle, \langle 2_H \rangle, \langle 3_H \rangle\}$. The canonical form, and the fact that it's unique, will prove useful for devising interventions (Sect. 3.4).

3 Underdeterministic counterfactuals

The main motivation behind the event algebra is modeling counterfactuals with disjunctive antecedents. I'll show how that works with underdeterministic counterfactuals (Wysocki, 2023b) and leave probabilistic counterfactuals for the future. Although the

[6] In both cases, construe the sets as sets of assignments over single variables. I.e., $(200, 300]_T = \{\langle t \rangle : 200 < t \leq 300\}$.

latter are much more familiar than the former, I'll focus on underdeterministic counterfactuals for two reasons: the mechanics is simpler than in the case of probabilistic models, and deterministic would-counterfactuals—the staple of causal analysis of counterfactuals—are edge cases of, and therefore easily handled as, underdeterministic counterfactuals.

3.1 Motivation

Since underdeterministic counterfactuals are a recent addition to the causal literature, I'll briefly motivate them.[7] Say, the following three might-counterfactuals are true: if Scylla hurls the rock, the ship may sink; if she hurls the rock, the ship may stay afloat; if she doesn't hurl her rock, the ship will stay afloat. She hurls her rock, the ship sinks, and the former is a cause of the latter—but for the rock, the ship wouldn't have sunk. I call such causes underdeterministic, for the causal claim holds even though the situation is neither deterministic nor expressed in terms of probabilities.

Underdeterministic token causes belong to a broader family of underdeterministic concepts that include, among others, conditional independence, causal counterfactuals (which I focus on below), type causes, and decision theory. Why should we care about investigating these concepts? First, there's a simple argument: discovering the nature of causation is a worthwhile philosophical endeavor; underdeterministic causes are a species of cause; therefore, discovering the nature of underdeterministic causes is a worthwhile philosophical endeavor. Moreover, since underdeterministic relations are the weakest relations that deserve to be called causal, investigating them will tell us something about the necessary conditions of causation. Second, there are scientific theories, such as the theory of eternal inflation, where non-determinism is essentially non-probabilistic (Norton, 2003; 2021). Many equilibrium models allow for multiple equilibria given boundary conditions but don't specify how likely these equilibria are; such models are genuinely underdeterministic (Spence, 1973). Behavior of agents under bounded uncertainty counts as underdeterministic, as they know the possible effects of their actions but not these effects' probabilities (Deœux, 2019; Knight, 1921). Finally, it seems that we sometimes reason with underdeterministic causes ("I better take an umbrella: if I don't take one, I may get wet because it may rain"), which makes causal underdeterminism interesting also for cognitive science.

3.2 Deterministic and underdeterministic models

The basic notion that allows for analyzing underdeterministic causal phenomena is that of an underdeterministic causal model; I'll introduce it starting with its familiar counterpart, the deterministic model.

Any causal model includes variables and their ranges, which I have already discussed. A (deterministic or underdeterministic) model also includes structural equations, which encode direct causal relationships between atomic events represented by

[7] For an extensive motivation, see (Wysocki, 2023b).

Table 1 The solution to the deterministic model

	S	C	R	H	D
$\vec{\sigma}$	1	1	1	2	1

variable values. If a variable is an argument in another's equation, the former is the latter's parent, and the latter is the former's child. A variable's ancestor is either a parent or an ancestor of a parent of the variable. Like most authors, I'll limit myself to acyclic models—ones where no variable is its own ancestor. In deterministic models, equations encode primitive would-counterfactuals, which state what value a variable would have taken on for any combination of values of its parents. A variable is endogenous if it has parents, and exogenous otherwise.

Recall the plight of Odysseus. Scylla will notice him, and so will Charybdis. If Scylla notices him, she'll throw one rock, and she'll stay put otherwise. If Charybdis notices him, two of her heads will attack, and none otherwise. He'll die if he's attacked by at least three rocks or heads combined.[8] These are the primitive counterfactuals, and they are encoded by:

$$S \leftarrow 1, \quad C \leftarrow 1, \quad R \leftarrow S, \quad H \leftarrow 2C, \quad D \leftarrow R{+}H \geq 3. \tag{16}$$

Odysseus' world is deterministic, and his fate is sealed—given that the monsters will notice him, Scylla will throw a rock, Charybdis will attack with two heads, and he'll die. How the world will unfold is represented by the solution to the equations, and deterministic acyclic models have exactly one solution (tab. 1):

A solution is therefore an assignment over all variables that satisfies the equations. It corresponds to the conjunctive event over all variables that happens in the situation described—the possible history. Hence, deterministic models always tell a single possible history.

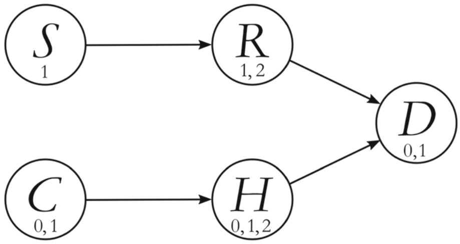
According to the standard semantics of deterministic would-counterfactuals with conjunctive antecedents, to evaluate "had $\vec{X}{=}\vec{x}$ happened, φ would have happened," you bring about $\vec{X}{=}\vec{x}$ with an intervention, replacing every X's equation with one that sets X's value to its corresponding value from $\vec{x}$: $X \leftarrow \vec{x}[X]$ for every X from $\vec{X}$. The would-counterfactual holds *iff* φ is satisfied by the solution to the post-intervention model. For instance, it's true (in the model) that Odysseus would survive if Scylla were not to notice the ship and Charybdis were to stay put. Bringing about the antecedent event $S = 0 \wedge H = 0$ produces a new set of equations,

$$S \leftarrow 0, \quad C \leftarrow 1, \quad R \leftarrow S, \quad H \leftarrow 0, \quad D \leftarrow R{+}H \geq 3,$$

on whose sole solution, $\langle 0_S, 1_C, 0_R, 0_H, 0_D \rangle$, Odysseus survives ($D = 0$). Notice that the interventions bringing about conjunctive events always make the target variables exogenous; that sometimes won't be the case with bringing about disjunctive events.

[8] See footnote 9 for why I use the future tense.

Fig. 1 The strait case



Representing underdeterministic situations requires just one change: structural equations can now return multiple values and thus encode primitive might-counterfactuals. Say, Odysseus' fate is more nuanced: Scylla will notice the ship, but Charybdis only may. If Scylla notices the ship, she may hurl one rock or she may hurl two; if she doesn't notice the ship, she may stay put, but she also may spontaneously throw one rock. If Charybdis doesn't notice the ship, no heads will attack the ship; if she notices, one or two heads may attack the vessel. Again, Odysseus will die if attacked by at least three rocks or heads combined, and won't otherwise.[9]

These primitive would- and might-counterfactuals are encoded by (Fig. 1):

$$S \leftarrow 1, \quad C \leftarrow 0, 1, \quad R \leftarrow S,\ S{+}1, \quad H \leftarrow C,\ 2C, \quad D \leftarrow R{+}H \geq 3. \tag{17}$$

S's equation, as before, is exogenous and deterministic. C's, exogenous and underdeterministic, states that Charybdis may ($C = 1$) or may not ($C = 0$) notice Odysseus. The equation constrains C's values to 0 and 1 but doesn't determine between them.[10] R's, endogenous and underdeterministic, states that Scylla may hurl one ($R = 1$) or two ($R = 2$) rocks if she notices Odysseus ($S = 1$), and none ($R = 0$) or one ($R = 1$) if she doesn't ($S = 0$).[11] H's equation determines H's value to be 0 if $C = 0$—no heads will attack if Charybdis doesn't notice the ship—and constrains H's values to 1 and 2 if $C = 1$—one or two heads may attack if she sees the vessel. D's equation remains unchanged.

Since every equation is a correspondence, which always returns at least one value, an underdeterministic model always tells at least one history.[12] The strait model tells six (tab. 2):

In the first history, Scylla notices the ship but Charybdis doesn't; having noticed it, Scylla hurls one rock; not having noticed the ship, Charybdis attacks with no heads; Odysseus survives. In the last history, both monsters notice the ship, a rock is thrown,

[9] In describing non-deterministic cases, I prefer to use the future tense, which makes it easier to think about the future as open. Once you know that Charybdis noticed the ship, you're less likely to find not noticing the ship a genuine possibility. But the account doesn't distinguish between counterfactuals and future conditionals.

[10] Just read it as $C \in \{0, 1\}$.

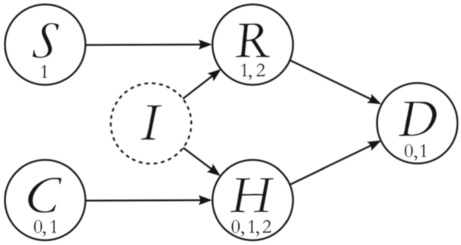
[11] Again, just read it as $R \in \{S, S + 1\}$, where the set depends on S's value.

[12] Formally, I'll construe structural equations as functions returning non-empty subsets of the variable's range, $f_X : \mathcal{R}_{P_1} \bowtie \cdots \bowtie \mathcal{R}_{P_n} \rightarrow 2^{\mathcal{R}_X}$. Therefore, a model is essentially a system of constraints in the form $X \in f_X(\vec{P})$. An equation is deterministic if it returns a singleton for every input.

Table 2 Solutions to the strait model

	S	C	R	H	D		S	C	R	H	D
$\vec{\sigma}_1$	1	0	1	0	0	$\vec{\sigma}_4$	1	1	2	1	1
$\vec{\sigma}_2$	1	0	2	0	0	$\vec{\sigma}_5$	1	1	1	2	1
$\vec{\sigma}_3$	1	1	1	1	0	$\vec{\sigma}_6$	1	1	2	2	1

Fig. 2 The diagram for $\mathfrak{M}_{R=1 \vee H=2}$



two heads attack, and Odysseus dies. And so on; all and only the six assignments satisfy the model's equations.

