# Simplifying Counterfactuals

Malte Willer

University of Chicago
willer@uchicago.edu

**Abstract**

The fact that counterfactuals in general license simplification of disjunctive antecedents is a familiar problem for the traditional Lewis-Stalnaker variably strict analysis of counterfactuals. This paper argues that recent semantic attempts to solve this problem in a variably strict setting do not address related simplification patterns and demonstrates that the data are well explained by a dynamic strict analysis of counterfactuals using ideas from the inquisitive semantic tradition.

## 1 The Plot

On a textbook variably strict analysis ([11],[18]) a *would*-counterfactual $\ulcorner\phi \mathrel{\Box\!\!\rightarrow} \psi\urcorner$ is true at some possible world $w$ iff $f_c(w, [\![\phi]\!]) \subseteq [\![\psi]\!]$, where $f_c(w, [\![\phi]\!])$ denotes the $\phi$-worlds closest to $w$ given some contextually provided similarity relation between worlds. The fact that counterfactuals in general license Simplification of Disjunctive Antecedents (SDA) is a familiar problem for this analysis ([5],[13]):

**SDA** $(\phi \vee \psi) \mathrel{\Box\!\!\rightarrow} \chi \vDash \phi \mathrel{\Box\!\!\rightarrow} \chi, \psi \mathrel{\Box\!\!\rightarrow} \chi$

Given classical disjunction, SDA inferences are unexpected in a variably strict setting since $f_c(w, [\![\phi \vee \psi]\!]) \neq f_c(w, [\![\phi]\!]) \cup f_c(w, [\![\psi]\!])$ unless the closest $\phi$-worlds are just as close to $w$ as are the closest $\psi$-worlds, leaving it unexplained why (1) entails that the party would have been fun if Alice had come *and* that the party would have been fun if Bert had come:

(1) If Alice or Bert had come to the party, it would have been fun.
 a. $\rightsquigarrow$ If Alice had come to the party, it would have been fun.
 b. $\rightsquigarrow$ If Bert had come to the party, it would have been fun.

The first observation of this paper—discussed in the remainder of this section—is that the trouble with simplification goes beyond SDA in interesting ways. The second observation is that the data are well explained by a dynamic strict analysis of counterfactuals that uses ideas from the inquisitive semantic tradition.

Semantic attempts to deal with the SDA trouble in a variably strict setting exist but do not address the full range of simplification data. One idea is to go with a Hamblin-style analysis of disjunction and then let *if*-clauses be universal quantifiers so that $\ulcorner\phi \mathrel{\Box\!\!\rightarrow} \psi\urcorner$ is now true at some possible world $w$ iff $f_c(w, p) \subseteq [\![\psi]\!]$ for all propositions $p$ in the set of alternatives denoted by the antecedent $\phi$—a singleton in case of non-disjunctive antecedents; a set containing all the atomic propositional disjuncts if the antecedent is a disjunct (see [2]). Another idea is to adopt an existential analysis of disjunction—the antecedent of (1), for instance, would be of the form '$\exists x.\, \mathrm{C}x \wedge (x = \text{Alice} \vee x = \text{Bert})$' with 'C' denoting the property of coming to the party—and a treatment of indices of evaluation as world-assignment pairs. If we now say that two such pairs are unconnected (and hence none of them more similar to the index of evaluation than the other) if their assignments differ, we predict that the counterfactual selection function includes indices at which Alice comes to the party as well as indices at which Bert comes to the party (see [16]).

Both approaches explain why (1) simplifies but are hand-tailored to handle simplifications of *disjunctive antecedents*: what does the explanatory lifting in each approach is the special interaction between a non-classical analysis of disjunction with whatever is involved in interpreting *if*-clauses. And this cannot be the whole story since a counterfactual like (2) with a *negated conjunction* as antecedent also simplifies (see [14]) and since a *might*-counterfactual such as (3) allows for simplification of its disjunctive *consequent*:

(2) If Nixon and Agnew had not both resigned, Ford would never have become president.
  a. $\rightsquigarrow$ If Nixon had not resigned, Ford would never have become president.
  b. $\rightsquigarrow$ If Agnew had not resigned, Ford would never have become president.

(3) If Mary had not gone to Pisa, she might have gone to Lisbon or Rome.
  a. $\rightsquigarrow$ If Mary had not gone to Pisa, she might have gone to Lisbon.
  b. $\rightsquigarrow$ If Mary had not gone to Pisa, she might have gone to Rome.

It will not do to just stipulate that (2) is evaluated by checking its consequent against the union of the closest worlds in which Nixon and Agnew do not resign, respectively: this fact calls for an explanation in terms of negation and conjunction just as much as SDA called for an explanation in terms of disjunction.[1] Explaining the simplification pattern exhibited by *might*-counterfactuals such as (3) is also challenging in a variably strict setting under the reasonable assumption that '$\Diamond\!\!\rightarrow$' and '$\Box\!\!\rightarrow$' are duals ($\phi \Diamond\!\!\rightarrow \psi =_{\text{def}} \neg(\phi \Box\!\!\rightarrow \neg\psi)$). For observe that given some context $c$, the truth of $\ulcorner\neg(\phi \Box\!\!\rightarrow \neg(\psi \vee \chi))\urcorner$ at some possible world $w$ only requires that $f_c(w, [\![\phi]\!]) \cap [\![\psi]\!] \neq \varnothing$ or $f_c(w, [\![\phi]\!]) \cap [\![\chi]\!] \neq \varnothing$, not both. In any case, a comprehensive story about how and why counterfactuals simplify cannot be dependent on the interpretation of *if*-clauses (or on considerations about the similarity relation) alone since simplification is also a feature of disjunctive counterfactual consequents.

