Incautious interventions *

(An)Thony S. Gillies
University of Arizona

Abstract There are open questions about the logical landscape when it comes to comparing *ordering semantics* of counterfactuals and *interventionist semantics* of counterfactuals. This paper offers a somewhat surprising partial answer: the divergences are both more basic and more robust than previously noted.

Keywords: counterfactual conditionals, conditional logic, ordering semantics, truth maker semantics, causal models, internventionist semantics

1 Introduction

In fact, it didn't rain and the seats in the convertible are dry even though the top was down. Even so, (1) is true in many contexts:

(1) If it had rained, the seats would have been wet.

Such a counterfactual temporarily sets aside certain facts (e.g., that it didn't rain), supposes instead that some other facts are true (e.g., that it rained), and then asserts in such a counterfactual context that the seats are wet.

By far the dominant theory for how this works is based on *comparative similarity*, for some contextually determined ordering of similarity. Roughly, the counterfactual in (1) is true at a world w iff the worlds most similar to w (according to contextually determined ordering from w) but in which it rained are all worlds in which the seats are wet.

* This project got started in the Fall 2020 pandemic edition of Sabine Iatridou, Kai von Fintel, and Justin Khoo's joint seminar on conditionals "at" MIT. It was then back burnered for a while (with a brief resurgence when I was cast as critic at the APA Author Meets Critics session on Khoo 2022, see Gillies 2023). The gist of the current paper was presented at PhLiP8 (Tarrytown, November 2023), where it received characteristically helpful feedback (apologies to the attendees for the unfortunate typsesetting choices made on that occassion). Generous support from the Center for the Philosophy of Freedom at the University of Arizona is gratefully acknowledged.

This has proven to be both a robust and general theory. There are truth-preserving maps between seemingly different theories expressed in terms of selected antecedent worlds (Stalnaker & Thomason 1967), systems of spheres (Lewis 1973), premise sets (Kratzer 1981, Veltman 1976), and orderings over worlds (Pollock 1976a).[1] Given the broad equivalence, call this family of theories collectively the *ordering semantics* for counterfactuals.

There is however a seemingly different intuition, and a subsequent family of theories implementing it. According to this different picture, setting aside a fact and hypothetically adopting another fact amounts instead to making reference to specific states of affairs where the counter-fact obtains.[2] Roughly, (1) is true iff the seats being wet is *made true* by any (counterfactual) state "singly referred to" by *it had rained* (Fine 1975: 454). Fleshing this out requires making sense of the idea of an antecedent "singly referring" to various counterfactual states and it requires making sense of the relation of truth-making that holds between a state and a formula or fact or what have you.

One class of implementations of this second intuition has been the development of causal models for counterfactuals where the single reference involved is operationalized as the upshot of *intervening* in a model to change the value of some variable in it. For the case of (1) the variable in question represents that it did not rain. The antecedent connection between states, represented by a systems of equations recording the dependencies in the model, together with the intervention determines whether the consequent is or is not made true and so whether the original (1) is true or false. While other theories fit into this picture to varying degrees, the theoretical framework and implementation of it most apt for comparison with ordering semantics of counterfactuals can be found in Fine 2012 and Briggs 2012 (respectively).[3] Despite the various differences in detail, call this family of theories collectively the *interventionist semantics* of counterfactuals.[4]

There are open questions about the logical landscape when it comes to comparing these two approaches to the semantics of counterfactuals. The

1 See van Fraassen 1974, Lewis 1981, and Veltman 1985.

2 I believe this originates in Fine 1975.

3 This is not least because other prominent interventionist semantics have limited expressive resources. More on that in a minute.

4 Other implementations more or less belonging to this family that I won't discuss in detail here include Kaufmann 2013, Santorio 2019, and Ciardelli et al. 2018. I think, without too much fiddling, most of what I have to say (in spirit if not in letter) carries over more or less to these.

aim here is to offer a somewhat surprising partial answer: to show that the divergences are both more basic and more robust than previously noted. More basic: they diverge not only compared to the relatively powerful logics of $\mathbb{CO}$ (Stalnaker & Thomason) and $\mathbb{VC}$ (Lewis), but already at the weak logic of $\mathbb{P}$, widely taken to be the minimal logic determined by the ordering semantics framework. More robust: there are validities of ordering semantics that on pain of triviality can't be validated by interventionist semantics.

2 A recap of ordering semantics

The logic of conditionals based on orderings or premise sets is very well understood.[5] For concreteness, we focus on a language $\mathbf{L}$, which features the counterfactual conditional operator $\boxrightarrow$, built on top of a *descriptive* propositional language $\mathbf{L_o}$.[6] What I have to say can be said with respect to counterfactuals which are flat (unembedded) but which permit arbitrary descriptive antecedents and consequents.[7]

I'll assume that for every descriptive $A, B, \ldots$ there is a corresponding proposition $\mathbf{A}, \mathbf{B}, \ldots$ of the set of worlds where A is true, B is true, ... and moreover, that $\overline{\mathbf{A}}$ is the proposition that $\neg A$ is true and that $\mathbf{A} \cap \mathbf{B}$ and $\mathbf{A} \cup \mathbf{B}$ are the propositions that $A \wedge B$ and $A \vee B$ are true (respectively).

Ordering semantics assumes that for every world w there is an associated (contextually determined) ordering $\leq_w$ that is at a minimum *reflexive* and *transitive*.

Definition 1 (Ordering truth conditions)**.** $A \boxrightarrow C$ is true at w iff the most similar worlds (according to $\leq_w$) in $\mathbf{A}$ to w are worlds in $\mathbf{C}$.

Equivalently: $A \boxrightarrow C$ is true at w (according to $\leq_w$) iff there are no worlds in $\mathbf{A} \cap \overline{\mathbf{C}}$ that are more similar to w than every world in $\mathbf{A} \cap \mathbf{C}$.

This setup lends itself to some nice pictures. In fact, Bizet was French and Verdi was Italian. But consider in addition to this three counterfactual possibilities in which they were compatriots: x in which they were both French, y in which they were both Italian, and z in which they were both Austrian. We might represent relative similarities between these possibilities

5 I am happy to slip back and forth between talking about a *semantics* and the *logic* which it generates. No harm comes from this in the current context and it eases exposition.

6 Official definitions will, for the most part, be relegated to the appendices.

7 This in fact is not much of a restriction: even in a language allowing arbitrary embeddings, the rich logics $\mathbb{CO}$ and $\mathbb{VC}$ put almost no constraints on the behavior of iterated counterfactuals.