I have so far avoided defining underdeterministic models formally—it's high time. A model is a quadruple $\mathfrak{M} = \left\langle \vec{\mathcal{V}}, \vec{\mathcal{I}}, \mathcal{R}, \mathcal{E} \right\rangle$. It contains two kinds of *nodes*: *variables*, $\vec{\mathcal{V}}$, and *intervention vertices*, $\vec{\mathcal{I}}$; $\vec{\mathcal{I}}$ is empty before any interventions are executed. $\mathcal{R}$ contains every node's range, and $\mathcal{E}$ contains every node's structural equation. Informally, you can think of variable values as denoting what happens in the situation, and of vertex values as denoting what may be done to the situation from without; I'll further develop this explanation once I get to interventions. A *solution* to a model is an assignment over all nodes that satisfies the model's equation. A *(possible) history* told by a model is an assignment over all variables, which is the projection of some solution onto variables alone—that is, histories are solutions oblivious to what has been done to the situation from the outside (Fig. 1).

Underdeterministic models can be helpfully represented with diagrams. Nodes (i.e., variables and intervention vertices together) correspond to nodes in the diagram. Directed edges go from a variable to all its children. A node in the diagram contains the values the node takes on in some solution.[13] Nodes, $\vec{\mathcal{I}} \cup \vec{\mathcal{V}}$, are ordered topologically, which means that every node comes before its every child.

3.3 Causal modalities and conditional dependence

Underdeterministic situations may develop in multiple ways, which suggests a natural semantics of causal-modal claims (Wysocki, 2023b). Event ψ is *causally possible* in model $\mathfrak{M}$,

$$\mathfrak{M} \models \Diamond\, \psi \quad \textit{iff} \quad \psi \text{ happens on some history told by } \mathfrak{M}. \tag{18}$$

[13] Unfortunately, this convention won't do justice to the entire model, as the diagram doesn't show which values from different nodes belong to the same solution. I also won't list the values of intervention vertices on the diagram.

For example, it's causally possible that the ship is attacked by more heads than rocks because $H > R$ on $\vec{\sigma}_5$.

Event ψ is *causally necessary* in model $\mathfrak{M}$,

$$\mathfrak{M} \models \Box\, \psi \quad \textit{iff} \quad \psi \text{ happens on all histories told by } \mathfrak{M}. \tag{19}$$

Scylla's noticing Odysseus is causally necessary because $S = 1$ on all solutions. The operators are duals,

$$\mathfrak{M} \not\models \Diamond\, \psi \quad \textit{iff} \quad \mathfrak{M} \models \Box\, \neg\psi \tag{20}$$

and necessity implies possibility. In deterministic models, all events that will happen will happen necessarily, and the two modal notions collapse into one.[14]

Causal possibility underlies the notion of independence.[15] Where $\vec{X}$ and $\vec{Y}$ don't overlap with each other but are allowed to overlap with $\vec{Z}$, $\vec{X}$ and $\vec{Y}$ are *conditionally independent* given $\vec{Z}$ in $\mathfrak{M}$ *iff* any possible conjunctive events over $\vec{X} \cup \vec{Z}$ and over $\vec{Y} \cup \vec{Z}$ that agree on $\vec{Z}$ are co-possible,

$$\begin{aligned} \vec{X} \perp \vec{Y} \mid \vec{Z} \qquad \textit{iff} \qquad & \mathfrak{M} \models \Diamond \left(\vec{X}{=}\vec{x} \wedge \vec{Z} = \vec{z}\right) \text{ and } \mathfrak{M} \models \Diamond \left(\vec{Y}{=}\vec{y} \wedge \vec{Z} = \vec{z}\right) \\ & \text{implies} \quad \mathfrak{M} \models \Diamond \left(\vec{X}{=}\vec{x} \wedge \vec{Y}{=}\vec{y} \wedge \vec{Z} = \vec{z}\right) \qquad \text{for any } \vec{x}, \vec{y}, \vec{z}. \end{aligned} \tag{21}$$

$\vec{Z} = \emptyset$ yields the definition of unconditional independence: any two possible conjunctive events over $\vec{X}$ and over $\vec{Y}$ are possible together. For instance, R and H are independent, $R \perp H \mid \emptyset$, because $R = 1$ and $R = 2$ are possible, $H = 0$, $H = 1$, and $H = 2$ are possible, and any conjunction of the former and the latter events is also possible. However, once you conditionalize on D, the variables become dependent, $R \not\perp H \mid D$, e.g., because $R = 2 \wedge D = 0$ and $H = 1 \wedge D = 0$ are possible, but $R = 2 \wedge H = 1 \wedge D = 0$ isn't. If $\vec{X}$ and $\vec{Y}$ don't share ancestors, they are independent.[16]

3.4 Semaphore interventions

Finally, I can formulate the semantics of counterfactuals. "Had φ happened, ψ could have happened" holds *iff* ψ is possible after an intervention brings about φ,

$$\mathfrak{M} \models \varphi \mathrel{\Diamond\!\!\rightarrow} \psi \quad \textit{iff} \quad \mathfrak{M}_{\varphi} \models \Diamond\, \psi, \tag{22}$$

[14] I am referring here to deterministic causal models as used by, for example, Hitchcock (2001) and Weslake (2015), where both exogenous and endogenous variables are treated on par, having their values set by equations (just that exogenous equations are nullary). But you can treat exogenous variables differently. Like Halpern (2016:155) or Beckers and Vennekens (2017:11), you can set exogenous values with value vectors, called contexts, and ask what set of solutions is induced by some set of contexts. In such formalism, you can recover genuine notions of causal necessity and possibility, although not disjunctive interventions (because an intervention in this formalism always sets the target variable to a single value). As an aside, I'll flag that I find it unbearably artificial to allow for underdeterminacy in exogenous variables (i.e., contexts) but not endogenous equations: whatever the motivation for introducing multiple contexts (e.g., to model either epistemic or genuinely metaphysical indeterminacy), it extends to endogenous variables.

[15] The notion satisfies Verma and Pearl's (1988) semi-graphoid axioms (Wysocki, 2023b).

[16] For proofs, see (Wysocki, 2023b).

where $\mathfrak{M}_\varphi$ denotes the post-intervention model obtained from $\mathfrak{M}$ with an intervention bringing about φ. “Had φ happened, ψ would have happened” holds *iff* ψ is necessary after an intervention brings about φ,

$$\mathfrak{M} \models \varphi \,\Box\!\!\rightarrow \psi \quad \textit{iff} \quad \mathfrak{M}_\varphi \models \Box\, \psi. \tag{23}$$

The definitions make sense only for non-contradictory antecedents. The two counterfactuals are duals,

$$\mathfrak{M} \not\models \varphi \,\Diamond\!\!\rightarrow \psi \quad \textit{iff} \quad \mathfrak{M} \models \varphi \,\Box\!\!\rightarrow \neg\psi, \tag{24}$$

which follows from (20); a would-counterfactual entails the corresponding might-counterfactual.

All the work in the semantics is offloaded on interventions, and so I need to show how to produce $\mathfrak{M}_\varphi$ for any pre-intervention model $\mathfrak{M}$ and any non-contradictory antecedent φ. The deterministic framework shows how to do that for any conjunctive event, but that’s of little help when the event is disjunctive and can happen in multiple ways. In the context of causal models, one strategy has been followed. Briggs (2012) has proposed to produce a post-intervention model for every way in which the antecedent can happen; a would-counterfactual is true *iff* the consequent holds in every such *piecemeal model*. Briggs’s semantics of events (which they, after Fine, call truthmakers) builds events up as finite boolean combinations of atomic sentences of the form $X = x$, which—if you recall—means the semantics can’t handle infinitely disjunctive events. However, the strategy from piecemeal models could be easily married with the event algebra I proposed here and extended to might-counterfactuals: a would- (might-)counterfactual is true *iff* the consequent holds in all (some) piecemeal models.[17] But I think we could do better and keep everything to a single model. (In Sect. 4, I elaborate on why using a single model counts as doing better.)

The motivating thought is: if you’re tasked with bringing about a disjunctive event, you have freedom, for you *may* bring about any of the ways in which the event can happen; therefore, bringing about a disjunctive event can be naturally modeled with underdeterministic means. And the implementation: let φ be the event to be brought about and $\mathfrak{M}$ be the pre-intervention model; the task is to produce $\mathfrak{M}_\varphi$. If φ is $\bot$, the operation can’t be executed. If φ is $\top$, $\mathfrak{M}$ is also the post-intervention model—$\top$ represents antecedents like “if nothing were to change...” Otherwise, find φ’s canonical form $[\![\varphi]\!] = \bowtie_{i=1}^{n} J_i$ and deal with the juncts separately. I’ll use examples to show how to do that.

Let J_i contain assignments over variables $\vec{X}$; some assignments can be over a subset of $\vec{X}$. First, add a new intervention vertex I to the model; its role is to assign new values to target variables. Encode all assignments from J_i as values in I’s range and, at the

[17] Although Günther (2017) also uses piecemeal models, his strategy is a little different. He proposes to evaluate a would-counterfactual with antecedent φ with a family of piecemeal models, each produced bringing about one of (what he calls) disjunctive situations of φ: “$A = a \vee B = b$ is satisfied if three possible situations are satisfied: (i) $A = a \wedge B = b$, (ii) $A = a \wedge B = \neg b$, and (iii) $A = \neg a \wedge B = b$. We refer to (i)-(iii) as the disjunctive situations or possibilities of the formula $A = a \vee B = b$.” Günther (2017):7. But like Briggs’s semantics, this strategy does not apply to infinite ranges.