The Lewis-Stalnaker analysis has a problem with simplification that goes beyond the familiar observations about SDA. One may, of course, couple such an analysis with a powerful pragmatic supplement (see for instance [7],[17]) but in the following §2 I demonstrate that the simplification data already receive a straightforward explanation in a suitably elaborated strict analysis of counterfactuals ([6],[8],[24]). §3 shows how to avoid the most pressing problems with any theory that treats SDA as valid, including the observation that certain counterfactuals seem to resist simplification.

## 2 Basic Framework

The proposal to be developed here is that a *would*-counterfactual $\ulcorner\phi \Box\!\!\rightarrow \psi\urcorner$ is a strict material conditional $\ulcorner\Box(\phi \supset \psi)\urcorner$ over a contextually determined but dynamically evolving domain of quantification presupposing that its antecedent $\phi$ is a possibility in that domain. Assuming duality and that presuppositions project out of negation, a *might*-counterfactual $\ulcorner\phi \Diamond\!\!\rightarrow \psi\urcorner$ amounts to an assertion of $\ulcorner\Diamond(\phi \wedge \psi)\urcorner$ under the presupposition that $\phi$ is possible. Once we predict that $\ulcorner\Diamond(\phi \vee \psi)\urcorner$ and $\ulcorner\Diamond\neg(\neg\phi \wedge \neg\psi)\urcorner$ entail $\ulcorner\Diamond\phi\urcorner$ *as well as* $\ulcorner\Diamond\psi\urcorner$ via a semantic free choice effect, we predict that (1)–(3) simplify in the way they do.

The target language $\mathcal{L}$ contains a set of sentential atoms $\mathcal{A} = \{p, q, ...\}$ and is closed under negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$), the modal possibility operator ($\Diamond$), and the *would*-counterfactual ($\Box\!\!\rightarrow$). Other connectives are defined in the usual manner. I begin by stating a simple semantics for $\mathcal{L}$ that predicts the free choice effect and the key simplification observations. The semantics does not give us everything one might hope for but—as I will show in the next section—many potential shortcomings are avoided once we add some bells and whistles to the basic story.

[1] Not all counterfactuals with negated conjunctions simplify, as an anonymous reviewer helpfully points out: 'If John had not had that terrible accident last week and died, he would have been here today' does not license 'If John had not died, he would have been here today' since he still might have had that accident. What underlies this observation, I suggest, is that negated conjunctions sometimes give rise to a 'neither' rather than a 'not both' reading (see [20] for detailed discussion) and that, unsurprisingly, only the latter licenses simplification.

## 2.1 Modals

The analysis of modals is in the spirit of Veltman's approach ([21]) but here we do not treat input contexts as sets of possible worlds but as sets of consistent propositions (which I label here *alternatives*).

**Definition: Possible Worlds, Propositions.** $w$ is a *possible world* iff $w\colon \mathcal{A} \to \{0,1\}$. $W$ is the set of such $w$'s, $\mathcal{P}(W)$ is the powerset of $W$. The function $[\![\cdot]\!]$ assigns to nonmodal formulas of $\mathcal{L}$ a *proposition* in the familiar fashion. $\underline{\perp}$ is the contradictory proposition (the empty set of possible worlds) while $\underline{\not\perp}$ is any consistent proposition.

**Definition: States, Alternatives.** A *state* $s \subseteq \mathcal{P}(W)\backslash \underline{\perp}$ is any set of consistent propositions (*alternatives*). $S$ is just the set of all such states. The *information* carried by a state $s$ is the set of possible worlds compatible with it so that $\mathtt{info}(s) = \{\bigcup \sigma\colon \sigma \in s\}$. We refer to $\varnothing$ as the *absurd* state and speak of $s_0 = \mathcal{P}(W)\backslash \underline{\perp}$ as the *initial* state.

States thus have informational content in the sense that they rule out certain ways the world could be. In addition, they encode this information as a set of alternatives (which do not have to be mutually exclusive).

States are updated by updating each of their alternatives. Updates on an alternative $\sigma$ are sensitive to the state $s$ containing it since modals perform tests on the state's informational content. Furthermore, I will think of update rules as *relations* between alternatives to capture the inquisitive effect of disjunction and distinguish between a *positive* acceptance inducing update relation $[\cdot]^+_s$ and a *negative* rejection inducing update relation $[\cdot]^-_s$ to allow for inquisitive negation (inspired by [1],[9]). So for instance we shall say:

$$(\mathcal{A}) \quad \begin{array}{l} \sigma[p]^+_s\tau \text{ iff } \tau = \sigma \cap [\![p]\!] \\ \sigma[p]^-_s\tau \text{ iff } \tau = \sigma\backslash[\![p]\!] \end{array} \qquad (\neg) \quad \begin{array}{l} \sigma[\neg\phi]^+_s\tau \text{ iff } \sigma[\phi]^-_s\tau \\ \sigma[\neg\phi]^-_s\tau \text{ iff } \sigma[\phi]^+_s\tau \end{array}$$

A positive update with a sentential atom $p$ eliminates from an alternative all possible worlds at which $p$ is false while a negative update with $p$ eliminates all possible worlds at which $p$ is true. A positive update with $\ulcorner\neg\phi\urcorner$ is just a negative update with $\phi$ and a negative update with $\ulcorner\neg\phi\urcorner$ is just a positive update with $\phi$.