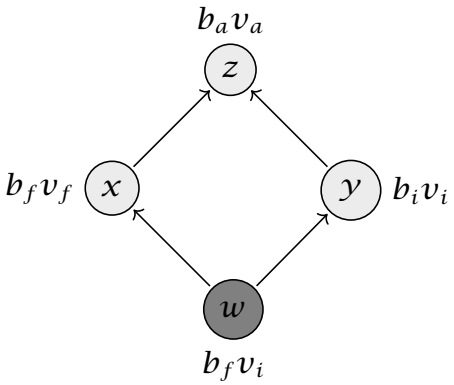


Figure 1 Bizet and Verdi

as in Figure 1 where $u \rightarrow v$ between worlds u and v iff $u \leq_w v$. Consider the counterfactuals (2a) and (2b):

(2) a. If Bizet and Verdi had been compatriots, they would have both been French. $c \mathrel{\square\!\!\rightarrow} f$

b. If Bizet and Verdi had been compatriots, they would have both been Italian. $c \mathrel{\square\!\!\rightarrow} i$

Given this ordering, neither of these are true. No world in $\mathbf{c} \cap \mathbf{f}$ is more similar to w than every world in $\mathbf{c} \cap \bar{\mathbf{f}}$ because $x \not\leq_w y$ and so (2a) is false. No world in $\mathbf{c} \cap \mathbf{i}$ is more similar to w than every world in $\mathbf{c} \cap \bar{\mathbf{i}}$ because $y \not\leq_w x$ and so (2b) is false.

If in addition $\leq_w$ is also *connected*[8] then we can represent $\leq_w$ by way of a system of spheres (Lewis 1973). For example, in Figure 2 worlds are distributed in nested spheres around w: v belongs to every sphere u belongs to iff $v \leq_w u$. Since the closest sphere which overlaps with the proposition that Bizet and Verdi were compatriots has a mix of both-French and not-both-French worlds, (2a) is false.

The minimal ordering semantics and truth conditions above is what is common to ordering semantics once we set aside disputes about the specific

8 For every v, u: either $v \leq_w u$ or $u \leq_w v$. It is worth noting that while there is a difference between two worlds being tied in an ordering and two worlds being incomparable in an ordering, that difference is not visible to the truth conditions ordering semantics gives for counterfactuals.

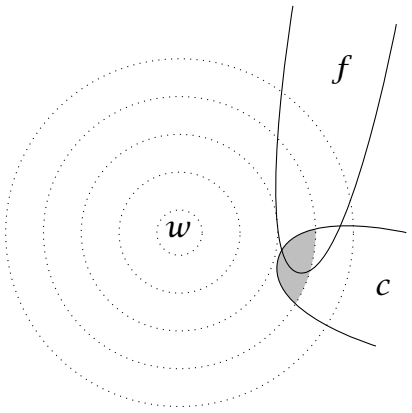


Figure 2 Bizet and Verdi in spheres

properties the orderings must have: (weakly) centered versus not, uniqueness versus ties, ties versus quasi-connectedness, quasi-connectedness versus incomparabilities, and all the rest. Famously, it is exactly characterized by the logic $\mathbb{P}$.[9]

Definition 2 (Containing $\mathbb{P}$)**.** An entailment relation $\models_{\mathbb{S}}$ determined by semantics $\mathbb{S}$ is at least as strong as $\models_{\mathbb{P}}$ iff:

(ID) $\models_{\mathbb{S}} A \boxrightarrow A$;

(RW) $A \boxrightarrow C \models_{\mathbb{S}} A \boxrightarrow C'$ if C entails C';

(OR) $A \boxrightarrow C, B \boxrightarrow C \models_{\mathbb{S}} (A \vee B) \boxrightarrow C$;

(RM) $A \boxrightarrow C, A \boxrightarrow B \models_{\mathbb{S}} A \boxrightarrow (B \wedge C)$;

(CM) $A \boxrightarrow C, A \boxrightarrow B \models_{\mathbb{S}} (A \wedge B) \boxrightarrow C$; and

(LE) $A \boxrightarrow C \models_{\mathbb{S}} A' \boxrightarrow C'$ if $A \equiv A'$ and $C \equiv C'$.

If $\models_{\mathbb{S}}$ is at least as strong as $\models_{\mathbb{P}}$ then $\mathbb{P}$ is included in $\mathbb{S}$.[10]

9 The axiomatization below is due to Burgess 1981 where the logic is referred to as $\mathbb{B}$. Modern convention (following Krauss et al. 1990) refers to it as $\mathbb{P}$ instead to avoid confusion with the modal logic **B**(rouwer).

10 Here entailment between descriptive sentences is classical and equivalence is co-entailment.

Both $\mathbb{CO}$ and $\mathbb{VC}$ strictly include $\mathbb{P}$.

Unsurprisingly absent as a validity of $\mathbb{P}$ is antecedent strengthening — this is top of the list of things ordering semantics is designed to avoid. For current purposes we say that a counterfactual logic is trivial if it validates antecedent strengthening.

Definition 3. A counterfactual logic $\mathbb{S}$ is trivial if it validates (AS):

(AS) $A \boxright C \models (A \wedge B) \boxright C$

As desired, none of the logics based on ordering semantics validate (AS). This of course is the bare minimum when it comes to counterfactuals:

(3) a. If it had rained, the seats would have been wet. $r \boxright w$

b. If it had rained and the car had been parked in the garage, the seats would have been wet. $(r \wedge g) \boxright w$

Whatever else is contested, that $r \boxright w$ does not entail $(r \wedge g) \boxright w$ isn't.[11]

A punchline below is that a characteristic validity of $\mathbb{P}$, namely (CM), is not valid in interventionist semantics and that furthermore the interventionist semantics can't be amended to validate it without also validating (AS).

3 A recap of causal models

Interventionism is built on the idea that counterfactual antecedents refer to states of affairs that exactly correspond to ways of making those antecedents true. We will, as suggested in Fine 2012 and developed in Briggs 2012, take the relevant interventions to be interventions on a causal model representing the contextually relevant facts and dependencies between them. The interventions in turn determine the relevant states of affairs that serve as referents for counterfactual antecedents.

Causal models consist of a set of atomic states ("variables") together with a set of dependencies ("structural equations") and an assignment of values to

11 Note that dynamic strict conditional theories (such as those in von Fintel 2001 and Gillies 2007) are not trivial in this sense either: those theories are designed to treat reverse Sobel sequences like $(r \wedge g) \boxright \neg w, r \boxright w$ as (dynamically) inconsistent without licensing a dynamic entailment from $r \boxright w$ to $(r \wedge g) \boxright w$.

(exogenous) variables.[12] The dependencies in the structural equations then propagate values to the rest of the variables.[13]

Typically we will take the variables $\mathbf{V}$ in a causal model to be some subset of atomic formulas. The values of exogenous variables is determined by an assignment: a function v mapping each exogenous variable to a truth value. This is like setting the initial conditions. Finally, causal models associate each endogenous variable with a value, given the values of the other variables. This is achieved through a system $\mathbf{F}$ of structural equations.