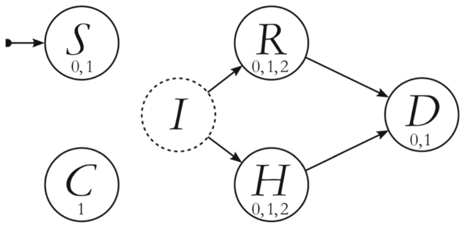


Fig. 3 Simplifying semaphore interventions

same time, the values the vertex may take on:

$$I \leftarrow \{\ulcorner \vec{x} \urcorner \colon\ \vec{x} \in J_i\}, \tag{25}$$

where $\ulcorner \vec{x} \urcorner$ denotes the encoding of the entire assignment $\vec{x}$ as one value, and $\mathcal{R}_I$ spans the same values. Next, for every target variable X from $\vec{X}$, replace its equation. There are two cases: either every value of I encodes X's value or not. In the first case, set X's new equation to:

$$X \leftarrow I[X], \tag{26}$$

where $I[X]$ denotes the value of X encoded in the value of I. In the post-intervention diagram, X is cut from its pre-intervention parents. In the second case, for I's values that don't encode X's, X behaves as if no intervention were performed. That is,

$$X \leftarrow \begin{cases} I[X] & \text{if } I[X] \text{ is set} \\ f_X\big(\vec{P}\big) & \text{otherwise} \end{cases}, \tag{27}$$

where $\vec{P}$ are X's pre-intervention parents, and $f_X\big(\vec{P}\big)$ its pre-intervention equation. You have to repeat this procedure for every junct J_i from φ's canonical form.

An illustration is much needed. The task is to bring about, in the strait model (17), the event "Scylla throws one rock or Charybdis attacks with two heads," whose canonical form consists of one junct, $J = \{\langle 1_R \rangle, \langle 1_R, 2_H \rangle, \langle 2_H \rangle\}$. First, add to the model a new intervention vertex I and set its equation to $I_i \leftarrow \ulcorner 1_R \urcorner, \ulcorner 1_R, 2_H \urcorner, \ulcorner 2_H \urcorner$; let I's range consist of the very same values. Next, modify the equations for R and for H. The new model reads (fig. 2):

$$I \leftarrow \ulcorner 1_R \urcorner, \ulcorner 1_R, 2_H \urcorner, \ulcorner 2_H \urcorner, \qquad S \leftarrow 1, \qquad C \leftarrow 0, 1,$$
$$R \leftarrow \begin{cases} I[R] & \text{if } I[R] \text{ is set} \\ S, S{+}1 & \text{otherwise} \end{cases}, \quad H \leftarrow \begin{cases} I[H] & \text{if } I[H] \text{ is set} \\ C, 2C & \text{otherwise} \end{cases}, \quad D \leftarrow R{+}H \geq 3.$$

R's equation now states: if R's value is encoded in the value of I, i.e., if $I[R]$ is set, assign this new value to R; otherwise, use the old equation $R \leftarrow S, S{+}1$ as if no intervention had taken place. Analogously, H's equation states: assign to H its value decoded from I's value, but if I's value doesn't encode H's, pretend there was no intervention. I call this kind of intervention a *semaphore intervention* because it steers the flow of values in the diagram depending on the value of the intervention vertex.

Table 3 Solutions to $\mathfrak{M}_{R=1\vee H=2}$

	I	S	C	R	H	D		I	S	C	R	H	D
$\vec{\sigma}_1$	$\ulcorner 1_R \urcorner$	1	0	1	0	0	$\vec{\sigma}_6$	$\ulcorner 2_H \urcorner$	1	0	1	2	1
$\vec{\sigma}_2$	$\ulcorner 1_R \urcorner$	1	1	1	1	0	$\vec{\sigma}_7$	$\ulcorner 2_H \urcorner$	1	0	1	2	1
$\vec{\sigma}_3$	$\ulcorner 1_R \urcorner$	1	1	1	2	1	$\vec{\sigma}_8$	$\ulcorner 2_H \urcorner$	1	1	1	2	1
$\vec{\sigma}_4$	$\ulcorner 1_R, 2_H \urcorner$	1	0	1	2	1	$\vec{\sigma}_9$	$\ulcorner 2_H \urcorner$	1	1	2	2	1
$\vec{\sigma}_5$	$\ulcorner 1_R, 2_H \urcorner$	1	1	1	2	1							

This semaphore intervention produces a post-intervention model $\mathfrak{M}_{R=1\vee H=2}$, which agrees with $\mathfrak{M}$ on variables and their ranges but has an additional intervention vertex, one extra equation (for the vertex), and new equations for R and H. The new model has quite a few solutions (tab. 3), but I think it's worthwhile to list them all:

I'll explain what goes on in some of them. In $\vec{\sigma}_1$, the intervention vertex encodes only a value for R, $R = 1$. S's always assigns to it 1, and C's equation allows C to take 0 or 1, and here $\vec{\sigma}_1[C] = 0$. Because $\vec{\sigma}_1[I]$ does encode R's value, R's equation assigns this value to R: $\vec{\sigma}_1[R] = 1$. Because $\vec{\sigma}_1[I]$ doesn't encode H's value, H's equation assigns to H a value using the pre-intervention equation; because $\vec{\sigma}_1[C] = 0$, $\vec{\sigma}_1[H] = 0$. Lastly, $\vec{\sigma}_1[D] = 0$ from $\vec{\sigma}_1[R] = 1$ and $\vec{\sigma}_1[H] = 0$. In $\vec{\sigma}_4$ and $\vec{\sigma}_5$, the value of the intervention vertex encodes both R's and H's values, and hence R, H, and D don't differ between these solutions. The only difference is in C, which still may take 0 or 1 per its equation. In $\vec{\sigma}_6$-$\vec{\sigma}_9$, only H's value is encoded by I's value, and the rest of the model behaves like before the intervention.

Now, I can use the post-intervention model to evaluate some counterfactuals. "Were Scylla to throw one rock or Charybdis to attack with two heads, Odysseus might die" is true in the model, for he dies in the third solution, and therefore the consequent is causally possible in the post-intervention model. "Were Scylla to throw one rock or Charybdis to attack with two heads, Odysseus would die" is false, for he survives in the first solution, and therefore the consequent isn't causally necessary in the post-intervention model.

I still want to show how the strategy works when applied to deterministic models (though then they lose their deterministic status), how it works with canonical forms with multiple juncts, and how to do without intervention vertices in interventions targeting single variables. I'll do that all at once with a single example. Recall the deterministic story of Odysseus and the deterministic model I used to represent it (16). As the counterfactual to evaluate, take "if as many rocks fly as heads attack, and Scylla does or does not notice Odysseus, he will die." The antecedent's canonical form is $[\![(R = H) \wedge (S = 0 \vee S = 1)]\!] = \{\langle 0_R, 0_H\rangle, \langle 1_R, 1_H\rangle, \langle 2_R, 2_H\rangle\} \bowtie \{\langle 0_S\rangle, \langle 1_S\rangle\}$. Therefore, you can bring about $\{\langle 0_R, 0_H\rangle, \langle 1_R, 1_H\rangle, \langle 2_R, 2_H\rangle\}$ with one semaphore intervention, and $\{\langle 0_S\rangle, \langle 1_S\rangle\}$ with another. Because the intervention vertex always assigns values to R and H, in virtue of (26), R and H lose their pre-intervention parents (fig. 3). Second, because the set contains assignments over the same variable, you can do without adding a new intervention vertex; instead, assign these values directly to the variable using an underdeterministic exogenous equation. Generally, the replacement equation is

$$X \leftarrow \{x : \langle x\rangle \in J_i\} \tag{28}$$

whenever all assignments from junct J_i are over X only. The new equations thus read:

$$I \leftarrow \ulcorner 0_R, 0_H \urcorner, \ulcorner 1_R, 1_H \urcorner, \ulcorner 2_R, 2_H \urcorner, \qquad S \leftarrow 0, 1, \qquad C \leftarrow 1,$$
$$R \leftarrow I[R], \qquad H \leftarrow I[H], \qquad D \leftarrow R+H \geq 3.$$

This is how the simplifications work,[18]

The post-intervention model has six solutions and thus is deterministic no longer. In some of them, e.g., $\langle (\ulcorner 1_R, 1_H \urcorner)_I, 1_S, 1_C, 1_R, 1_H, 0_D\rangle$, Odysseus survives, which means his dying isn't necessary in the post-intervention model, and therefore the target would-counterfactual is false. The underdeterminism of the new model doesn't come from the original situation being underdeterministic—which it isn't—but from the fact that the antecedent of the would-counterfactual admits many ways of being brought about. Some of these ways (e.g., $R = 2 \wedge H = 2 \wedge S = 0$) deterministically lead to his death, yet others (e.g., $R = 1 \wedge H = 1 \wedge S = 1$) deterministically lead to his escape, which makes the counterfactual fail.

Another one, due to Quine (1983:23) though not truly a *causal* counterfactual: if Bizet and Verdi had been compatriots, Verdi would have been French. Let B stand for Bizet's nationality, V for Verdi's, and let $\mathcal{R}_B = \mathcal{R}_V$ contain the two-hundred-plus nationalities denoted as natural numbers. Where 33 denotes the French and 39 the Italians, the deterministic model has two exogenous equations,

$$B \leftarrow 33, \qquad V \leftarrow 39,$$

and encodes no counterfactual relationships between the events. Expressed in the language of this model, Quine's counterfactual reads $\mathfrak{M} \models B = V \;\Box\!\!\rightarrow\; V = 33$. It's false. Since $[\![B = C]\!]$ contains all assignments of the same nationality to both composers, $\mathfrak{M}_{B=C}$ now reads[19]:

$$I \leftarrow 1, 2, \ldots, \qquad B \leftarrow I, \qquad V \leftarrow I.$$

Quine's counterfactual is false because only in one of many solutions is Verdi French. What's true, however, is that if Bizet and Verdi had been compatriots, Verdi could

[18] Therefore, an underdeterministic semaphore intervention takes model $\mathfrak{M} = \left\langle \vec{\mathcal{V}}, \vec{\mathcal{I}}, \mathcal{RE} \right\rangle$ and a set of assignments J_i over variables $\vec{X}$ ($\vec{X} \subseteq \vec{\mathcal{V}}$, J cannot be further decomposed into a junction of juncts over non-overlapping variables) and returns a post-intervention model. In cases where $\emptyset \neq J \neq \{\langle\rangle\}$ and the assignments from J aren't all over the same single variable, the post-intervention model is given by $\mathfrak{m} = \left\langle \vec{\mathcal{V}}, \vec{\mathcal{I}} \uplus \{I\}, \mathcal{R} \cup \{\mathcal{R}_I\}, \epsilon \right\rangle$, where $\uplus$ is exclusive union, I is a new intervention node, and ϵ is the new set of equations. ϵ contains the same structural equations as $\mathcal{E}$ except for equations for I, which is given by (25), and for $\vec{X}$, which are given by (26) and (27). If J only contains assignments over the same single variable X, X's new equation is given by (28), and no intervention vertex is added to $\mathfrak{m}$. If $J = \{\langle\rangle\}$, $\mathfrak{m} = \mathfrak{M}$. If $J = \emptyset$, the operation cannot be performed.