The basic idea about possibility modals is that they test whether their prejacent relates the information carried by a state to a contradiction $\underline{\perp}$ or to a consistent proposition $\underline{\not\perp}$:

$$(\Diamond) \quad \begin{array}{l} \sigma[\Diamond\phi]^+_s\tau \text{ iff } \tau = \{w \in \sigma\colon \langle \mathtt{info}(s), \underline{\perp}\rangle \notin [\phi]^+_s\} \\ \sigma[\Diamond\phi]^-_s\tau \text{ iff } \tau = \{w \in \sigma\colon \langle \mathtt{info}(s), \underline{\not\perp}\rangle \notin [\phi]^+_s\} \end{array}$$

For an alternative in a state $s$ to pass the test imposed by a *positive* update with $\ulcorner\Diamond\phi\urcorner$, the information carried by $s$ must not be related to the *inconsistent* proposition via a positive update with $\phi$. For an alternative in a state $s$ to pass the test imposed by a *negative* update with $\ulcorner\Diamond\phi\urcorner$—that is, a positive update with $\ulcorner\Box\neg\phi\urcorner$—the information carried by $s$ must not be related to a *consistent* proposition via a positive update with $\phi$.

So far we only have a rewrite of classical Update Semantics but the present setup allows us to combine a test semantics for modals with an inquisitive analysis of disjunction and negated conjunction. Start by coupling each update rule for alternatives with a corresponding update procedure for states:

**Definition: Updates on States.** Define two update operations $\uparrow, \downarrow\colon (\mathcal{L} \times S) \mapsto S$:

1. $s \uparrow \phi = \{\tau \neq \underline{\perp}\colon \exists\sigma \in s.\, \sigma[\phi]^+_s\tau\}$
2. $s \downarrow \phi = \{\tau \neq \underline{\perp}\colon \exists\sigma \in s.\, \sigma[\phi]^-_s\tau\}$

A positive/negative update of some state $s$ with $\phi$ delivers all the alternatives that are positively/negatively related to some element of $s$ via $\phi$.

The proposal for disjunction is then the following one:

$$(\vee)\quad \begin{array}{l}\sigma[\phi \vee \psi]^+_s \tau \text{ iff } \sigma[\phi]^+_s \tau \ \underline{\text{or}}\ \sigma[\psi]^+_{s\downarrow\phi}\tau \\ \sigma[\phi \vee \psi]^-_s \tau \text{ iff } \exists\nu\colon \sigma[\phi]^-_s \nu \ \underline{\text{and}}\ \nu[\psi]^-_{s\downarrow\phi}\tau\end{array}$$

This analysis captures two important intuitions about disjunctions: first, in addition to ruling out certain possibilities they raise each of their disjuncts as an issue in discourse. We capture this by letting a disjunction relate an input alternative to two potentially distinct alternatives: the result of updating with the first and the result of updating with the second disjunct. Moreover, in a sentence such as "Mary is in Chicago or she must be in New York" the modal in the second disjunct naturally receives a modally subordinated interpretation: it is interpreted under the supposition that Mary is not in Chicago. We achieve this result by saying that whenever a disjunction is processed in light of some state $s$, its second disjunct is processed in light of a negative update of $s$ with the first disjunct.

Given some state $s$, a positive update with a conjunction $\ulcorner\phi \wedge \psi\urcorner$ proceeds via a positive update with $\phi$ light of $s$ and then via a positive update with $\psi$ in light of $s \uparrow \phi$:

$$(\wedge)\quad \begin{array}{l}\sigma[\phi \wedge \psi]^+_s \tau \text{ iff } \exists\nu\colon \sigma[\phi]^+_s \nu \ \underline{\text{and}}\ \nu[\psi]^+_{s\uparrow\phi}\tau \\ \sigma[\phi \wedge \psi]^-_s \tau \text{ iff } \sigma[\phi]^-_s \tau \ \underline{\text{or}}\ \sigma[\psi]^-_{s\uparrow\phi}\tau\end{array}$$

The rules for negative updates with disjunctions and conjunctions enforce the validity of De Morgan's Laws.

Let me highlight some predictions before moving on to conditionals. As a preparation, define the notions of support, entailment, and consistency in the familiar dynamic fashion:

**Definition: Support, Entailment, Consistency.** Take any $s \in S$ and formulas of $\mathcal{L}$:

1. $s$ supports $\phi$, $s \Vdash \phi$, iff $s \uparrow \phi = s$
2. $\phi_1, \ldots, \phi_n$ *entails* $\psi$, $\phi_1, \ldots, \phi_n \vDash \psi$, iff for all $s \in S$, $s \uparrow \phi_1 \ldots \uparrow \phi_n \Vdash \psi$
3. $\phi_1, \ldots \phi_n$ is *consistent* iff for some $s \in S$: $s \uparrow \phi_1 \ldots \uparrow \phi_n \neq \varnothing$

A state supports $\phi$ just in case a positive update of $s$ with $\phi$ idles. Entailment is just guaranteed preservation of support and the consistency of a sequence requires that a positive update with it sometimes results in a non-absurd state. It would, of course, be possible to define the notions of entailment and consistency on the basis of $\downarrow$ but I set an exploration of this interesting avenue aside for now.