Some examples will help. Suppose a light c is wired to two switches, a and b. The light is on if either a is up or b is up and it is off otherwise. As a matter of fact, a is up and b is down and c is on. Making the obvious choices about $\mathbf{V}$ we get the following:

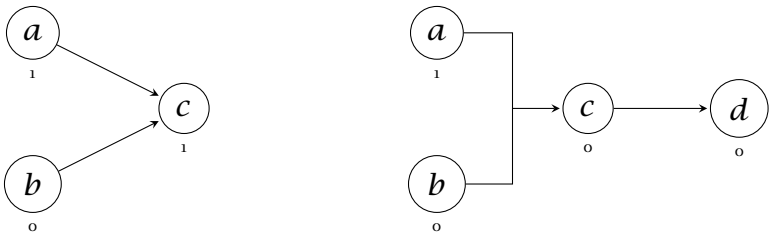
(4) $\mathbf{M} = \langle \mathbf{V}, \mathbf{F}, v \rangle$:

$$\begin{aligned} \mathbf{V_X} &= \{a, b\} \\ \mathbf{V_E} &= \{c\} \\ \mathbf{F} &= \{\langle c, v(c) = \max(v(a), v(b))\rangle\} \\ v &= \{\langle a, 1\rangle, \langle b, 0\rangle\} \end{aligned}$$

Note that c is the lone endogenous variable and its value is determined by the lone function f_c in $\mathbf{F}$, which stipulates that $v(c) = \max(v(a), v(b))$. You'll recognize that as boolean code for the classical truth conditions for $\vee$. Hence since a is up ($v(a) = 1$) and b is down ($v(b) = 0$), we know that light c is on

12 For a thorough introduction see Pearl 2009. Both Galles & Pearl 1998 and Halpern 2013 report some partial results on the relationship between ordering semantics (in particular, some of the $\mathbb{V}$ logics) and interventionism: they are reported to be (roughly) equivalent. However, the underlying language used does not permit interesting counterfactual antecedents like $a \vee b$ and $\neg(a \wedge b)$. In the context of a more expressive language like $\mathbf{L}$ these results are arguably misleading. Both ordering semantics and interventionism are characterized by entailment patterns that essentially involve complex boolean counterfactual antecedents: (OR) and (instances of) (CM) for ordering semantics, and simplification of disjunctive antecedents (SDA) for interventionism. More on (SDA) shortly.

13 Usually, atomic states are represented by simple equations like $X = x$, saying that the variable X in the model has value x, where x can be just about anything (a probability, an objective chance, a truth value, etc.). For our purposes, this is more generality than is needed: atomic states will here correspond to atomic formulas and so truth values as values. Thus rather than writing the cumbersome $\mathsf{a} = 1$ or $\mathsf{a} = 0$ to represent that the atomic formula a is true or false, we will opt to generally write the more familiar a to represent the former and $\neg a$ to represent the latter.



(a) a or b (b) a and b

Figure 3 Switches

($v(c) = 1$). This is represented in the graph of the causal model in Figure 3(a): the nodes represent the variables and the (independent) arrows from a to c and from b to c together represent the structural equation.

But what if the set-up were different, say, that c is on only if both a is up and b is up? And what if there is a second light d that goes on whenever c is on? In that case we would have:

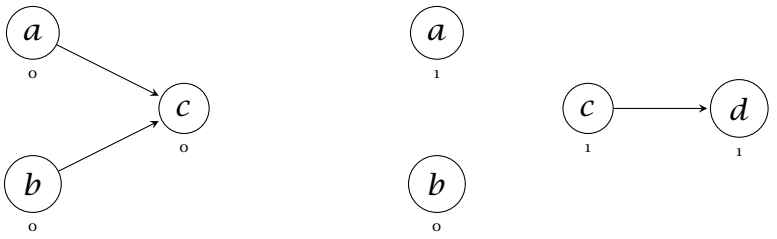
(5) $\mathbf{M}' = \langle \mathbf{V}, \mathbf{F}', v \rangle$

$$\begin{aligned} \mathbf{V_X} &= \{a, b\} \\ \mathbf{V_E} &= \{c, d\} \\ \mathbf{F}' &= \{\langle c, v(c) = \min(v(a), v(b))\rangle, \langle d, v(d) = v(c)\rangle\} \\ v &= \{\langle a, 1\rangle, \langle b, 0\rangle\} \end{aligned}$$

In $\mathbf{M}'$, notice, $v(c)$ is the smallest value of the values of a and b. You'll recognize that as boolean code for classical $\wedge$ and so in this model c is true iff both a and b are true. This is represented in Figure 3(b) by the joint arrow from a and b to c. And since d is completely determined by c, light d is also off.[14]

Intervening on a model $\mathbf{M}$ is a matter of setting a particular variable to a particular value, and then seeing how that changes the values of other variables given the dependencies in the model. Let's describe the basic operation of intervention on specific variables here, and then extend that to complex interventions below.

14 Truth for arbitrary boolean formulas in a model is defined in the expected (classical) way.



(a) Intervening on a

(b) Intervening on c

Figure 4 Interventions on switches

To perform an x-intervention on $\mathbf{M}$, we need to know if x is exogenous or endogenous since x's value is set in different ways in each case. The idea is simple: find the reason x is true, and stipulate that $v(x) = 1$ (or $v(x) = 0$, as the case may be). The resulting model is $[\mathbf{M}]^x$.

For endogenous x, interventions replace the old structural equation with the equation that stipulates that x is true (or false, as the case may be). For an exogenous x, interventions replace the bit of the old valuation about x with a new bit that stipulates that x is true (or false, as the case may be). Finally, performing an intervention on a conjunction of literals reduces to a sequence of interventions on the conjuncts.

Again some examples will help. Consider an intervention to make a false in the model in Figure 3(a). Since a is exogenous this leaves $\mathbf{F}$ unchanged but alters the assignment: in $[v]^{\bar{a}}$ we find $\langle a, 0\rangle$. Since it is still the case that $v(c) = \min(v(a), v(b))$ and both a and b are false in $[\mathbf{M}]^{\bar{a}}$, the result is that c is false post-intervention (Figure 4(a)).

Next, consider an intervention that changes the structural equations from (3(b)) as in (6):

(6) $\quad [\mathbf{F}']^c = \{\langle c, v(c) = 1\rangle, \langle d, v(d) = v(c)\rangle\}$

Here we are considering what would have been the case had c been on. The equation that ties d's value to c's is still present but there is no longer a

connection between a and b and whether c is on and that dependency is eliminated from the graph of the situation in Figure 4(b).[15]

4 A recap of truthmaker semantics

In some implementations, a counterfactual $A \boxdot\!\!\rightarrow C$ invites an intervention to change a model so that A is the case, and the counterfactual is true if C is true in the post-intervention model. This would seem to require: (i) interventions defined for more complex expressions than (conjunctions of) literals and (ii) some guarantee that, no matter A's complexity, there is always a unique post-intervention model. A key advantage of the truthmaker implementation of interventionist semantics is that it sees no reason to insist on (ii) and therefore no reason to pursue (i). Instead, $A \boxdot\!\!\rightarrow C$ invites us to collect *all* post-intervention models that exactly make true or refer to A states and see whether they all support or inexactly make true C.