[19] Notice that I simplified how I encodes the values of B and C. An identical trick is possible with the previous model.

have been French, and that if Bizet and Verdi had been compatriots, they could have been both Polish.

Why do I insist on starting with the canonical form? It is, in principle, possible to encode all assignments from the target event as values of a single intervention vertex, which then would parent all target variables. Possible, but unbecoming. Since variables that don't share ancestors are unconditionally independent, if two variables are sired by different intervention vertices, they are independent. The resulting diagram is more informative.

This is, I feel, the right place to say more about the interpretation of intervention vertices and their values. Intervention vertices have an odd status. When you use a new intervention vertex, you add it to the family of intervention vertices $\vec{\mathcal{I}}$; because it's a brand-new node, you also add its equation to the family of equations $\mathcal{E}$ and its range to the family of ranges $\mathcal{R}$. Therefore, like variables, intervention vertices have their own equations and figure in the diagram of the model. But unlike variables, you can't intervene on them, and they cannot figure in an event sentence, a modal statement, or a counterfactual because the values of intervention nodes don't denote events that happen *in* the situation represented by the model. Recall the metaphor, which some take more (Menzies & Price, 1993) and others less seriously (Woodward, 2003:123): an intervention corresponds to an action of an (ideal, hypothetical) intervener that she performs on the situation from without in such a way that affects all and only the target variables. Woodward, for instance, explicitly models an intervention with an extra variable that parents only the target variable. In the underdeterministic framework, the metaphor is all the more fruitful. While variables represent events that may happen within the situation, (different) intervention vertices represent possible actions of (different) hypothetical interveners who change the situation from the outside. The interveners are underdeterministic in that they may choose to bring about any assignment from the junct. Actions of different interveners are independent from each other,[20] while a single intervener coordinates the values of the variables under her influence. For instance, in the current example, you can think that one intervener's task is to coordinate the values of R and H, but she may choose between $R = H = 0$, $R = H = 1$, and $R = H = 3$; another intervener's task is to choose the value for S, and she may choose between $S = 0$ and $S = 1$.

While the theory so far evaluates causal counterfactuals correctly, there's another reason to trust the formalism. Semaphore interventions satisfy the following *equivalence requirement*, which seem to capture what we want from bringing disjunctive events about: every history told by $\mathfrak{M}_\varphi$ is told by some piecemeal model $\mathfrak{M}_{\vec{X}=\vec{x}}$, produced by bringing about a single way $\vec{x}$ in which φ can happen, $\vec{x} \in [\![\varphi]\!]$, *and* every history told by any piecemeal model $\mathfrak{M}_{\vec{X}=\vec{x}}$ is told by $\mathfrak{M}_\varphi$. That is, taken together, piecemeal models tell all and only stories told by the post-intervention model:

$$\text{for any } \vec{v}, \qquad \mathfrak{M}_\varphi \models \Diamond \vec{\mathcal{V}} = \vec{v} \quad \textit{iff} \quad \text{for some } \vec{x} \text{ from } [\![\varphi]\!],\ \mathfrak{M}_{\vec{X}=\vec{x}} \models \Diamond \vec{\mathcal{V}} = \vec{v}, \tag{29}$$

[20] According to the notion of independence (21) applied to intervention vertices.

where $\vec{\mathcal{V}} = \vec{v}$ is a history, and $\vec{x}$ is over $\vec{X}$. The equivalence requirement follows the spirit of Briggs's theory, which couches the effects of bringing about a disjunctive event purely in terms of what happens in the family of piecemeal models.

Semaphore interventions do satisfy the equivalence requirement. Take $\vec{Y}{=}\vec{y}$, one of the ways in which φ can happen: $\vec{y} \in [\![\varphi]\!]$, where $\vec{Y} \subseteq \vec{X}$. Per (25), exactly one value of the intervention vertex I encodes this assignment, $I = \ulcorner\vec{y}\urcorner$. Per (27), in any solution $\vec{\sigma}$ where $\vec{\sigma}[I] = \ulcorner\vec{y}\urcorner$, $\vec{\sigma}[\vec{Y}] = \vec{y}$, while the values of the remaining variables $\vec{X} \backslash \vec{Y}$ are determined by their pre-intervention equations. That just means event $\vec{\mathcal{V}} = \vec{\sigma}[\vec{\mathcal{V}}]$ is possible on $\mathfrak{M}_\varphi$ *iff* it's also possible on $\mathfrak{M}_{\vec{Y}=\vec{y}}$. As I encodes all and only the elements of $[\![\varphi]\!]$, the equivalence requirement holds. This also means that theory agrees with Briggs's for any counterfactual that the latter makes a prediction about. For an illustration, look at the solutions to the original post-intervention model $\mathfrak{M}_{R=1\vee H=2}$ (tab. 3). The first three histories (i.e., $\vec{\sigma}_1$-$\vec{\sigma}_3$ projected onto the variables alone) are all and only the histories told by $\mathfrak{M}_{R=1}$, the fourth and fifth are told by $\mathfrak{M}_{R=1\wedge H=2}$, and the remaining ones are told by $\mathfrak{M}_{H=2}$.

The power of the event algebra, as I touted it, lies in modeling infinitely disjunctive events. I began with: had the tower been any taller than 320 ms, it might have collapsed. Let T denote the height of the tower, $\mathcal{R}_T = [0, +\infty)$, and C denote whether it collapses, $\mathcal{R}_C = \{0, 1\}$. Nouguier's calculations have it that the tower may collapse if it's taller than 318 ms and will collapse if taller than 400 ms. In fact, the tower is built 300 ms tall,

$$T \leftarrow 300, \qquad C \leftarrow \begin{cases} 1 & \text{if } 400 < T \\ 0, 1 & \text{if } 318 < T \leq 400 \\ 0 & \text{if } T \leq 318 \end{cases},$$

and on the model's sole solution, the tower doesn't collapse. Within the model, the counterfactual is expressed as $\mathfrak{M} \models T{>}320 \Diamond\!\!\rightarrow C = 1$. $[\![T{>}320]\!] = (320, +\infty)_T$, which is already in the canonical form. In the post-intervention model $\mathfrak{M}_{T>320}$, T gets a new equation, $T \leftarrow (320, +\infty)$, and in some of the model's infinitely many solutions, $C = 1$. Hence, the target counterfactual is evaluated as true.

Another case sets the stage for an underdeterministic causal decision theory. The witch can make only one potion, and she's choosing between two. The first one may extend her youth by $2 - 1/n$ decades, and the second by $2 - 1/n$ or $2 + 1/n$ decades for any n from $\mathbb{N}$. The witch is an agent under bounded uncertainty: she knows the possible consequences of drinking each potion, but not how probable they are. In fact, nothing indicates that one outcome is more likely than another, yet assigning equal probability to all outcomes of each potion leads to a contradiction, which means that no probabilistic model can accurately describe the situation.[21] Moreover, say that the witch is extremely risk averse and always wants to guard herself against the worst possible outcome—i.e., she applies the leximin strategy, at least when it comes to potions. To decide which potion to make, she'll construct an underdeterministic model first. Where $P = 1$ denotes making the first potion, $P = 2$ making the second,

[21] If the probabilities are 0, they sum to 0; if they are anything more than 0, their sum doesn't exist (i.e., it's infinite). In either case, they don't sum to 1 and thus violate the probability axioms.

and E denotes the effect of the potion, the witch's model states:

$$P \leftarrow \text{-}1, 1, \qquad E \leftarrow \{2 - 1/n : n \in \mathbb{N}\} \cup \{2 + 1/n : n \in \mathbb{N} \wedge P = 2\}\,.$$

(The second set in E's equation is $\emptyset$ if $P = 1$.)