Disjunctions embedded under a possibility operator then exhibit the free choice effect:

**Fact 1.** $\Diamond(p \vee q) \vDash \Diamond p \wedge \Diamond q$

The underlying observation here is that $s \uparrow \Diamond(p \vee q) \neq \varnothing$ only if $\langle \texttt{info}(s), \underline{\bot}\rangle \notin [p \vee q]^+_s$. But suppose that $\texttt{info}(s)$ fails to contain both $p$- and $q$-worlds: then $[p]^+_s$ or $[q]^+_s$ *does* relate $\texttt{info}(s)$ to $\underline{\bot}$, hence $\texttt{info}(s)[p \vee q]^+_s \underline{\bot}$ and thus $\langle \texttt{info}(s), \underline{\bot}\rangle \in [p \vee q]^+_s$ after all. So if $s \uparrow \Diamond(p \vee q) \neq \varnothing$ then $s \uparrow \Diamond p = s$ and $s \uparrow \Diamond q = s$.

Note that $\Diamond(p \vee q) \nvDash \Diamond(p \wedge q)$ since passing the test conditions under consideration does not require the presence of a $p \wedge q$-world in $\texttt{info}(s)$. Furthermore, it is easy to see that the free choice effect also arises if $\Diamond$ scopes over a negated conjunction since for all choices of $s \in S$ we have $[\neg(\neg\phi \wedge \neg\psi)]^+_s = [\phi \vee \psi]^+_s$ by design.

We also account for the observation (see [1] and references therein) that embedding a disjunctive possibility under negation reverts disjunction to its classical behavior:

**Fact 2.** $\neg\Diamond(p \vee q) \vDash \neg\Diamond p \wedge \neg\Diamond q$

Observe that $s \uparrow \neg\Diamond(p \vee q) \neq \varnothing$ only if $\langle \texttt{info}(s), \underline{\not\perp}\rangle \notin [p \vee q]^+_s$. But suppose that $\texttt{info}(s)$ contains a $p$- or a $q$-world: then $[p]^+_s$ or $[q]^+_s$ *does* relate $\texttt{info}(s)$ to $\underline{\not\perp}$, hence $\texttt{info}(s)[p \vee q]^+_s \underline{\not\perp}$ and thus $\langle \texttt{info}(s), \underline{\not\perp}\rangle \in [p \vee q]^+_s$ after all. So if $s \uparrow \neg\Diamond(p \vee q) \neq \varnothing$ then $s \uparrow \neg\Diamond p = s$ and $s \uparrow \neg\Diamond q = s$.

We may also observe that the framework developed here preserves key insights from the dynamic analysis of modals, including the internal dynamics of conjunction:

**Fact 3.** $\neg p \wedge \Diamond p$ is inconsistent

Here it pays off that updates are defined relative to a shifty state parameter $s$. Clearly $\texttt{info}(s \uparrow \neg p)$ does not contain any $p$-worlds and so any update with with '$\Diamond p$' in light of $s \uparrow \neg p$ is guaranteed to result in the absurd state.

Finally, let me just state some observations about the material conditional and the necessity operator that are of relevance for the upcoming discussion:

**Fact 4.** For all $s \in S$: $[\Box(\phi \supset \psi)]^+_s = [\neg\Diamond(\phi \wedge \neg\psi)]^+_s$ and $[\Box(\phi \supset \psi)]^-_s = [\Diamond(\phi \wedge \neg\psi)]^+_s$

These identities follow immediately if we treat '$\Diamond$' and '$\Box$' as duals and adopt the standard analysis of the material conditional in terms of conjunction and negation.

In sum, the proposal developed so far combines a test semantics for modals with an inquisitive approach to disjunction and negated conjunction in a way that captures the scope as well as the limits of the free choice effect. Let me now turn to counterfactuals.

## 2.2 Counterfactuals

A *would*-counterfactual is a strict material conditional presupposing that its antecedent is possible. Following standard protocol I treat *might*- and *would*-counterfactuals as duals and presuppositions as definedness conditions on updating ([10],[3]):

$$(\Box\!\!\rightarrow) \quad \begin{array}{l} \sigma[\phi \Box\!\!\rightarrow \psi]^+_s \tau \text{ iff } \sigma[\Diamond\phi]^+_s \sigma \text{ \underline{and} } \sigma[\Box(\phi \supset \psi)]^+_s \tau \\ \sigma[\phi \Box\!\!\rightarrow \psi]^-_s \tau \text{ iff } \sigma[\Diamond\phi]^+_s \sigma \text{ \underline{and} } \sigma[\Box(\phi \supset \psi)]^-_s \tau \end{array}$$

Given some state $s$, a positive or negative update with $\ulcorner \phi \Box\!\!\rightarrow \psi \urcorner$ fails to relate an input alternative $\sigma$ to any output in case the information carried by $s$ is incompatible with $\phi$ (the presupposition thus projects out of negation). Assuming that the presupposition is satisfied, a positive update with $\ulcorner \phi \Box\!\!\rightarrow \psi \urcorner$ then tests whether $s$ supports $\ulcorner \phi \supset \psi \urcorner$ while a negative update effectively asks whether $\ulcorner \phi \wedge \neg\psi \urcorner$ is compatible with $s$.[2]

For convenience, let me state explicitly the update rules for *might*-counterfactuals:

$$(\Diamond\!\!\rightarrow) \quad \begin{array}{l} \sigma[\phi \Diamond\!\!\rightarrow \psi]^+_s \tau \text{ iff } \sigma[\Diamond\phi]^+_s \sigma \text{ \underline{and} } \sigma[\Diamond(\phi \wedge \psi)]^+_s \tau \\ \sigma[\phi \Diamond\!\!\rightarrow \psi]^-_s \tau \text{ iff } \sigma[\Diamond\phi]^+_s \sigma \text{ \underline{and} } \sigma[\Diamond(\phi \wedge \psi)]^-_s \tau \end{array}$$

These update rules are an immediate consequence of treating '$\Box\!\!\rightarrow$' and '$\Diamond\!\!\rightarrow$' as duals.

