Reasonable Inferences for Counterfactuals[1]

Ginger Schultheis

1 Introduction

Consider the following inference patterns.

Transitivity
$A \to B, B \to C \vDash A \to C$

Simplification
$A \vee C \to B \vDash A \to B$ and $C \to B$

Contraposition
$A \to B \vDash \neg B \to \neg A$

Antecedent Strengthening
$A \to B \vDash A \wedge C \to B$

Transitivity, Simplification, and Contraposition are intuitively compelling. Although Antecedent Strengthening may seem less attractive at first, close attention to the full range of data reveals that it too has considerable appeal.

An adequate theory of conditionals should account for these facts. The **strict theory** of conditionals does so by validating the four inferences.[2] It says that natural language conditionals are necessitated material conditionals: $\ulcorner A \to B \urcorner$ is true if and only if $\ulcorner A \supset B \urcorner$ is true throughout a set of accessible worlds. As a result, it validates many classical inferences, including Transitivity, Simplification, Contraposition, and Antecedent Strengthening. In what follows I will refer to these as the **strict inferences**.

The **variably strict theory** does not say that natural language conditionals are necessitated material conditionals: the set of worlds throughout which B must be true in order to make the conditional $\ulcorner A \to B \urcorner$ true depends partly on A—it *varies* from antecedent to antecedent. As a result, the variably strict theory invalidates many classical inference patterns, including all four strict inferences.

So the variably strict theorist faces a question: why do we find these inferences so compelling? My task in this paper is to suggest an answer on her behalf: that they are **reasonable inferences** in the sense introduced by Stalnaker (1975).

Given two compelling, widely acceptable principles—Or-to-If and If-to-Or—it follows that the strict inferences are reasonable for *indicatives*. A variable strict theory of the indicative,

[1]Thanks to Andrew Bacon, Melissa Fusco, Arc Kocurek, Anubav Vasudevan, Malte Willer, and especially, David Boylan and Matt Mandelkern for helpful feedback.

[2]See Warmbrod (1981), Veltman (1985), von Fintel (2001), Gillies (2007), and Gillies (2009) for defenses of strict theories.

like Stalnaker's, that secures these principles therefore predicts that the strict inferences are reasonable. I show all of this in §4, building on Stalnaker's passing note that his theory predicts that Transitivity and Contraposition are reasonable inferences.[3]

To my knowledge, no one has explored whether a variably strict theory can predict that the strict inferences are reasonable for *counterfactuals*. In §5, I show that given two plausible principles—counterfactual analogues of Or-to-If and If-to-Or—it follows that the strict principles are reasonable for counterfactuals. In §6, I sketch a variably strict theory of counterfactuals that secures these two principles, and therefore predicts that the strict inferences are reasonable.

I begin, in §2, by stating the strict and variably strict theories in a formal framework. In §3, I vindicate the intuitions that support the strict inferences for both indicatives and counterfactuals.

2 Two Theories

According to the strict theory, natural language conditionals are necessitated material conditionals. The flavor of necessity depends on the kind of conditional. For indicatives it is epistemic necessity. An indicative conditional $\ulcorner A > B \urcorner$ is true if and only if $\ulcorner A \supset B \urcorner$ is true throughout the epistemically possible worlds. For counterfactuals, the necessity is metaphysical necessity. A counterfactual $\ulcorner A \mathrel{\Box\!\!\rightarrow} B \urcorner$ is true if and only if $\ulcorner A \supset B \urcorner$ is true throughout the metaphysically possible worlds.

Let W be a non-empty set of possible worlds. Let R_c be a contextually-supplied, reflexive accessibility relation over W. Let '→' stand for both indicatives and counterfactuals. Then we state the strict theory as follows.

Strict Theory
$[\![A \to B]\!]^{c,w} = 1$ if and only if $R_c(w) \cap [\![A]\!]^c \subseteq [\![B]\!]^c$

Earlier I said that the strict theory validates Transitivity, Simplification, Contraposition, and Antecedent Strengthening. This needs to be qualified. Conditionals carry a **compatibility presupposition**: a conditional presupposes that there are accessible antecedent-worlds. Often this presupposition is formalized in a trivalent framework—sentences whose presuppositions are not satisfied are neither true nor false. On these theories, Simplification, Contraposition, and Antecedent Strengthening are merely Strawson-valid: if the premises are true and the conclusion is either true or false, the conclusion is true.[4] (Transitivity is still classically valid.) Others prefer a multidimensional treatment of presupposition: sentences are always true or false, and presupposition is an independent dimension of meaning.[5] On this theory, the strict inferences are classically valid. For ease of exposition, I'll assume a multidimensional treatment.

[3] See Stalnaker (1975).

[4] See von Fintel (1999, 2001). The term 'Strawson entailment' comes from Strawson (1952).

[5] See Herzberger (1973) and Kartunnen and Peters (1979).