Now, this is not the place for presenting the entire underdeterministic causal theory, and hence I'll just sketch the way forward. Following previous applications of causal models to causal decision theory (Hitchcock, 2016; Meek & Glymour, 1994), represent the actions available to the witch with interventions: $P = 1$ for deciding to make the first potion, and $P = 2$ for deciding to make the second one instead. These yield two models, $\mathfrak{M}_{P=1}$ and $\mathfrak{M}_{P=2}$, whose solutions she must now compare with respect to the value of E (the higher, the better). As every solution from $\mathfrak{M}_{P=1}$ has a counterpart in some solution from $\mathfrak{M}_{P=2}$, and every solution from $\mathfrak{M}_{P=2}$ that doesn't have a counterpart is better than every solution from $\mathfrak{M}_{P=1}$, the witch should make the second potion. In the worst case, each potion will give her an extra decade of youth, but only the second potion may extend her youth by more than two decades.[22]

While the potion case doesn't require using semaphore interventions, there are decision-theoretic notions that do. First, there's an obvious underdeterministic analog of mixed strategies. Say, the witch has recipes for seven potions, $P \in \{1, ..., 7\}$ and is vacillating between making the first or second on the one hand, or the sixth or seventh on the other. She can compare these mixed strategies by comparing the solutions to $\mathfrak{M}_{P=1 \vee P=2}$ with the solutions to $\mathfrak{M}_{P=6 \vee P=7}$. Second, Meek and Glymour (1994) represent deliberate actions as values of an intervention vertex added to the model, while the agent acting automatically or instinctively (i.e., the action that would have happened, had the agent not deliberated) is represented by the pre-intervention model. In this context, semaphore interventions are well suited, for instance, to represent what may happen under any action, deliberate or habitual: where $I = i$ for $i = 1, \ldots, n$ denote the actions available to the agent, $\top \vee \bigvee_i I = i$ is the event to be brought about, so that the post-intervention model can represent all consequences available to the agent.[23]

3.5 Composing counterfactuals; structural and parametric interventions

With little effort it's possible to extend the framework to accommodate composing counterfactuals and to devise underdeterministic analogs of structural and parametric interventions—the probabilistic originals are used in the context of causal discovery. The rules of interpreting counterfactuals, (22) and (23), don't tell you how to parse counterfactuals with other counterfactuals in the antecedent, but two extra rules will:

$$\mathfrak{M} \models \varphi \mathrel{\Diamond\!\!\rightarrow} (\psi \mathrel{\Diamond\!\!\rightarrow} \theta) \quad \textit{iff} \quad \mathfrak{M}_\varphi \models \psi \mathrel{\Diamond\!\!\rightarrow} \theta, \tag{30}$$

[22] The theory has all the hallmarks of a causal decision theory. For instance, it's possible to formulate an underdeterministic Newcomb's case: I may or may not be predestined for hell; if I'm predestined for hell, and only then, will I sin and will I go to hell. Though I enjoy sinning, no time spent in hell is worth it. Going to hell and sinning are conditionally dependent per (21), yet I should sin, for doing so doesn't cause my going to hell (Wysocki ms1).

[23] This is an example of an underdeterministic parametric intervention, which I am about to discuss.

and

$$\mathfrak{M} \models \varphi \mathrel{\Box\!\!\rightarrow} (\psi \mathrel{\Box\!\!\rightarrow} \theta) \quad \textit{iff} \quad \mathfrak{M}_{\varphi} \models \psi \mathrel{\Box\!\!\rightarrow} \theta. \tag{31}$$

To evaluate a compound counterfactual, intervene on the original model $\mathfrak{M}$ with an intervention corresponding to the first antecedent, then intervene on the resulting model again with an intervention corresponding to the second antecedent, and so on, until you can apply (22) or (23).

As an illustration, consider: "if Scylla throws one rock or Charybdis attacks with two heads, then if Scylla fails to see Odysseus, he may survive," $\mathfrak{M} \models R = 1 \vee H = 1 \mathrel{\Diamond\!\!\rightarrow} \big(S = 0 \mathrel{\Diamond\!\!\rightarrow} D = 0\big)$. Apply (30) to get $\mathfrak{M}_{R=1\vee H=1} \models S = 0 \mathrel{\Diamond\!\!\rightarrow} D = 0$. Apply it again to get $\mathfrak{M}_{R=1\vee H=1, S=0} \models \Diamond\, D = 0$, where the interventions in $\mathfrak{M}_{R=1\vee H=1, S=0}$, separated by the comma, are to be executed from left to right—bring about $R = 1 \vee H = 1$ and then, in the model you just obtained, bring about $S = 0$. In some of the solutions to the post-intervention model, e.g., $\langle \ulcorner 1_H \urcorner_I, 0_S, 1_C, 0_R, 1_H, 0_D \rangle$, Odysseus survives, and the counterfactual is true.

The theory of underdeterministic counterfactuals requires only semaphore interventions, yet the framework can afford more—every intervention possible in the probabilistic framework has (I boldly surmise) an analog in the underdeterministic framework. For instance, Eberhardt and Scheines (2006) distinguish between structural and parametric interventions in causal discovery.[24] A probabilistic *structural* intervention introduces an intervention vertex that works as a switch: when the switch is off, the target variables behave as they did in the pre-intervention model. When the switch is on, the target variables become independent from any other variables, and are governed by a new probabilistic distribution. For an underdeterministic analog, take a semaphore intervention that brings about $\top \vee \varphi$, where φ is the counterpart of the new distribution (i.e., φ specifies the new possible values for the target variables). Per (2) and (9), $[\![\top \vee \varphi]\!] = \{\langle\rangle\} \cup [\![\varphi]\!]$, and therefore for $I = \ulcorner\langle\rangle\urcorner$, the entire model behaves as before the intervention—the switch is off. For all other values of the intervention vertex, the switch is on. A probabilistic *parametric* intervention is like the structural intervention, but when the switch is on, the target variables still depend on pre-intervention parents, although the conditional distribution is new. To implement an underdeterministic parametric intervention, you can modify semaphore interventions into a fourth species. Say, for every target variable X, its new possible values are given by $g_X\big(\vec{P}, I\big)$. Then for every X, replace $I[X]$ in (27) with $g_X\big(\vec{P}, I\big)$, so that when the switch is on (i.e., I's value is set for X), X's value is determined by a new equation. When the switch is off, however, X behaves according to its pre-intervention equation.

Two remarks to crown the theory: first, since semaphore interventions can encode any set of tuples, it's possible to modify the event algebra or even use a completely different system of deriving such sets. The resulting account would still have the advantage of producing one post-intervention model. Second, I promise that the algebra and semaphore interventions can be adapted to the probabilistic framework (Fenton-Glynn, 2017; Twardy & Korb, 2011). An event would have to be accompanied by a probability distribution over all the ways in which it can be brought about, and the equations of intervention vertices would have to reflect this probability. In fact, Rosella and Sprenger (2022) develop a probabilistic analog of Briggs's theory, associating with

[24] Also see Korb et al. (2004).

each piecemeal model a probability proportional to how similar the model is to the pre-intervention original; it seems that their theory could easily be reformulated with semaphore interventions. As this project comes with its own philosophical and technical challenges, however, delivering on this promise will have to wait for another occasion. All I want to convey for now is that the significance and utility of the event algebra and semaphore interventions isn't meant to be limited to underdeterministic causation.

4 Doubts

Two elements of my formalism are in need of further motivation: developing the underdeterministic framework as a tool for investigating causal underdeterminism; and using semaphore interventions rather than families of piecemeal models (Briggs, 2012; Günther, 2017) or other theories capable of modeling infinitely disjunctive antecedents (Hiddleston, 2005; Huber, 2013; Vandenburgh, 2022).

I have already argued why underdeterministic causal phenomena are worth paying attention to (Sect. 3.1). However, the probabilistic framework could ostensibly handle underdeterministic counterfactuals: instead of underdeterministic equations, use counterfactual probabilities that encode necessity as 1, possibility as any value from (0, 1), and impossibility as 0. (For uncountably infinite ranges, an impossible event would be instead encoded as one of zero density, and a possible event as one of any positive density; a possible event would count as necessary *iff* all other events had zero density.) But this appropriation would violate the third desideratum for a semantics of counterfactuals, hindering both our understanding of underdeterminism and the practical applications of the theory.

So, first, there's a philosophical reason against modeling underdeterminism this way. Using the probabilistic framework for analyzing probabilistic causal phenomena carves them at their joints, or approximately so, but using it for analyzing underdeterministic phenomena unnecessarily obfuscates the picture. For instance, it is of some philosophical importance that there's a simple qualitative notion of conditional independence (21); this fact would be lost on you if you limited yourself to the probabilistic framework. No one advocates abandoning deterministic causal models because they can be simulated by probabilistic models, even if it's in principle possible[25]; if so, we shouldn't eschew underdeterministic models either.

Second, there's a practical reason against modeling underdeterminism this way. Probabilistic models are much more complex than underdeterministic ones, for each structural equation returns a probability distribution rather than a set of possible values (Fenton-Glynn, 2017). If the exact probabilities were to be eventually ignored, simplicity clearly suggests using the underdeterministic models. Using underdeterministic models is also computationally cheaper. Say, you want to know if some event is causally impossible, necessary, or merely possible. Even if you eventually ignore its specific probability, you have to compute it first, or else you won't know whether it's

[25] This would require using degenerate counterfactual distributions assigning 1 to the outcome of the structural equation and 0 to any other value from the range.

0, 1, or something in between. But computing underdeterministic structural equations is much faster than computing corresponding counterfactual probabilities.

Third, there's a scientific reason against modeling underdeterminism with the probabilistic framework. We do seem to reason about many everyday non-deterministic situations explicitly without invoking probabilities. It is an open empirical question whether explicit non-probabilistic inferences are implemented as implicit probabilistic causal inferences or as its own mechanism—for instance, in the form of underdeterministic models. But even to formulate this question, a cognitive scientist needs underdeterministic models as a separate device from probabilistic models. In the context of the cognitive science of causal reasoning, simulating underdeterministic models with probabilistic ones means a priori subscribing to the first alternative.

The last reason is technical: you can't ascribe probability to elements of an infinite set such that, for some set positive threshold, no matter how small, every element is more likely than the threshold. Because our minds and our machines are finite and thus use finite numerical representations of probability, the probability of some possible events will be calculated as 0.[26] This means that some underdeterministic models can't be simulated in practice by probabilistic models.

In need of further motivation is also my use of semaphore interventions. Briggs's (2012) strategy easily extends to underdeterministic models. Since we have one familiar and intuitive solution, we don't need another, which seems complicated, as it requires a new type of nodes that take on values that encode other values. The event algebra is still necessary for identifying the piecemeal models to be constructed, but once that's done, we can use Briggs's original mechanism. But there's a philosophical reason to go with semaphore interventions. First, bringing about a disjunctive event counts as underdeterministic—it may be done in many different ways—and so it's only natural to apply the mechanics already present in underdeterministic models. It would seem inconsistent, that is, to represent one type of underdeterminacy with underdeterministic equations but another type with multiple models. Second, it's common in the literature to introduce extra nodes to represent interventions (2003:47) and deliberative actions of agents (Meek & Glymour, 1994), and structural interventions, which behave a lot like semaphore interventions, have already been used (Eberhardt & Scheines, 2006). Under the metaphor that different values of the intervention node correspond to different possible actions of the ideal intervener, representing her actions as values of one variable follows previous practice.