It is of course uncontroversial that this analysis predicts that counterfactuals simplify if their antecedents involve a disjunction or a negated conjunction:

**Fact 5.** $(p \vee q) \Box\!\!\rightarrow r \models p \Box\!\!\rightarrow r, q \Box\!\!\rightarrow r$ and $\neg(p \wedge q) \Box\!\!\rightarrow r \models \neg p \Box\!\!\rightarrow r, \neg q \Box\!\!\rightarrow r$

This is an immediate consequence of analyzing *would*-counterfactuals as strict material conditionals. The claim that counterfactuals presuppose the possibility of their antecedents, however, immediately predicts that $s \uparrow (p \vee q) \Box\!\!\rightarrow r = \varnothing$ unless $\texttt{info}(s)$ includes $p$- as well as $q$-worlds due to the free choice effect. I will come back to this fact momentarily, but we can already at this stage observe the following fact about *might*-counterfactuals:

[2] Bringing presuppositions into the picture also raises the question of how they project and a proper answer requires minor modifications to some of our original update rules. For instance, in order to predict that presuppositions project out of the possibility operator one would need to say that $\sigma[\Diamond\phi]^+_s \tau$ holds just in case $\tau = \{w \in \sigma \colon \langle \texttt{info}(s), \underline{\perp} \rangle \notin [\phi]^+_s\}$ and, moreover, $\exists\nu.\, \sigma[\phi]^+_s \nu$. Likewise for the negative entry: $\sigma[\Diamond\phi]^-_s \tau$ holds just in case $\tau = \{w \in \sigma \colon \langle \texttt{info}(s), \underline{\not\perp} \rangle \notin [\phi]^+_s\}$ and, moreover, $\exists\nu.\, \sigma[\phi]^-_s \nu$. I set these additional complexities, which would also affect the update rules to disjunction, aside to streamline the notation and since getting all the facts about presupposition projection right goes beyond the scope of this investigation.

**Fact 6.** $p \diamondto (q \vee r) \vDash p \diamondto q, p \diamondto r$

Clearly a state $s$ supports '$p \diamondto (q \vee r)$' just in case it supports '$\Diamond p$' and '$\Diamond(q \vee r)$' and hence—due to the free choice effect—both '$\Diamond q$' and '$\Diamond r$'. So we predict that *might*-counterfactuals such as (3) simplify in the way the do. In contrast *would*-counterfactuals with disjunctive consequents rightly fail to simplify: a nonempty state supporting '$p \wedge (q \wedge \neg r)$', for instance, supports '$p \boxright (q \vee r)$' without supporting '$(p \boxright r)$'.

This is all good news but there remain some open problems. Fine and Warmbrōd worry that SDA entails the validity of ANTECEDENT STRENGTHENING (AS) assuming substitution of logical equivalents in conditional antecedents (see [5],[22]). While AS fails to be valid here since a state may support the possibility presupposition carried by $\ulcorner \phi \boxright \chi \urcorner$ without supporting the one carried by $\ulcorner (\phi \wedge \psi) \boxright \chi \urcorner$, the framework developed so far fails to leave room for the consistency of Sobel sequences such as (4):

(4) If Mary had come to the party, it would have been fun. But if Bert had come too, it would not have been fun.

No consistent state compatible with Alice and Bert coming to the party can support the asserted contents of both counterfactuals, and every state that is incompatible with that possibility inevitably fails to satisfy the presuppositions carried by the first or the second member of the sequence. In any case updating with the sequence in (4) is guaranteed to result in the absurd state and thus counts as inconsistent, which is not a good result.

The second worry is that there are familiar cases in which simplification of disjunctive antecedents seems to fail. McKay and van Inwagen ([12]) consider the following case:

(5) If Spain had fought for the Axis or the Allies, she would have fought for the Axis.
a. $\rightsquigarrow$ If Spain had fought for the Axis, she would have fought for the Axis.
b. ??? If Spain had fought for the Allies, she would have fought for the Axis.

(5a) is of course trivial but (5b) is objectionable, contrary to what SDA predicts.

Both problems can be solved by adding a few complexities to our basic story about presupposition and assertion. Presuppositions in general and possibility presuppositions in particular are normally *accommodated* as discourse proceeds, allowing counterfactual domains of quantification to evolve dynamically—this is what underlies the consistency of Sobel sequences. Assertions sometimes conflict with other bits of information taken for granted in discourse and thus may *pragmatically* trigger modifications of the input context so that the proposed update may be processed—this is what explains why SDA is semantically valid but occasionally defeated by intervening pragmatic factors.

# 3 Hyperstates

The twist to the basic story is the idea that modals and counterfactuals are quantifiers over a *minimal* and *dynamically evolving* domain of quantification. To make sense of this idea we let context determine a slightly more complex background for processing counterfactuals: instead of thinking of a context as providing a single state, we will think of it as providing a *set* of (nonempty) states. Intuitively, the information carried by each state can be understood as a domain of quantification, and counterfactuals then pertain to whatever states come with the strongest informational content.

**Definition: Hyperstates.** A *hyperstate* $\pi \subseteq \mathcal{P}(S) \backslash \varnothing$ is any set of nonempty states. We say that $s' \preceq_\pi s$, $s'$ is *at least as strong as* $s$ in $\pi$, iff $s, s' \in \pi$ and $\texttt{info}(s') \subseteq \texttt{info}(s)$. $\Pi$ is the set of all hyperstates. We refer to $\varnothing$ as the *absurd* hyperstate and treat $\pi_0 = \mathcal{P}(S) \backslash \varnothing$ as the *initial* hyperstate.