The first step is to say how possible interventions create a space of states based on a given model.[16]

Definition 4 (State spaces)**.** A state space based on causal model $\mathbf{M}$ is the set $\mathbf{U}$ of states obtainable by intervening on $\mathbf{M}$ with a consistent formula A together with a partial order $\sqsubseteq$ where $\mathbf{s} \sqcup \mathbf{t} = \mathbf{t}$ iff $\mathbf{s} \sqsubseteq \mathbf{t}$. Whenever $\mathbf{s}$ and $\mathbf{t}$ are states in $\mathbf{S_M}$ so is $\mathbf{s} \sqcup \mathbf{t}$.[17]

The state $\mathbf{s} \sqcup \mathbf{t}$ is the *fusion* of states $\mathbf{s}$ and $\mathbf{t}$: the state of it being windy and cold ($\mathbf{s} \sqcup \mathbf{t}$) that is comprised of the state of it being windy ($\mathbf{s}$) fused together with the state of it being cold ($\mathbf{t}$).

Again, an example will help. Take the model $\mathbf{M}$ represented in Figure 4(a) and consider $[\mathbf{M}]^{a \vee b}$. Three states ("submodels") are then relevant, none of which alter any of the structural equations, and all of which are fusions of atomic states (or their negations): one in which a is up and b is down (and c is on), one in which b is up and a is down (and c is up), and one in which they are both up (and c is up).

15 Without amendments the framework is not well-suited to backtracking counterfactuals, which we set aside here.

16 These states are sometimes called "submodels" of the starting model. This is somewhat misleading since the "submodels" typically assign different truth values to (for instance) the atomic formulas than the origin model does and so are not your logician's submodels at all.

17 We also assume that $[\mathbf{M}]^{A \wedge B}$ is $[\mathbf{M}]^{A} \sqcup [\mathbf{M}]^{B}$.

The cornerstone concept in truthmaker semantics is when a state verifies (or falsifies) a formula. This is non-classical (allowing gaps) and hyperintensional (classically equivalent formulas can have different sets of verifiers). For our purposes, the verification clauses for disjunction and conjunction are most relevant.[18]

A disjunction $A \vee B$ is verified by a state $\mathbf{s}$ iff $\mathbf{s}$ verifies A ($\mathbf{s} \Vdash A$) or $\mathbf{s}$ verifies B ($\mathbf{s} \Vdash B$) or both. Dually: $\mathbf{s}$ falsifies $A \vee B$ iff it falsifies both A and B.

A conjunction $A \wedge B$ on the other hand is verified by a state $\mathbf{s}$ iff $\mathbf{s}$ is the fusion of states $\mathbf{t}$ and $\mathbf{r}$ that in turn verify the conjuncts separately. That is: $\mathbf{s} \Vdash A \wedge B$ iff there are states $\mathbf{t}$ and $\mathbf{r}$ such that $\mathbf{t} \Vdash A$ and $\mathbf{t} \Vdash B$ where $\mathbf{t} \sqcup \mathbf{r} = \mathbf{s}$. Dually: $\mathbf{s}$ falsifies $A \wedge B$ iff it falsifies either A or B or both.

Finally, $\mathbb{TM}$ says that a counterfactual is true iff its consequent is true at every post-intervention state that verifies its antecedent.

Definition 5 (Truth maker truth conditions ($\mathbb{TM}$))**.** Given a causal model $\mathbf{M}$ and counterfactual $A \boxdot\!\!\rightarrow C$:

$$\mathbf{M} \models_{\mathbb{TM}} A \boxdot\!\!\rightarrow C \text{ iff for every } \mathbf{s} \in [\mathbf{M}]^A \text{ there is a } \mathbf{t} \sqsubseteq \mathbf{s} \text{ such that } \mathbf{t} \Vdash C$$

In truthmaker lingo: a counterfactual is true iff every way of exactly making the antecedent true contains a way of making the consequent true.

The straightforward prediction in the example in Figure 4(a) is that every state in $[\mathbf{M}]^{a \vee b}$ contains a state in which the light is on. Hence (7) is true:

(7) If a or b had been up, c would have been on.

This is of course no accident since interventionism is designed explicitly to verify (SDA).

Observation 1 (Fine 2012)**.** $\mathbb{TM}$ validates

(SDA) $(A \vee B) \boxdot\!\!\rightarrow C \models A \boxdot\!\!\rightarrow C$

On the other hand, it is of course well known that $\mathbb{P}$ does not validate (SDA): the nearest worlds in $\mathbf{A} \cup \mathbf{B}$ to w may well all be worlds in $\overline{\mathbf{A}} \cap \mathbf{B}$. In that case it can happen that the nearest worlds in $\mathbf{A} \cup \mathbf{B}$ to w are in $\mathbf{C}$ (and so $(A \vee B) \boxdot\!\!\rightarrow C$ is true at w) while some of the nearest worlds in $\mathbf{A}$ to w are not in $\mathbf{C}$ (and so $A \boxdot\!\!\rightarrow C$ is false at w). So $\mathbb{P}$ does not contain $\mathbb{TM}$.

Of course $\mathbb{TM}$ does not validate (AS).

18 A full recursive definition of both verification and falsification is in Definition 18 in Appendix B.

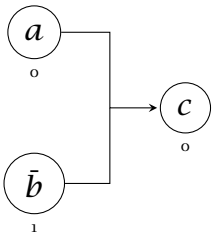


Figure 5 A useful countermodel

Proof. A simple counterexample:

(8) $\mathbf{M} = \langle \mathbf{V}, \mathbf{F}, v \rangle$:

$$\begin{aligned} \mathbf{V_X} &= \{a, b\} \\ \mathbf{V_E} &= \{c\} \\ \mathbf{F} &= \{\langle c, v(c) = v(a)[1 - v(b)]\rangle\} \\ v &= \{\langle a, 0\rangle, \langle b, 0\rangle\} \end{aligned}$$

□

Here, for the light to be on, switch a must be up and switch b must be down. In fact, both switches are down and so the light is off. Still (9) is true.

(9) If a had been up, c would have been on.

However $(a \wedge b) \boxdot\!\!\rightarrow c$ is not true in $\mathbf{M}$: at no state in $[\mathbf{M}]^{a \wedge b}$ is the light on.

Observation 2. $\mathbb{TM}$ does not validate (AS).

Proof. Figure 5. □

Since, in the presence of (LE), (SDA) entails (AS) it follows that $\mathbb{TM}$ does not validate (LE). The key is that in general A and $(A \wedge B) \vee (A \wedge \neg B)$ have different truthmakers even though they are (classically) equivalent.[19]

Observation 3. $\mathbb{TM}$ does not validate (LE).