Another reason for the new formalism is that, unlike families of piecemeal models, semaphore interventions satisfy the last desideratum for a semantics of counterfactuals. So, semaphore interventions let the modeler understand the target situation better. Using the strategy of piecemeal models leaves her with multiple models characterized by different directed graphs. Testing, for instance, whether some variables are dependent post-intervention won't be straightforward because they may be independent in every piecemeal model but not in the situation itself, as it would be represented by

[26] Of course, no mind or machine will store at once infinitely many solutions to a model. But the problem described applies, for instance, to the physically feasible task of checking whether some assignment is a solution to a model. (I'll come back to this point soon.) Finite numerical representations will force a physically implemented probabilistic model to produce false negatives for infinitely many solutions.

a single model after a semaphore intervention.[27] More generally, bringing about a disjunctive event leaves us with one underdeterministic situation (i.e., one situation that may develop in many ways), and one situation deserves to be modeled with one model. Piecemeal models give us piecemeal understanding.

Moreover, in the context of cognitive science, it seems more plausible (or at least a hypothesis worth investigating) that we use a single model to think of counterfactuals with disjunctive antecedents rather than a possibly very numerous family of models. Therefore, semaphore interventions seem like a potentially useful tool for psychologists working on causal cognition.

There's also a computational advantage to using semaphore interventions: generating one model rather than many is just computationally cheaper, and the difference becomes most acute when you're facing a choice between one and infinitely many models. In the case of models with infinitely many solutions, no machine will be able to store all of their solutions, but this doesn't make the point moot. Consider, for instance, a simple algorithm that tests whether a solution satisfies a model. If you were using a single model, you could generate it once and run the algorithm many times on different inputs. If you were using piecemeal models, you'll likely have to generate a different model for different inputs. Or take a non-deterministic algorithm that generates solutions to a model. The strategy from piecemeal models, unlike the one from semaphore interventions, would require producing a randomly chosen piecemeal model each time before producing a solution.

The theory of underdeterministic counterfactuals forces me to take a stance on some issues discussed in the counterfactual literature. Like Briggs's, my account doesn't respect the principle of conditional excluded middle—I side with Lewis (1973), Bennett (2003), and Hall (2004) against DeRose (1999), Stalnaker (1980), and Williams (2010); this also means that underdeterministic might-counterfactuals aren't equivalent to would-counterfactuals hedged with the epistemic possibility operator.[28] Relatedly, I again agree with Lewis (1973) and Hájek (2021) rather than Edgington (2004), Schaffer (2004), and Walters (2009) that if a genuinely nondeterministic coin landed heads, "had it been flipped, it would have landed heads" is false. Defending underdeterministic counterfactuals from these authors requires a full-fledged argument, which I cannot make here, but I think I can at least say this. The main motivation behind underdeterministic counterfactuals is to model causal phenomena and aid decision making. The motivation behind the authors I disagree with is to explain linguistic intuitions (Stalnaker, 1980; DeRose, 1999) and assist in reasoning about the past (Edgington, 2004). Differences in motivation translate into differences in tools, and a tool

[27] Two variables may share no ancestors in every piecemeal model—and therefore be unconditionally independent—yet share an ancestor, an intervention node, in the model produced by the semaphore intervention.

[28] The epistemic interpretation implies that an event is either causally necessary or causally impossible. "φ could have happened" is equivalent to the counterfactual "had everything been as it was, φ could have happened," $\top \mathbin{\diamond\!\!\rightarrow} \varphi$; therefore, according to the epistemic interpretation, "φ could have happened" is equivalent to "it's consistent with the speaker's knowledge that φ would have happened." This means that observing that φ happens (doesn't happen) compels the speaker to think that φ was causally necessary (impossible). In short: the epistemic interpretation entails that statements of the form "the coin landed heads but could have landed tails" are infelicitous Moore's sentences. In contrast, the underdeterministic treatment judges such sentences as felicitous.

shouldn't be criticized for how it performs when applied to what it wasn't intended for.

I can now discuss, at last, alternative interventionist theories of counterfactuals that can deal with infinitely disjunctive antecedents, Huber's (2013), Vandenburgh's (2022), and Hiddleston's (2005): how do they differ from mine, and why do we need mine? The answers depend on the theory, though the common theme is that they all rely on similarity measures, transplanted onto the interventionist ground, which makes them violate some of the desiderata and renders them unusable for modeling genuine underdeterminism.

First, take Huber's (2013). The theory is complex, and I'm forced to present just enough to point out the differences. Huber is motivated by the developments in the causal literature that extend causal models with normality orderings over events to account for otherwise unaccountable causal intuitions (Gallow, 2021; Hall, 2007; Halpern & Hitchcock, 2015; Halpern, 2016). His goal is to unify normality orderings and counterfactual structure into a single similarity ordering that yields both counterfactuals and causal judgments. On his theory, worlds are identified with assignments over all variables, similarity is a complete order over worlds that satisfies certain conditions, and a would-counterfactual is true *iff* its consequent is true in all worlds that satisfy the antecedent and are most similar to the actual world. Because of this 'all worlds' clause, the theory can in principle handle infinitely disjunctive antecedents—you simply survey all value assignments that satisfy the antecedent without resorting to anything like the event algebra.

This theory is wanting, however, in the light of what I expect of a theory of counterfactuals. First, neither does it help the modeler understand the modeled situation, nor will it be useful to cognitive scientists, since you trade structural equations, which are easy to reason with, for orderings over all value assignments, which are not. That is, while Huber obtains the (extremely interesting and metaphysically significant) result that structural equations and normality can be unified into a single modality, his formalism isn't very handy—which isn't a problem given his aims but is given mine. Second, and worse, the theory doesn't apply to underdeterminism. One of the indispensable assumptions in the theory is strong centering: any non-actual world is strictly less similar to the actual world than the actual world itself. But that means that his account leaves no room for causal modalities and genuine might-counterfactuals, as the coin example above shows. If a genuinely non-deterministic coin lands heads, "it could have landed tails" is false on Huber's theory, for the world where it landed tails fails the similarity condition and therefore doesn't fall under the 'all worlds' clause. This means, in turn, that Huber's formalism doesn't satisfy the second desideratum. Third, the assumption behind the theory is that the most promising strategy for defining token causation involves normality orderings. However, not everyone agrees (Blanchard & Schaffer, 2017; McDonald, 2023a). In fact, Wysocki (2023a) has recently shown that you can combine the cases that have motivated normality orderings so that no normality ordering can deliver all the correct judgments. Normality orderings, that is, fail for the very cases that they were promised to solve. But if they cannot deliver, Huber's semantics won't help even with theories of deterministic causation, pace, again, the second desideratum.

Next is Vandenburgh's theory (2022). The main difference between his and other interventionist theories (including mine) is that his bars interventions on endogenous nodes. This move allows him to model backtracking counterfactuals (which other theories typically cannot model) and to account for disjunctive antecedents. To bring about a disjunctive event, you are allowed to make minimal interventions—in terms of the number of exogenous variables targeted—that make the event true in the model. This works even for infinitely disjunctive events.

However, this theory doesn't satisfy my desiderata either. First, it's not conservative: without interventions on endogenous nodes, the theory cannot be used in standard accounts of token causation (Glymour & Wimberly, 2007; Hitchcock, 2001; Woodward, 2003; Halpern & Pearl, 2005; Weslake, 2015) without significantly modifying these accounts. Second, like Briggs's solution, Vandenburgh's also requires building one piecemeal model for each intervention, which for infinitely disjunctive events potentially means an infinity of models, violating the third desideratum that the semantics should be machine-friendly and potentially helpful to cognitive scientists.

And take Hiddleston's (2005) semantics; again, I can afford to introduce only the relevant pieces. Like Vandenburgh's, this one constrains what interventions are allowed: to bring about an event, you can intervene on a variable only if no intervention on its descendants could make this event true. This constraint serves Hiddleston as a similarity measure: the more downstream (in terms of the graph's directed edges) the intervention, the more similar the model is to the pre-intervention original. Now, similarity makes the familiar problem appear. If a genuinely non-deterministic coin lands heads, his theory predicts that it could have landed tails; but "it landed heads but could have landed tails" sounds perfectly reasonable. The theory, that is, can't account for genuine underdeterminism. The other problems are present too. In cases where the event still must be brought about with separate interventions, the procedure produces a family of piecemeal models, which for infinitely disjunctive events means potentially an infinity of models.

Last, a problem that I find decisive: strong centering, to which all three theories are committed, yields inappropriate consequences when it comes to disjunctive antecedents. Say, I will sing if it rains but won't if it's sunny. It rains, I sing. Entertain: if it were sunny or rainy, I *would* sing. The counterfactual is false, for I wouldn't sing in the sun. However, all three theories evaluate the counterfactual as true, as you need no interventions to bring about the antecedent. It's already raining, and any changes to the world to bring about this event would be, on their theories, superfluous. This is no good. On my proposal, in contrast, the counterfactual is false, for in the possible history where it's sunny the consequent doesn't hold.

However, dispensing with strong centering also has negative consequences. Notice that both "if the tower is taller than 320 ms, it may collapse" and "if the tower is no smaller than 1 m, it may collapse" turn out true on my theory, yet the second claim sounds infelicitous, as the risk of collapse begins at 318 ms only. I think this infelicity can be given a Gricean explanation: if you know that the tower won't collapse if it's no taller than 318 ms, your uttering the latter might-counterfactual is misleading, even if not strictly false. It is not contradictory to say "if the tower is no smaller than 1 m, it may collapse, but I don't mean to say that if it's smaller than 318 ms, it may collapse," especially if you helpfully add "because, e.g., 400 ms is more than 1 m, and a 400-meter-tall tower may collapse." That is, "if the tower is no smaller than 1 m, it may collapse" comes with the implicature "if the tower is exactly 1 m tall, it may collapse," but the implicature is cancellable, and hence (ought we to trust Grice) doesn't bear on the truth conditions of the target counterfactual.