Thinking of contexts as hyperstates requires some modifications to our update system. Fortunately, our update procedures for alternatives can stay the same. But states are now updated in light of a hyperstate and updates with modals now pertain to those elements of a hyperstate whose informational content is strongest. We achieve this by slightly modifying the update functions for states in the following manner:

**Definition: Hyper-updates on States** Define update functions $\uparrow_\pi, \downarrow_\pi : \mathcal{L} \mapsto (S \mapsto S)$ as follows:

1. $s \uparrow_\pi \phi = \{\tau \neq \perp : \exists \sigma \in s\ \exists s' \preceq_\pi s.\, \sigma[\phi]^+_{s'} \tau\}$
2. $s \downarrow_\pi \phi = \{\tau \neq \perp : \exists \sigma \in s\ \exists s' \preceq_\pi s.\, \sigma[\phi]^-_{s'} \tau\}$

To see why these modifications matter, consider how a modal now interacts with some $s \in \pi$: an update of $s$ with '$\Diamond p$' now tests whether there is some $s' \preceq_\pi s$ whose informational content $\texttt{info}(s')$ includes a $p$-world. Clearly, this is so just in case $\texttt{info}(s)$ includes a $p$-world as well, and so possibility modals work exactly as before. But the twist does matter when it comes to an update with '$\Box p$': this now tests whether there is some $s' \preceq_\pi s$ whose informational content $\texttt{info}(s')$ exclusively consists of $p$-worlds, and that may be so even if $\texttt{info}(s)$ itself includes a possible world at which $p$ is false (though of course no state stronger than $s'$ may contain a $\neg p$-world). So in this sense '$\Box$' becomes a strict quantifier over the informational content of the strongest members of a hyperstate. Updating with nonmodal formulas of $\mathcal{L}$ stays the same.

We now define what it takes for a context understood as selecting a hyperstate to accept and admit $\phi$ and define updates on hyperstates on that basis:

**Definition: Acceptance, Admission, Updates on Hyperstates.** Consider arbitrary $\pi \in \Pi$ and $\phi \in \mathcal{L}$:

1. $\pi$ *accepts* $\phi$, $\pi \Vdash \phi$, iff for all $s' \in \pi$ there exists some $s \preceq_\pi s'$: $s \uparrow_\pi \phi = s$
2. $\pi$ *admits* $\phi$, $\pi \rhd \phi$, iff $\pi \nVdash \neg\phi$
3. $\pi + \phi = \{s \uparrow_\pi \phi : s \in \pi \;\&\; \pi \rhd \phi\} \backslash \varnothing$

Acceptance of $\phi$ amounts to support by the strongest states in a hyperstate. An update with $\phi$ is admitted as long as its negation is not accepted. And finally, a hyperstate is updated with $\phi$ by updating each of its elements with $\phi$ and collecting the nonempty results, provided that an update with $\phi$ is admissible.

Entailment and consistency are again understood in the familiar dynamic fashion:

**Definition: Entailment and Consistency (Hyperstates).** Take any $\pi \in \Pi$ and formulas of $\mathcal{L}$:

1. $\phi_1, \ldots, \phi_n$ *entails* $\psi$, $\phi_1, \ldots, \phi_n \vDash \psi$, iff for all $\pi \in \Pi$, $\pi + \phi_n \ldots + \phi_n \Vdash \psi$
2. $\phi_1, \ldots \phi_n$ is *consistent* iff for some $\pi \in \Pi$: $\pi + \phi_1 \ldots + \phi_n \neq \varnothing$

This setup preserves everything said in §2 but in addition allows for a Sobel sequence like '$p \mathrel{\Box\!\!\rightarrow} r$' followed by '$(p \wedge q) \mathrel{\Box\!\!\rightarrow} \neg r$' to be consistent. To see why, assume that $w_1 \in [\![p \wedge \neg q \wedge r]\!]$ and that $w_2 \in [\![p \wedge q \wedge \neg r]\!]$, and let $\pi = \{s, s'\}$ be such that $\texttt{info}(s) = \{w_1\}$ while $\texttt{info}(s') = \{w_1, w_2\}$. Then clearly both $s$ and $s'$ satisfy the presupposition carried by the first counterfactual in the sequence and since $s \preceq_\pi s'$ and $s \uparrow_\pi p \mathrel{\Box\!\!\rightarrow} r = s$, we have $\pi + p \mathrel{\Box\!\!\rightarrow} r = \pi$. Notice furthermore that $\pi \nVdash \neg((p \wedge q) \mathrel{\Box\!\!\rightarrow} r)$ since we have $s \uparrow_\pi \neg((p \wedge q) \mathrel{\Box\!\!\rightarrow} r) \neq s$: here the underlying observation is that $s$ is the strongest state in $\pi$ but fails support the counterfactual's possibility presupposition '$\Diamond(p \wedge q)$'. So $\pi$ admits an update with the second member of the Sobel sequence, resulting in a consistent hyperstate $\pi' = \{s'\}$, as desired. More complex Sobel sequences can be consistently processed in more complex hyperstates. I conclude that the validity of SDA is compatible with the fact that counterfactuals resist AS so that Sobel sequences are consistent.

It remains to comment on the fact that counterfactuals with disjunctive antecedents do not seem to simplify across the board. Here I maintain that SDA is semantically valid and that there is a principled story for why pragmatic factors sometimes intervene. Part of the picture is that the problematic counterfactual (repeated here for convenience) intuitively communicates that Spain would never have fought for Allies.