19 Those in the $\mathbb{TM}$ world see hyperintensionalism is a feature not a bug.

Proof. The model in (8) serves as a counterexample here, too: as we saw, $a \boxrightarrow c$ is true while $(a \wedge b) \boxrightarrow c$ is false. And yet $(a \wedge b) \boxrightarrow c$ is entailed (by SDA) by $((a \wedge \neg b) \vee (a \wedge b)) \boxrightarrow c$ and so the latter must be false as well. □

5 A counterexample to (CM)

$\mathbb{TM}$ and $\mathbb{P}$ also diverge in ways that are logically independent of (LE) and issues of hyperintensionality. In particular, while $\mathbb{P}$ validates (CM), $\mathbb{TM}$ does not.[20]

Two other validities are relevant to seing this. As noted by Fine, both (ID) and (RW) are $\mathbb{TM}$-valid.

Observation 4. $\mathbb{TM}$ validates both (ID) and (RW).

Observation 5. $\mathbb{TM}$ does not validate (CM).

Proof. We note that the model (8), pictured in Figure 5, is once again a countermodel. First, as we have seen, $\mathbf{M} \models a \boxrightarrow c$. Now, since $a \boxrightarrow a$ by (ID), it follows that $a \boxrightarrow (a \vee b)$ by (RW). To see that $\mathbf{M}$ is a countermodel to (CM) we now show:

$$\mathbf{M} \not\models (a \wedge (a \vee b)) \boxrightarrow c$$

Suppose otherwise. Note that $a \wedge (a \vee b)$ and $(a \wedge a) \vee (a \wedge b)$ have the same set of truthmakers and therefore:

$$(a \wedge (a \vee b)) \boxrightarrow c \models ((a \wedge a) \vee (a \wedge b)) \boxrightarrow c$$

Hence, by (SDA):

$$(a \wedge (a \vee b)) \boxrightarrow c \models (a \wedge b) \boxrightarrow c$$

But clearly $\mathbf{M} \not\models (a \wedge b) \boxrightarrow c$ and hence $\mathbf{M} \not\models (a \wedge (a \vee b)) \boxrightarrow c$. □

6 Incompatibility and triviality

Above I asserted that (CM) and (LE) are logically independent. They are, but you don't have to take my word for it.

20 The counterfactual developed in Khoo 2022 is a sort of hybrid conditional which blends elements of interventionism with elements of the more traditional theories of conditional modality in Kratzer 1991. That conditional also (plausibly) invalidates (CM) but for somewhat different reasons: it is not a hyperintensional conditional, it does not validate (SDA), but it also invalidates substitution of conditional equivalents. See Gillies 2023.

Observation 6. (CM) does not imply (LE).

Proof. It is possible for a conditional to validate (CM) but not validate (LE). For instance, the conditional $\rightarrow$ of intuitionistic logic ($\mathbb{Int}$) is monotonic and therefore (CM) is $\mathbb{Int}$-derivable:

$$\dfrac{\dfrac{\dfrac{a \wedge b \vdash a \wedge b}{a \wedge b \vdash a}\wedge\text{E} \qquad a \rightarrow c \vdash a \rightarrow c}{a \wedge b, a \rightarrow c \vdash c}\rightarrow\text{E}}{a \rightarrow c \vdash (a \wedge b) \rightarrow c}\rightarrow\text{I} \qquad a \rightarrow b \vdash a \rightarrow b}{a \rightarrow c, a \rightarrow b \vdash ((a \wedge b) \rightarrow c) \wedge (a \rightarrow b)}\wedge\text{I}$$
$$\dfrac{a \rightarrow c, a \rightarrow b \vdash ((a \wedge b) \rightarrow c) \wedge (a \rightarrow b)}{a \rightarrow c, a \rightarrow b \vdash (a \wedge b) \rightarrow c}\wedge\text{E}$$

But (owing to the lack of total DeMorgan's equivalences) $\neg(a \wedge b) \rightarrow c$ is not equivalent with $(\neg a \vee \neg b) \rightarrow c$: there is a sketch of a countermodel in Figure 6. In this model: (i) $\neg(a \wedge b)$ is true at v while c is false there and so $\neg(a \wedge b) \rightarrow c$ is false at w; but (ii) $\neg a \vee \neg b$ is not true at v and hence no world downstream of w can falsify $(\neg a \vee \neg b) \rightarrow c$ at w and so it is true and hence (CM) does not entail (LE).[21] Hence (CM) does not imply (LE).[22] □

Observation 7. (LE) does not imply (CM).

Proof. It is possible for a conditional to validate (LE) without validating (CM). For instance, consider a conditional $\Rightarrow$ interpreted at w with respect to two sequences of worlds $\overrightarrow{w}$ and $\overleftarrow{w}$, where $\overrightarrow{w}_1$ is the first world in the sequence $\overrightarrow{w}$ and $\overleftarrow{w}_1$ is the first world in the sequence $\overleftarrow{w}$. Say that $A \Rightarrow C$ is true at w iff either $\overrightarrow{w}_1$ is an $A \wedge C$-world or $\overleftarrow{w}_1$ is an $A \wedge C$-world. Clearly, such a conditional is intensional and not hyperintensional and so validates (LE). But it does not validate (CM). Figure 7 depicts a countermodel where: (i) $\overrightarrow{w}_1$ is an $a \wedge c$-world that is not a b-world; and (ii) $\overleftarrow{w}_1$ is an $a \wedge b$-world that is not a c-world. Hence while $a \Rightarrow c$ and $a \Rightarrow b$ are true at w, neither $(a \wedge b) \Rightarrow c$ nor $(a \wedge c) \Rightarrow b$ is. □

21 The key intuitionist truth conditions here are for negation: for $\neg A$ to be true at w, no successor state to w can verify A.

22 For another example, see Berto & Özgün 2021 where the conditional validates (CM) without validating either (AS) or (LE).

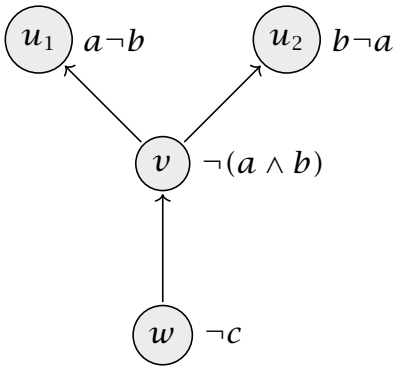


Figure 6 $\mathbb{Int}$ counterexample to (LE)

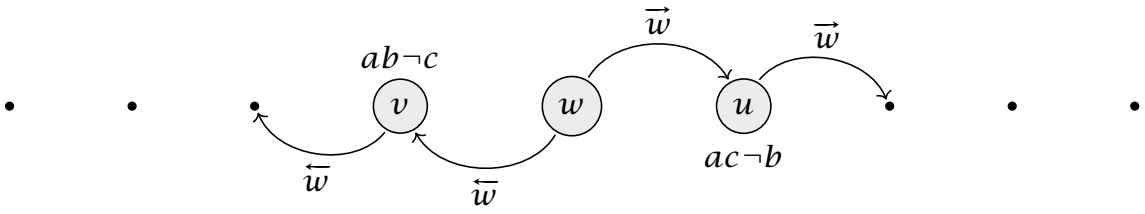


Figure 7 Counterexample to (CM)

So $\mathbb{TM}$ and $\mathbb{P}$ are fundamentally incomparable: each invalidates a core pattern the other validates. But the differences run deeper: neither can be amended, on pain of triviality, to validate its missing entailment pattern. This is of course well-known for $\mathbb{P}$ since (LE) and (SDA) jointly entail (AS). This is mirrored for $\mathbb{TM}$: (SDA) and (CM), with minimal additional assumptions, jointly entail (AS).