5 Conclusions

I have proposed two somewhat independent devices: the event algebra and semaphore interventions. The algebra provides rules for identifying all the ways in which an event can happen, allowing for modeling infinitely disjunctive events. Moreover, every event has a unique canonical representation as the conjunction of the maximal number of independent disjunctive events. Semaphore interventions allow for bringing about a set of assignments within a single model, which has both theoretical and practical advantages over using a family of models. Together, the devices offer a natural semantics of underdeterministic counterfactuals with disjunctive antecedents and promise further applications to probabilistic counterfactuals (Rosella & Sprenger, 2022), interventionist theories of decision-making (Hitchcock, 2016; Meek & Glymour, 1994), and deterministic (Andreas & Günther, 2021; Beckers & Vennekens, 2017; Gallow, 2021; Halpern & Pearl, 2005; Weslake, 2015), probabilistic (Fenton-Glynn, 2017; Twardy & Korb, 2011), and underdeterministic causation (Wysocki, ms2).

6 Appendix

6.1 The syntax and semantics of model statements

Here, I present the syntax and semantics of event names and model statements—the statements that can be true or false in a model—in one place.

First, the grammar:

		production rule
brick sentence	$B \rightarrow$	an expression over some variables from $\vec{\mathcal{V}}$; the exact syntax depends on the ranges
event name	$E \rightarrow$	B,$(\neg E)$, $(E \wedge E)$, $(E \vee E)$, $(E \rightarrow E)$, $(E \equiv E)$,
modal sentence	$M \rightarrow$	$\Box\, E$, $\Diamond\, E$,
counterfactual	$C \rightarrow$	$E \mathrel{\Box\!\!\rightarrow} E$, $E \mathrel{\Box\!\!\rightarrow} (C)$, $E \mathrel{\Diamond\!\!\rightarrow} E$, $E \mathrel{\Diamond\!\!\rightarrow} (C)$,
model statement	$S \rightarrow$	(C), $(\neg S)$, $(S \wedge S)$, $(S \vee S)$.

In all cases used throughout, brick sentences were arithmetic equations or inequalities. The grammar allows for model statements like $(\neg \Diamond \varphi) \wedge (\varphi \mathrel{\Diamond\!\!\rightarrow} \psi \mathrel{\Diamond\!\!\rightarrow} \gamma)$, but not for $\Box \Diamond \varphi$ or $\varphi \mathrel{\Box\!\!\rightarrow} \Diamond \psi$.

Let $\mathfrak{M}$ be the (pre-intervention) model of evaluation, $\mathfrak{M}^{\top}$ be the set of solutions to $\mathfrak{M}$, ψ be a name over $\vec{X}$ ($\vec{X} \subseteq \vec{\mathcal{V}}$), φ be any name of a consistent event (i.e., $[\![\varphi]\!] \neq \emptyset$)

over some subset of $\vec{\mathcal{V}}$, $\vec{\sigma}$ be any assignment over all its nodes, $\vec{\mathcal{V}} \cup \vec{\mathcal{I}}$,[29] and S_1, S_2 be any model statements.[30] Below is the full semantics; although it doesn't need to be expressed in terms of the event algebra, it can, and I provide the alternative rules too:

$$\mathfrak{M}, \vec{\sigma} \models \psi \qquad \textit{iff} \text{ substituting } \vec{x}[X] \text{ for all } X \text{ from } \vec{X} \text{ makes } \psi \text{ true} \qquad \textit{iff } \{\vec{\sigma}\} \bowtie [\![\psi]\!] = \{\vec{\sigma}\}, \tag{32}$$

$$\mathfrak{M} \models \Diamond\, \psi \qquad \textit{iff } \mathfrak{M}, \vec{\sigma} \models \psi \text{ for some } \vec{\sigma} \text{ from } \mathfrak{M}^\top \qquad \textit{iff } [\![\psi]\!] \bowtie \mathfrak{M}^\top \neq \emptyset, \tag{33}$$

$$\mathfrak{M} \models \Box\, \psi \qquad \textit{iff } \mathfrak{M}, \vec{\sigma} \models \psi \text{ for all } \vec{\sigma} \text{ from } \mathfrak{M}^\top \qquad \textit{iff } [\![\psi]\!] \bowtie \mathfrak{M}^\top = \mathfrak{M}^\top, \tag{34}$$

$$\mathfrak{M} \models \varphi \mathrel{\Diamond\!\!\rightarrow} \psi \qquad \textit{iff } \mathfrak{M}_\varphi \models \Diamond\, \psi \qquad \textit{iff } [\![\psi]\!] \bowtie \mathfrak{M}_\varphi^\top \neq \emptyset, \tag{35}$$

$$\mathfrak{M} \models \varphi \mathrel{\Box\!\!\rightarrow} \psi \qquad \textit{iff } \mathfrak{M}_\varphi \models \Box\, \psi \qquad \textit{iff } [\![\psi]\!] \bowtie \mathfrak{M}_\varphi^\top = \mathfrak{M}_\varphi^\top, \tag{36}$$

$$\mathfrak{M} \models \neg S_1 \qquad \textit{iff } \mathfrak{M} \not\models S_1, \text{ i.e., it's not the case that } \mathfrak{M} \models S_1, \tag{37}$$

$$\mathfrak{M} \models S_1 \wedge S_2 \qquad \textit{iff } \mathfrak{M} \models S_1 \text{ and } \mathfrak{M} \models S_2, \tag{38}$$

$$\mathfrak{M} \models S_1 \vee S_2 \qquad \textit{iff } \mathfrak{M} \models S_1 \text{ or } \mathfrak{M} \models S_2. \tag{39}$$

6.2 An event has at most one canonical form

Theorem *A non-empty set of assignments over at least one variable has a single canonical form, which is the junction of the maximum number of assignment sets (juncts) over non-overlapping variables:*

$$S_{\vec{X}} = \mathop{\bowtie}_{i=1}^{n} J_i, \tag{40}$$

where: $S_{\vec{X}}$ is over variables $\vec{X}$, $\vec{X} \neq \emptyset$, assignments from J_i don't share variables with assignments from J_j ($i \neq j$), and $J_i \neq \{\langle\rangle\}$.

Proof The proof will be by induction on the maximum number of juncts the event decomposes into. First, notice that if a set of assignments is a junction of other sets of assignments, then the projection of the former onto some variables is a junction of the projections of the latter:

$$\text{if} \quad U_{\vec{X}} = V_{\vec{Y}} \bowtie W_{\vec{Z}}, \quad \text{then} \quad U_{\vec{X}}[\vec{P}] = V_{\vec{Y}}[\vec{P} \cap \vec{Y}] \bowtie W_{\vec{Z}}[\vec{P} \cap \vec{Z}], \tag{41}$$

[29] The rules also work if $\vec{\sigma}$ is an assignment over $\vec{\mathcal{V}}$ only.

[30] I use $\mathfrak{M}^\top$ for the set of all solutions, as elsewhere (Wysocki, 2023b) I use $\mathfrak{M}^\psi$ for the set of all solutions that satisfy ψ.

where $\vec{X} = \vec{Y} \cup \vec{Z}$, $\vec{P} \subseteq \vec{X}$, and $U_{\vec{X}}[\vec{P}]$ is a projection of $U_{\vec{X}}$ onto $\vec{P}$ (read $V_{\vec{Y}}[\vec{P} \cap \vec{Y}]$ and $W_{\vec{Z}}[\vec{P} \cap \vec{Z}]$ analogously).[31] This property follows from (1), the definition of junction, because if an assignment is a spatial concatenation of two other assignments, then a projection of the former is a spatial concatenation of the corresponding projections of the latter.

Now, for the inductive part of the proof.

Base step The event denoted by any sentence over one variable is already in the canonical form. (The only non-empty set of assignments over no variables is $[\![\top]\!] = \{\langle\rangle\}$; as a convention, assume that this is its canonical form.)

Inductive step Assume property (40) holds for sentences over fewer than k variables. Take $S_{\vec{X}}$, a set of assignments over k variables. If $S_{\vec{X}}$ doesn't decompose any further, it trivially satisfies the property, as the set is already in the canonical form.