(5) If Spain had fought for the Axis or the Allies, she would have fought for the Axis.

In fact, explicitly acknowledging the possibility of Spain joining the Allies renders (5) unacceptable (see also [19]):

(6) Spain might have fought for the Allies. ???But if Spain had fought for the Axis or the Allies, she would have fought for the Axis.

Given this communicative effect it is not surprising that a context that has been strengthened with (5) fails to support 'If Spain had fought for the Allies, she would have fought for the Axis' since it fails to satisfy the counterfactual's possibility presupposition.

One may think that the previous observation immediately undermines the proposal that counterfactuals with disjunctive antecedents presuppose the possibility of each disjunct. But this is not so: an indicative conditional such as 'If John wins the competition, then I am the Flying Dutchman' communicates—in ordinary circumstances anyway—that John will not win the competition, but there is no serious doubt that indicative conditionals presuppose that their antecedent is a possibility in the common ground. What is needed is a general story about how conditional assertions may at times remove worlds at which their antecedents are true from the domain of quantification.

Here is how such a story might go. Consider the asserted content of (5) in a context in which the counterfactual's possibility presupposition has been accommodated: updating with that content would result in a context in which Spain might have fought on the Axis *and* the Allies side, but of course it is a background assumption that Spain would not have fought on both sides, and hence clear to all discourse participants that updating with the asserted content of (5) results in the absurd state. Yet speakers are in general cooperative and thus do not propose to add information to the common ground that is incompatible with what is taken for granted. In these lights, it makes good sense to say that a speaker may communicate $\psi$ by asserting $\phi$ in case updating the discourse with $\phi$ results in the absurd state but updating that context with $\psi$ and then with $\phi$ does not. More precisely:

> Suppose that $\pi+\phi = \varnothing$ but $\pi+\psi+\phi \neq \varnothing$ and for all $\chi$ such that $\pi+\chi+\phi \neq \varnothing$, $\pi + \chi \Vdash \psi$. Then an utterance amounting to an assertion of $\phi$ in $\pi$ by default pragmatically implies a proposal to update $\pi + \psi$ with $\phi$.

Suppose then that $\pi \models \Diamond \mathtt{Ax} \wedge \Diamond \mathtt{Al}$ yet also $\pi \models \neg(\mathtt{Ax} \wedge \mathtt{Al})$, and consider $\pi + \Box((\mathtt{Ax} \vee \mathtt{Al}) \supset \mathtt{Ax})$, that is, the result of updating $\pi$ with the asserted content of (5). Clearly $\pi + \Box((\mathtt{Ax} \vee \mathtt{Al}) \supset \mathtt{Al}) \Vdash \Diamond(\mathtt{Ax} \wedge \mathtt{Al})$ and hence the update results in the absurd state. But we can also observe that $\pi + \neg\mathtt{Al}$ *is* consistent and *can* be consistently updated with the asserted content of (5) and so we predict—given the pragmatic twist just proposed—that the assertion that comes with an utterance of (5) implies that Spain would not have sided with the Allies. Since this is an implicature, we expect it to be cancellable, which is just what we have in (6). Furthermore, $\pi + \neg\mathtt{Ax}$, while consistent, *cannot* be consistently updated with the asserted content of (5), and so we do not make the wrong prediction that there is an implicature that Spain would not have sided with the Axis.

I thus conclude that there is a principled pragmatic explanation for why certain counterfactuals resist simplification. The reason, in brief, is that certain counterfactuals imply information in addition to asserting a strict material conditional, and that this additional information may in fact eliminate possibilities that have been brought into view by presupposed content. While this account taps into pragmatic resources to account for the problematic data, the needed assumption appears to be fairly modest and well-motivated.

## 4 Conclusion

A dynamic strict analysis that exploits insights from inquisitive semantics predicts why counterfactuals simplify in the way they do. The explanation is compatible with the consistency of Sobel sequences and with the fact that certain counterfactuals resist simplification for principled pragmatic reasons.

Let me conclude the discussion by pointing to two remaining tasks left for another day. First, simplification is not a feature particular to counterfactual conditionals: indicative conditionals are just as amenable to simplification as are their counterfactuals cousins. What is needed then is a story about how the framework developed here can be generalized so that it covers other conditional constructions as well. While this is not a trivial task, the basic idea is clear: all conditionals are strict and come with a possibility presupposition; their differences amount to differences in the domain of quantification.

Second, the attempt to combine a dynamic treatment of modals with an inquisitive treatment of disjunction brings to mind the question, discussed by [4] and [15], of how to distinguish between informative, inquisitive, and attentive content. Here I want to briefly observe that the notion of a hyperstate is fine-grained enough to keep track of various kinds of discourse information. First, there is the *informational content* of a hyperstate $\pi$, understood as the possible worlds compatible with what is taken for granted: $\mathtt{Info}(\pi) = \bigcup\{\mathtt{info}(s)\colon s \in \pi\}$. Second, we may associate with $\pi$ an *issue* understood as the set of its maximal alternatives: $\mathtt{Issue}(\pi) = \{\sigma\colon \sigma \in \mathtt{Alt}(\pi) \;\&\; \neg\exists\tau \in \mathtt{Alt}(\pi).\, \sigma \subset \tau\}$, where $\mathtt{Alt}(\pi) = \bigcup\{s\colon s \in \pi\}$. Looking at the initial hyperstate $\pi_0$ we can then say that an atomic sentence $p$ has $[\![p]\!]$ as its *informational* content in the sense that the informational content of $\pi_0 + p$ is just $[\![p]\!]$. For parallel reasons we can say that '$p \vee q$' has $[\![p]\!] \cup [\![q]\!]$ as its informational content and $\{[\![p]\!], [\![q]\!]\}$ as its *inquisitive* content since $[\![p]\!]$ and $[\![q]\!]$ are in the issue of $\pi_0 + p \vee q$. The distinction between the informativeness of a formula and its inquisitiveness is thus analyzed in terms of the potential to eliminate possibilities and to raise issues in discourse. And this strategy also allows us to look at hyperstates in a way that identifies yet another kind of content. Let me explain.