Most theories take $\wedge$ to expresses a meet and $\vee$ a join in a distributive lattice over descriptive contents. $\mathbb{TM}$ falls into this category[23], but so do other non-classical semantics such as inquisitive semantics.[24] For present purposes a slightly weaker assumption, also clearly validated by $\mathbb{TM}$, will do.

Definition 6. A counterfactual logic $\mathbb{S}$ is distributive iff it validates

(DS) $(A \wedge (B \vee D)) \boxrightarrow C \models ((A \wedge B) \vee (A \wedge D)) \boxrightarrow C$

Observation 8. No nontrivial semantics that validates (RW), (DS), and (SDA) also validates (CM).

Proof. Suppose otherwise and suppose $a \boxrightarrow c$. By (RW) we have for any b that $a \boxrightarrow (b \vee c)$ and hence by (CM) that $(a \wedge (b \vee c)) \boxrightarrow c$. By (DS) it then follows that $((a \wedge b) \vee (a \wedge c)) \boxrightarrow c$. Thus, by (SDA), $(a \wedge b) \boxrightarrow c$ and so $\models$ validates (AS) and so is trivial. □

So, sure, the interventionist worldview is at odds with the ordering worldview because the ordering worldview is intensional and the interventionist worldview is hyperintensional. This is revealed in the incompatibility between (SDA) and (LE): the price of (SDA) is hyperintensionality. Or put in more a $\mathbb{TM}$-friendly way: the price of intensionality is invalidating (SDA). This has certainly been Fine's focus. And, from a certain point of view, this does look like a big difference.

But, from another point of view, whether counterfactuals with a as their antecedents are equivalent to corresponding counterfactuals with $(a \wedge b) \vee (a \wedge \neg b)$ as their antecedents seems like pretty high church discourse. Or, at any rate, it is natural to wonder if intensionality-vs-hyperintensionality

23 Note that $\mathbf{s} \Vdash a \wedge (b \vee c)$ iff for some $\mathbf{t}, \mathbf{r}$: $\mathbf{s} = \mathbf{t} \sqcup \mathbf{r}$ s.t. $\mathbf{t} \Vdash a$ and $\mathbf{r} \Vdash b \vee c$. That is, iff: $\mathbf{t} \Vdash a$ and $\mathbf{r} \Vdash b$, or $\mathbf{t} \Vdash a$ and $\mathbf{r} \Vdash c$ and so iff $\mathbf{s} \Vdash a \wedge b$ or $\mathbf{s} \Vdash a \wedge c$. That is: iff $\mathbf{s} \Vdash (a \wedge b) \vee (a \wedge c)$.

24 Connectives in inquisitive semantics correspond to operations on a Heyting algebra (so the divergences from classical operators emerge with negation); see Ciardelli et al. 2019.

is really The Big Difference (capital letters) here. So it is significant that there is another, elementary incompatibility between the two worldviews, the incompatibility between (SDA) and (CM), that doesn't have anything at all to do with hyperintenionality as such. As it turns out, (SDA) is incompatible with a great many things ($n \geq 2$).

There are of course other assumptions involved here — (DS) and (RW) — and so maybe that is where the real tension lies and so maybe jettisoning one of those opens the door to extending existing implementations of interventionist semantics, in some as yet unknown way, to make them validate (CM) after all.

There is, however, precious little independent reason, from either a truthmaker perspective or any other, for denying distributivity for descriptive sentences, much less the weaker (DS). What state could make-true *It's windy and either rainy or snowy* without making-true *Either it's windy and rainy or it's windy and snowy*? None comes to mind.

The road to invalidating (RW) is similarly steep and rocky. Compare:

(10) a. If I had gone to office hours, I would have been on campus;

b. that's because if I had gone to office hours, I would have been in the department.

(11) a. If I had gone to office hours, I might not have been on campus;

b. ??even though if I had gone to office hours, I would have been in the department.

The discourse in (10) seems like a good, if a little pedantic, explanation as predicted by the validity of (RW). But the discourse in (11) is coherent only if the department is somehow not on campus. This pattern is backwards from what would be unexpected if (RW) were invalid. The pattern is the same with explicit disjunction:

(12) a. If I had gone to office hours, I would have been either in the department or in the library;

b. that's because if I had gone to office hours, I would have been in the department.

(13) a. If I had gone to office hours, I might not have been either in the department or in the library;

b. ??even though if I had gone to office hours, I would have been in the department.

More generally, if (RW) weren't valid, we would have a witnessing true counterfactual $a \boxrightarrow c$ and some b such that $a \boxrightarrow b$ is false even though c entails b. Assuming that $\neg(a \boxrightarrow b)$ entails the *might*-counterfactual $a \diamondrightarrow \neg b$, we would then expect to find coherent discourses that assert $a \diamondrightarrow \neg b$ while maintaining that $a \boxrightarrow c$. Such discourses do not seem forthcoming.

7 Conclusion

Interventionist semantics gives rise to logics which are fundamentally incomparable to and incompatible with even the weakest counterfactual logic determined by ordering semantics. My main interest has been in making this point without saying anything about whether these differences provide reasons to favor one worldview about counterfactuals over the other.

I will close by venturing some initial thoughts about that. On the face of it, (CM) has a lot to be said for it.

(14) a. If it had rained, it would have poured. $a \boxrightarrow b$

b. If it had rained, the seats would have been wet. $a \boxrightarrow c$

c. ??But that doesn't mean that if it had rained and it had poured the seats would have been wet. *maybe* $(\neg (a \wedge b) \boxrightarrow c)$

Obviously no number of positive instances is enough to establish that (CM) is valid. But the general vibe of them can't be denied either. How could it be that some of the counterfactual consequences of A couldn't also without interfering be used as counterfactual lemmas together with A?