Otherwise, decompose $S_{\vec{X}}$ as

$$S_{\vec{X}} = K_{\vec{Y}} \bowtie L_{\vec{Z}}, \tag{42}$$

where $\vec{X} = \vec{Y} \cup \vec{Z}$, $\vec{Y} \neq \emptyset \neq \vec{Z}$, and $L_{\vec{Z}}$ doesn't decompose any further. As $K_{\vec{Y}}$ is now over fewer than k variables, in virtue of the inductive assumption, $K_{\vec{Y}} = \bowtie_{i=1}^{n} J_i$ for some n and conjuncts $\{J_i\}_{i=1}^{n}$. Therefore, the canonical form for the target set is:

$$S_{\vec{X}} = \left(\bowtie_{i=1}^{n} J_i\right) \bowtie L_{\vec{Z}}. \tag{43}$$

What's left is to prove that this decomposition is unique. Decompose $S_{\vec{X}}$ into any two juncts:

$$S_{\vec{X}} = M_{\vec{P}} \bowtie N_{\vec{Q}}, \tag{44}$$

where $\vec{X} = \vec{P} \cup \vec{Q}$, and $\vec{P} \neq \emptyset \neq \vec{Q}$. Either $\vec{Z} \subseteq \vec{Q}$ or $\vec{Z} \subseteq \vec{P}$, because otherwise, per (41), $L_{\vec{Z}}$ could be decomposed further into two juncts: one over some variables from $\vec{Q}$ and the other over some variables from $\vec{P}$. For focus, assume $\vec{Z} \subseteq \vec{Q}$. This means that $N_{\vec{Q}} = O_{\vec{R}} \bowtie L_{\vec{Z}}$, where $\vec{R} = \vec{Q} \setminus \vec{Z}$. Therefore,

$$S_{\vec{X}} = M_{\vec{P}} \bowtie N_{\vec{Q}} = M_{\vec{P}} \bowtie O_{\vec{R}} \bowtie L_{\vec{Z}} = K_{\vec{Y}} \bowtie L_{\vec{Z}}. \tag{45}$$

In virtue of the inductive assumption, $M_{\vec{P}} \bowtie O_{\vec{R}}$ has a unique canonical form, and it's the same as the one for $K_{\vec{Y}}$, because $M_{\vec{P}} \bowtie O_{\vec{R}} = K_{\vec{Y}}$. That in turn means that (43) still gives the canonical form of $S_{\vec{X}}$. Therefore, $S_{\vec{X}}$ has a unique canonical form. This completes the inductive proof. □

Acknowledgements This paper benefited greatly from the suggestions from the two reviewers, who seriously engaged with the paper and helped me clarify its contribution much better; from Ray Briggs and

[31] I.e., $U_{\vec{X}}$ contains assignments over variables from $\vec{X}$, though not every assignment needs to be over $\vec{X}$. $U_{\vec{X}}[\vec{P}]$ contains assignments from $U_{\vec{X}}$ where the values for $\vec{X} \setminus \vec{P}$ have been trimmed off from the assignments.

Jonathan Vandenburgh, who were my APA commentators and gave me both good feedback and encouragement; from Jim Woodward, Dmitri Gallow, Clark Glymour, and Colin Allen, who helped me shape this theory when it was still a thesis chapter; and from Jason Kay and Stephen Mackereth.

Funding Funding was provided by Office for Science and Technology of the Embassy of France in the United States (Chateaubriand Fellowship 2022), Japan Society for the Promotion of Science London (Grant No. PE22039), and Deutsche Forschungsgemeinschaft (Reinhart Koselleck project WA 621/25-1).

Conflict of interest None declared.

References

Andreas, H., & Günther, M. (2021). Difference-making causation. *Journal of Philosophy, 118*(12), 680–701.
Bell, J. L. (2016). Infinitary Logic. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (Ed.). https://plato.stanford.edu/archives/win2016/entries/logic-infinitary/.
Barbero, F., & Sandu, G. (2021). Team semantics for interventionist counterfactuals: Observations vs. interventions. *Journal of Philosophical Logic, 50*(3), 471–521.
Beckers, S., & Vennekens, J. (2017). The transitivity and asymmetry of actual causation. *Ergo, 4*(1), 1–27.
Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.
Biswas, T. (1997). *Decision-making under uncertainty*. Cambridge: MIT Press.
Blanchard, T., & Schaffer, J. (2017). Cause without default. *Beebee* (pp. 175–214). Price, making a difference. Oxford: Hitchcock.
Briggs, R. A. (2012). Interventionist counterfactuals. *Philosophical Studies, 160*(1), 139–166.
Cohen, M., & Jaffray, J. (1983). Rational choice under complete ignorance. *Econometrica, 48*, 1281–1299.
Deœux, T. (2019). Decision-making with belief functions: A review. *International Journal of Approximate Reasoning, 109*, 87–110.
DeRose, K. (1999). Can it be that it would have been even though it might not have been? *Philosophical Perspectives, 13*, 385–413.
Eberhardt, F., & Scheines, R. (2006). Interventions and causal inference. *Philosophy of Science, 74*(5), 981–995.
Edgington, D. (2004). Counterfactuals and the benefit of hindsight. In Dowe & Noordhof (Eds.), *Cause and chance: Causation in an indeterministic world.* Routledge.
Ellis, B., Jackson, F., & Pargetter, R. (1977). An objection to possible-world semantics for counterfactual logics. *Journal of Philosophical Logic, 6*, 355–357.
Fenton-Glynn, L. (2017). A proposed probabilistic extension of the Halpern and pearl definition of 'actual cause'. *The British Journal for the Philosophy of Science, 68*(4), 1061–1124.
Fine, K. (2012). Counterfactuals without possible worlds. *The Journal of Philosophy, 109*(3), 221–246.
Gallow, D. (ms.). Dependence, defaults, and needs.
Gallow, D. (2021). A model-invariant theory of causation. *Philosophical Review, 130*(1), 45–96.
Glymour, C., & Wimberly, F. (2007). Actual causes and thought experiments. In Silverstein O'Rourke (Ed.), *Campbell* (pp. 43–67). Cambridge: Causation and explanation.
Gerstenberg, T., & Tenenbaum, J. (2017). Intuitive theories. In M. Waldmannn (Ed.), *Oxford handbook of causal reasoning* (pp. 515–548). Oxford University Press.
Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 1–31.
Günther, M. (2017). Disjunctive antecedents for causal models. *Proceedings of the 21st Amsterdam Colloquium*.
Hájek, A. (2021). Counterfactual scepticism and antecedent-contextualism. *Synthese, 199*, 637–659.
Hall, N. (2004). Two concepts of causation. In Collins & Hall (Eds.), *Counterfactuals and causation* (pp. 225–276). Cambridge: MIT Press.
Hall, N. (2007). Structural equations and causation. *Philosophical Studies, 132*(1), 109–136.
Halpern, J. (2016). *Actual causality*. Cambridge, MA: MIT Press.
Halpern, J. (2000). Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research, 12*, 317–337.
Halpern, J., & Pearl, J. (2005). Causes and explanations: A structural-model approach, part I: Causes. *British Journal for the Philosophy of Science, 56*(4), 843–87.

Halpern, J., & Hitchcock, C. (2015). Graded causation and defaults. *British Journal for the Philosophy of Science, 66*(2), 413–457.
Hiddleston, E. (2005). A causal theory of counterfactuals. *Noûs, 39*, 632–657.
Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy, 98*(6), 273–299.
Hitchcock, C. (2016). Conditioning, intervening, and decision. *Synthese, 193*(4), 1157–1176.
Huber, F. (2013). Structural equations and beyond. *Review of Symbolic Logic, 6*(4), 709–732.
Knight, F. (1921). *Risk, uncertainty, and profit*. Boston: Houghton Mifflin Company.
Korb, K., Hope, L., Nicholson, A., & Axnick, K. (2004). Varieties of causal intervention. In: Zhang, C. & Werner H. (Eds.), *Pacific rim conference on artificial intelligence*, New York, (pp. 322–31).
Lewis, D. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.
Lewis, D. (1986). *Events. In his Philosophical Papers* (Vol. II, pp. 241–69). New York: Oxford University Press.
Meek, C., & Glymour, C. (1994). Conditioning and intervening. *British Journal for the Philosophy of Science, 45*(4), 1001–1021.
Menzies, Peter, & Price, Huw. (1993). Causation as a secondary quality. *British Journal for the Philosophy of Science, 44*(2), 187–203.
McDonald, J. (2023, Forthcoming). Essential structure for causal models. *Australasian Journal of Philosophy*.
Norton, J. (2021). Eternal inflation: When probabilities fail. *Synthese, 198*, 3853–3875.
Norton, John. (2003). Causation as folk science. *Philosophers'. Imprint, 3*, 1–22.
Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge: Cambridge University Press.
Pearl, J. (2011). The algorithmization of counterfactuals. *Annals of Mathematics and Artificial Intelligence, 61*(1), 29–39.
Quine, W. (1982). *Methods of Logic*. Cambridge, MA: Harvard University Press.
Rosella, G., & Sprenger, J. (2022). Causal modeling semantics for counterfactuals with disjunctive antecedents. Draft. https://philpapers.org/archive/ROSCMS-2.pdf.
Ross, L., & Woodward, J. (2022). Irreversible (one-hit) and reversible (sustaining) causation. *Philosophy of Science, 89*(5), 889–898.
Sartorio, C. (2006). Disjunctive causes. *Journal of Philosophy, 103*(10), 521–538.
Schaffer, J. (2004). Counterfactuals, causal independence and conceptual circularity. *Analysis, 64*(4), 299–308.
Stalnaker, R. (1968). A theory of conditionals. In Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Oxford: Basil Blackwell.
Stalnaker, R. (1980). A defense of conditional excluded middle. In Harper, Stalnaker, & Pearce (Eds.), *IFS.* (Vol. 15). Dordrecht: Springer.
Twardy, C., & Korb, K. (2011). Actual causation by probabilistic active paths. *Philosophy of Science, 78*, 900–913.
Vandenburgh, J. (2022). Backtracking through interventions: An exogenous intervention model for counterfactual semantics. *Mind & Language*.
Verma, T., & Pearl. J. (1988). Causal Networks: Semantics and Expressiveness. *Proceedings, 4th Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, MN 352-359.
Walters, L. (2009). Morgenbesser's coin and counterfactuals with true components. *Proceedings of the Aristotelian Society, 109*(1), 365–379.
Weslake, B. (2015). A partial theory of actual causation. *British Journal for the Philosophy of Science*.
Williams, J. (2010). Defending conditional excluded middle. *Noûs, 44*(4), 650–668.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
Woodward, J. (2016). The problem of variable choice. *Synthese, 193*, 1047–1072.
Woodward, J. (2021). *Causation with a human face: Normative theory and descriptive psychology*. Oxford: Oxford University Press.
Wysocki, T. (2023a). Conjoined cases. *Synthese*, 201, 197.
Wysocki, T. (2023b). The Underdeterministic Framework. *The British Journal for the Philosophy of Science*. Forthcoming. 10.1086/724450
Wysocki, T. (ms1). Causal decision theory for the probabilistically blinded.
Wysocki, T. (ms2). Underdeterministic causation: a proof of concept.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.