Earlier I said that each element of a hyperstate is a potential domain of quantification. If we think of a domain of quantification as a candidate for the region of logical space that is *relevant* for the modal discourse under consideration, it makes sense to think of a hyperstate $\pi$ as identifying what I have called elswhere a set of *serious* or *live possibilities* (see [23]), that is, the possibilities compatible with every state in $\pi$: $\mathtt{Live}(\pi) = \{\sigma\colon \forall s \in \pi\; \exists w \in \mathtt{info}(s).\, w \in \sigma\}$. We can then say that a sentence has *attentive* content in virtue of its potential to bring hitherto ignored possibilities into view: '$\Diamond p$', for instance, has $\{[\![p]\!]\}$ as its attentive content since $\pi_0 + \Diamond p$ treats $[\![p]\!]$ as a live possibility, and for parallel reasons '$\Diamond(p \vee q)$' has $\{[\![p]\!], [\![q]\!]\}$ as its attentive content.

A comprehensive discussion would explore in more detail what predictions the setup sketched here makes about the interaction between informational, inquisitive, and attentive content, and how it predictions differ from those of other frameworks. For now, let me just conclude that the framework is of general semantic interest beyond its capacity to handle some empirical challenges pertaining to inferences licensed by counterfactual conditionals: it combines a very attractive dynamic semantic treatment of possibility modals as highlighting the significance of certain possibilities in discourse with a—no less attractive—inquisitive treatment of disjunction as refining issues in discourse. The fact that combining these treatments allows us to make substantial progress toward a better understanding of the free choice effect gives us all the more reason to think that the dynamic inquisitive story told here deserves further exploration.

**Acknowledgments** Special thanks to Daniel Lassiter, Patrick Munoz, Paul Portner, and three anonymous referees for comments and discussion.

## References


[1] Martin Aher. Free choice in deontic inquistive semantics (dis). *Lecture Notes in Computer Science*, 7218:22–31, 2012.

[2] Luis Alonso-Ovalle. Counterfactuals, correlatives, and disjunction. *Linguistics and Philosophy*, 32(2):207–244, 2009.

[3] David I. Beaver. *Presupposition and Assertion in Dynamic Semantics*. CSLI Publications, Stanford, 2001.

[4] Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. Attention! *Might* in inquisitive semantics. *Proceedings of SALT XIX*, pages 91–108, 2009.

[5] Kit Fine. Critical notice. *Mind*, 84(335):451–458, 1975.

[6] Kai von Fintel. Counterfactuals in a dynamic context. In Michael Kenstowicz, editor, *Ken Hale: A Life in Language*, pages 123–152. MIT Press, Cambridge, MA, 2001.

[7] Michael Franke. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, 4(1):1–82, 2011.

[8] Anthony S. Gillies. Counterfactual scorekeeping. *Linguistics and Philosophy*, 30(3):329–360, 2007.

[9] Jeroen Groenendijk and Floris Roelofsen. Towards a suppositional inquisitive semantics, 2015. Manuscript, University of Amsterdam.

[10] Irene Heim. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts, Amherst, 1982.

[11] David K. Lewis. *Counterfactuals*. Harvard University Press, Cambridge, MA, 1973.

[12] Thomas McKay and Peter van Inwagen. Counterfactuals with disjunctive antecedents. *Philosophical Studies*, 31(5):353–356, 1977.

[13] Donald E. Nute. Counterfactuals and the similarity of worlds. *Journal of Philosophy*, 72(21):773–778, 1975.

[14] Donald E. Nute. *Topics in Conditional Logic*. Reidel, Dordrecht, 1980.

[15] Floris Roelofsen. A bare bone attentive semantics for *Might*. In Maria Aloni, Michael Franke, and Floris Roelofsen, editors, *The Dynamic, Inquisitive, and Visionary Life of $\phi$, $?\phi$, and $\Diamond\phi$*, pages 190–215. ILLC Publications, Amsterdam, 2013.

[16] Robert van Rooij. Free choice counterfactual donkeys. *Journal of Semantics*, 23(4):383–402, 2006.

[17] Robert van Rooij. Conjunctive interpretation of disjunction. *Semantics and Pragmatics*, 3(11):1–28, 2010.

[18] Robert C. Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in Logical Theory*, pages 98–112. Blackwell, Oxford, 1968.

[19] William B. Starr. A uniform theory of conditionals. *Journal of Philosophical Logic*, 43(6):1019–1064, 2014.

[20] Anna Szabolsci and Bill Haddican. Conjunction meets negation: A study in cross-linguistic variation. *Journal of Semantics*, 21(3):219–249, 2004.

[21] Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.

[22] Ken Warmbrōd. Counterfactuals and substitution of equivalent antecedents. *Journal of Philosophical Logic*, 10(2):267–289, 1981.

[23] Malte Willer. Dynamics of epistemic modality. *Philosophical Review*, 122(1):45–92, 2013.

[24] Malte Willer. Indicative scorekeeping. *Proceedings of the 19th Amsterdam Colloquium*, pages 249–256, 2013.