More to the point for interventionism, is Fine's own reason for adopting what he calls "Conjunction":

(CON) $A \boxrightarrow C_1, A \boxrightarrow C_2, A \boxrightarrow C_3, \ldots \models A \boxrightarrow C_1 \wedge C_2 \wedge C_3 \wedge \ldots$

This is the closure principle in Pollock 1976b that generalizes (RM). In ordering semantics (CON) is equivalent to the Limit Assumption. In the absence of

(CON) there can be "paradoxical" assumptions: if there are ever-closer A-worlds then all of the things that would be the case if A had been the case can't all be the case.[25] Against Lewis's rejection of the Limit Assumption and hence (CON) Fine writes:

> I am deciding between bringing it about that A or bringing it about that B. To this end, I consider the counterfactual consequences of bringing about the one or the other and then make my decision on the basis of a comparison of the consequences. But what if one or both of the counterfactual suppositions that I bring about A and that I bring about B are paradoxical? How then can I compare them, given that I can form no coherent conception of what the consequences of one or both of them are? (Fine 2012: 227)

The idea is that in weighing what to do I need to form a coherent picture of what would be the case under various counterfactual assumptions. If any of those assumptions turn out to be paradoxical, it is hard to see how I could do this since there is no counterfactual situation in which all the counterfactual consequences of a paradoxical assumption hold.

Just as failures of (CON) create a category of paradoxical assumptions, failures of (CM) create a category of *interfering consequences* of a given assumption A. The trouble, from an interventionist point of view, is that paradoxical assumptions and interfering consequences are similarly confounding. The task of forming a coherent picture of what would be the case had A been the case seems to require some antecedent knowledge of which consequences of A are interfering and which are not. Absent this, how do I know whether or not to count C as part of the picture of what would be the case had A been the case when $A \boxrightarrow C$ is true?

This is made more awkward by the fact that (CON) and therefore (RM) is valid in interventionist semantics. An instance of a (CM) failure would then involve some counterfactual consequence B of A that interferes with $A \boxrightarrow C$ when taken as a counterfactual lemma (because $\neg((A \wedge B) \boxrightarrow C)$) even though it is compatible with $A \boxrightarrow C$ when taken as a counterfactual consequence (because $A \boxrightarrow (B \wedge C)$). So in trying to form a coherent picture of what would have been the case had A been the case, in the absence of (CM) I need to take care to only think of B as a consequence of A and not

25 See Herzberger 1979.

as something that may be coherently added to A as a hypothesis — only in thinking of B in the first way and not in the second way is C part of the coherent situation I am trying to entertain. I doubt I can reliably pull this mental accounting trick off and it's to a theory's relative discredit if it requires me to.

A Appendix: Ordering semantics

Below are the official definitions alluded to in the main body having to do with ordering semantics.

Definition 7. Let **At** be a finite set of atomic formulas $\{a, b, c \ldots\}$. The language $\mathbf{L_o}$ is the smallest set of formulas including **At** that is closed under negation ($\neg$), conjunction ($\wedge$), and disjunction ($\vee$) in the usual way.The language **L** is the smallest set containing $\mathbf{L_o}$ and is such that if $A, C \in \mathbf{L_o}$ then $A \mathrel{\Box\!\!\rightarrow} C \in \mathbf{L}$.

Definition 8 (Ordering models)**.** For a finite set **At** of atomic formulas, an ordering model $\mathbf{M} = \langle W, \langle \leq \rangle \rangle$ for **L** is a pair where:

i. $W = 2^{\mathbf{At}}$; and

ii. $\langle \leq \rangle$ assigns a reflexive, transitive relation $\leq_w$ over W for each $w \in W$.

Definition 9 (Ordering semantics)**.** For any ordering model $\mathbf{M} = \langle W, \langle \leq \rangle \rangle$ and counterfactual $A \mathrel{\Box\!\!\rightarrow} C$ in **L**:

$$\mathbf{M}, w \models_{\mathbb{P}} A \mathrel{\Box\!\!\rightarrow} C \text{ iff } \min(A, w) \subseteq C$$

where $\min(A, w) = \{v \colon\ v \in \mathbf{A} \text{ and there is no } u \in \mathbf{A} \text{ such that } u <_w v\}$.

B Appendix: Causal models and truthmaker semantics

Below are the official definitions alluded to in the main body having to do with causal models and truthmaker semantics.

Definition 10 (Variables)**.** The variables in a causal model is a pair $\mathbf{V} = \langle \mathbf{V_X}, \mathbf{V_E} \rangle$ consisting of a finite set $\mathbf{V_X}$ of exogenous variables and a disjoint finite set $\mathbf{V_E}$ of endogenous variables.

Definition 11 (Assignments)**.** An assignment $v \colon \mathbf{V_X} \to \{0, 1\}$ is a function from the exogenous variables in a model to the set of truth values.

Definition 12 (Structural equations)**.** A function **F** is a structural equation mapping if it assigns each variable $x \in \mathbf{V_E}$ a function f_x determining x's value in terms of the values of the variables in $\mathbf{V} \setminus \{x\}$.[26]

26 Again, abusing notation slightly we sometimes write $v(x)$ to denote x's value in a model, even if x is endogenous.

Definition 13 (Causal models)**.** A causal model is triple $\mathbf{M} = \langle \mathbf{V}, \mathbf{F}, v \rangle$ where $\mathbf{V}$ are the variables in M, $\mathbf{F}$ is a structural equation mapping, and v is an assignment.

Definition 14 (Truth in a model)**.** Let $\mathbf{M} = \langle \mathbf{V}, \mathbf{F}, v \rangle$ be any model where $\mathbf{V} \subseteq \mathbf{At}$, x any variable in $\mathbf{V}$, and A, B any sentences of $\mathbf{L_o}$. The relation $\models_M$ is defined recursively as follows:

i. $\mathbf{M} \models x$ iff $v(x) = 1$

ii. $\mathbf{M} \models \neg A$ iff $\mathbf{M} \not\models A$

iii. $\mathbf{M} \models A \wedge B$ iff $\mathbf{M} \models A$ and $\mathbf{M} \models B$

Definition 15 (Setting truth values)**.** Let $\mathbf{M} = \langle \mathbf{V}, \mathbf{F}, v \rangle$ be any model and let x any variable.

i. $[\mathbf{F}]^x = (\mathbf{F} \setminus \{\langle x, f_x \rangle\}) \cup \{\langle x, v(x) = 1 \rangle\}$.

ii. $[\mathbf{F}]^{\bar{x}} = (\mathbf{F} \setminus \{\langle x, f_x \rangle\}) \cup \{\langle x, v(x) = 0 \rangle\})$.

iii. $[v]^x = (v \setminus \{\langle x, 0 \rangle\}) \cup \{\langle x, 1 \rangle\}$.

iv. $[v]^{\bar{x}} = (v \setminus \{\langle x, 1 \rangle\}) \cup \{\langle x, 0 \rangle\}$.

Definition 16 (Interventions)**.** Let $\mathbf{M} = \langle \mathbf{V}, \mathbf{F}, v \rangle$ be any model and let x any variable in $\mathbf{V}$, and $A_1, \ldots, A_n$ be any literals.

i. If x is endogenous:

   a. $[\mathbf{M}]^x = \langle \mathbf{V}, [\mathbf{F}]^x, v \rangle$;

   b. $[\mathbf{M}]^{\bar{x}} = \langle \mathbf{V}, [\mathbf{F}]^{\bar{x}}, v \rangle$.

ii. If x is exogenous:

   a. $[\mathbf{M}]^x = \langle \mathbf{V}, \mathbf{F}, [v]^x \rangle$;

   b. $[\mathbf{M}]^{\bar{x}} = \langle \mathbf{V}, \mathbf{F}, [v]^{\bar{x}} \rangle$.

iii. $[\mathbf{M}]^{A_1 \wedge \cdots \wedge A_n} = [[[\mathbf{M}]^{A_1}] \cdots]^{A_n}$.

Definition 17 (State space, official)**.** The state space $\mathbf{S_M}$ based on a causal model $\mathbf{M}$ is a pair $\langle \mathbf{U}, \sqcup \rangle$ where:

i. $\mathbf{U} = \left\{ [\mathbf{M}]^A : \text{ for some } A \in \mathbf{L_o} \text{ such that } A \not\models \bot \right\}$;

ii. $\sqsubseteq$ is a partial order of $\mathbf{U}$ and $\mathbf{s} \sqcup \mathbf{t}$ iff $\mathbf{s} \sqsubseteq \mathbf{t}$;

iii. $[\mathbf{M}]^A \sqcup [\mathbf{M}]^B = [\mathbf{M}]^{A \wedge B}$.

Definition 18 (Verification and falsification)**.** Let $\mathbf{S_M} = \langle \mathbf{U}, \sqcup \rangle$ be any state space, s any state in $\mathbf{U}$. The relations $\Vdash$ (exact verification) and $\dashv\!\Vdash$ (falsification) are defined over $\mathbf{L_o}$ as follows:

i. $\mathbf{s} \Vdash x$ iff $\mathbf{s} = [\mathbf{M}]^x$

ii. $\mathbf{s} \dashv\!\Vdash x$ iff $\mathbf{s} = [\mathbf{M}]^{\bar{x}}$

iii. $\mathbf{s} \Vdash \neg A$ iff $\mathbf{s} \dashv\!\Vdash A$

iv. $\mathbf{s} \dashv\!\Vdash \neg A$ iff $\mathbf{s} \Vdash A$

v. $\mathbf{s} \Vdash A \wedge B$ iff $\mathbf{t} \Vdash A$ and $\mathbf{r} \Vdash B$ where $\mathbf{s} = \mathbf{t} \sqcup \mathbf{r}$

vi. $\mathbf{s} \dashv\!\Vdash A \wedge B$ iff $\mathbf{s} \dashv\!\Vdash A$ or $\mathbf{s} \dashv\!\Vdash B$ or $\mathbf{s} \dashv\!\Vdash A \vee B$

vii. $\mathbf{s} \Vdash A \vee B$ iff $\mathbf{s} \Vdash A$ or $\mathbf{s} \Vdash B$ or $\mathbf{s} \Vdash A \wedge B$

viii. $\mathbf{s} \dashv\!\Vdash A \vee B$ iff $\mathbf{t} \dashv\!\Vdash A$ and $\mathbf{r} \dashv\!\Vdash B$ where $\mathbf{s} = \mathbf{t} \sqcup \mathbf{r}$

References

Berto, Franz & Aybüke Özgün. 2021. Indicative conditionals: probabilities and relevance. *Philosophical Studies* 178. 3697–3730.

Briggs, R. A. 2012. Interventionist counterfactuals. *Philosophical Studies* 160. 139–166.

Burgess, John P. 1981. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic* 22(1). 76–84.

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2019. *Inquisitive semantics*, Oxford Surveys in Semantics and Pragmatics. Oxford University Press.

Ciardelli, Ivano, Linmin Zhang & Lucas Champollion. 2018. Two switches in the theory of counterfactuals. *Linguistics & Philosophy* 41. 577–621. http://dx.doi.org/10.1007/s10988-018-9232-4.

Fine, Kit. 1975. Critical notice of Counterfactuals by David Lewis. *Mind* 84. 451—58.

Fine, Kit. 2012. Counterfactuals without possible worlds. *Journal of Philosophy* 109(3). 221–146.

von Fintel, Kai. 2001. Counterfactuals in a dynamic context. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*. MIT Press.

van Fraassen, Bas C. 1974. Hidden variables in conditional logic. *Theoria* 40(3). 176–190.

Galles, David & Judea Pearl. 1998. An axiomatic characterization of causal counterfactuals. *Foundations of Science* 3. 151–182. http://dx.doi.org/10.1023/A:1009602825894.

Gillies, Anthony S. 2007. Counterfactual scorekeeping. *Linguistics & Philosophy* 30. 329–360. http://dx.doi.org/10.1007/s10988-007-90186.

Gillies, Anthony S. 2023. Some iffy comments on Khoo 2022. Central APA Author Meets Critics.

Halpern, Joseph Y. 2013. From causal models to counterfcatual structures. *The Review of Symbolic Logic* 6(2). 305–322. http://dx.doi.org/10.1017/s1755020312000305. URL http://dx.doi.org/10.1017/S1755020312000305.

Herzberger, Hans G. 1979. Counterfactuals and consistency. *Journal of Philosophy* 76(2). 83–88.

Kaufmann, Stefan. 2013. Causal premise semantics. *Cognitive Science* 37(6). 1136–1170. http://dx.doi.org/10.1111/cogs.12063. URL http://dx.doi.org/10.1111/cogs.12063.

Khoo, Justin. 2022. *The meaning of if*. Oxford University Press.

Kratzer, Angelika. 1981. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic* 10. 201–16. Reprinted in: Angelika Kratzer, *Modals and conditionals* (Oxford University Press, 2012).

Kratzer, Angelika. 1991. Modality. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantics: An international handbook of contemporary research*, 639–650. Berlin: de Gruyter.

Krauss, Sarit, Daniel Lehmann & Menachem Magidor. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44(1–2). 167–207.

Lewis, David. 1973. *Counterfactuals*. Blackwell.

Lewis, David. 1981. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic* 10. 217–234.

Pearl, Judea. 2009. *Causality: Models, reasoning, and inference*. Cambridge University Press, 2nd edn.

Pollock, John L. 1976a. The 'possible worlds' analysis of counterfactuals. *Philosophical Studies* 29(6). 469–476.

Pollock, John L. 1976b. *Subjunctive reasoning*. D. Reidel.

Santorio, Paolo. 2019. Interventions in premise semantics. *Philosophers' Imprint* 19(1). 1–27.

Stalnaker, Robert C. & Richmond H. Thomason. 1967. A semantic analysis of conditional logic. Tech. rep., Yale University.

Veltman, Frank. 1976. Prejudices, presuppositions and the theory of counterfactuals. In Jeroen Groenendijk & Martin Stokhof (eds.), *Amsterdam papers in formal grammar: Proceedings of the 1st amsterdam colloquium*, 248–281. University of Amsterdam.

Veltman, Frank. 1985. *Logics for conditionals*. Ph.D. thesis, University of Amsterdam.