I state the variably strict theory using a *selection function* f_c that takes a world w and an antecedent $[\![A]\!]^c$ to a set of worlds such that $f_c(w, [\![A]\!]^c) \subseteq R_c(w) \cap [\![A]\!]^c$.[6] Then:

Variably Strict Theory
$[\![A \rightarrow B]\!]^{c,w} = 1$ if and only if $f_c([\![A]\!]^c, w) \subseteq [\![B]\!]^c$

This says that $\ulcorner A \rightarrow B \urcorner$ is true at a world w if and only if B is true in all of the selected A-worlds, where the selected A-worlds are a subset of the accessible A-worlds.

The variably strict theory invalidates the strict inferences because the set of worlds throughout which B must be true in order for $\ulcorner A \rightarrow B \urcorner$ to be true—the set of selected A worlds—is a function of A. For example, consider Simplification. Let $R_c(w) = \{w_1, w_2, w_3\}$. Suppose that A is false in w_1 and w_2 and true in w_3. Suppose that B and C are true in w_1 and w_2 and that C is false in w_3. Let $f_c([\![A \vee B]\!]^c, w) = \{w_1, w_2\}$. And let $f_c([\![A]\!]^c, w) = \{w_3\}$. Then Simplification fails: $\ulcorner A \vee B \rightarrow C \urcorner$ is true in w, but $\ulcorner A \rightarrow C \urcorner$ is false. We can construct similar counterexamples to Transitivity, Contraposition, and Antecedent Strengthening.

3 Defending the Inferences

In §3.1 I present the prima facia case for all four of the strict inferences.[7] In §3.2, I respond to apparent counterexamples.

3.1 The Strict Inferences

Start with Transitivity. Consider:

(1) If Milo did not go to New York, he went to Boston.

(2) If Milo went to Boston, he saw the Red Sox play.

(3) Therefore, if Milo did not go to New York, he saw the Red Sox play.

This reasoning is flawless.

One more example. I am talking with Milo about seeding for the NBA playoffs. I say:

(4) If the Lakers win tonight, they will secure the fifth seed.

Milo tells me:

(5) If the Lakers secure the fifth seed, they will play the Warriors on Monday.

If I accept (5), then I will conclude:

[6] See Stalnaker (1968), Lewis (1973), McGee (1985), and Kratzer (1986) for classic variably strict theories.

[7] For defenses of Simplification, see Fine (2012) and Willer (2015). For a defense of Contraposition, see Warmbrod (1983), Gillies (2009), and Starr (2014). (Note that Gillies (2009) defends only the Strawon validity of Contraposition. He does say that it is classically valid. Starr defends only a limited version of Contraposition.) For defenses of Transitivity, see Warmbrod (1983) and von Fintel (2001). For defenses of Antecedent Strengthening, see von Fintel (2001), Gillies (2007), and Willer (2017).

(6) If the Lakers win tonight, they will play the Warriors on Monday.

Transitivity is equally compelling for counterfactuals. Consider our first example:

(7) If Milo had not gone to New York, he would have gone to Boston.

(8) If Milo had gone to Boston, he would have seen the Red Sox play.

(9) Therefore, if Milo had not gone to New York, he would have seen the Red Sox play.

And our second:

(10) If the Lakers had won tonight, they would have secured the fifth seed.

(11) If they had secured the fifth seed, they would have played the Warriors on Monday.

(12) Therefore, if the Lakers had won tonight, they would have played the Warriors on Monday.

Both inferences are impeccable.

Turn to Simplification. It also seems valid. Consider:

(13) If it rained or snowed, the picnic was cancelled.

If I accept (13), I am committed to both (14) and (15).

(14) If it rained, the picnic was cancelled.

(15) If it snowed, the picnic was cancelled.

Contraposing, if I reject either (14) or (15), I must also reject (13).

As we saw with Transitivity, Simplification is no less compelling for counterfactuals. Consider:

(16) If it had rained or snowed, the picnic would have been cancelled.

If I accept (16), I am committed to both (17) and (18).

(17) If it had rained, the picnic would have been cancelled.

(18) If it had snowed, the picnic would have been cancelled.

Contraposing, if I reject either (17) or (18), I must also reject (16).

Next, we have Contraposition. Suppose I dip my ring in a solution, and it will either turn green or red, depending on whether it is made of gold. You say:

(19) If the ring was made of gold, it turned green.

I conclude:

(20) Therefore, if it turned red, it wasn't made of gold.

Or suppose Milo and I are talking about how many minutes Kevin Durant played in last night's game. Milo tells me:

(21) If Durant played the whole first half, he didn't play the whole second half.

Then I am in a position to conclude:

(22) If Durant played the whole second, he didn't play the whole first.

Now, Contraposition is much less commonly used with counterfactuals than it is with indicative conditionals. But I don't think this alone gives us reason to doubt its validity. Counterfactuals tend to implicate that their antecedents and consequents are false, and so it's hard to find contexts in which $\ulcorner A \mathrel{\Box\!\!\rightarrow} B \urcorner$ and $\ulcorner \neg B \mathrel{\Box\!\!\rightarrow} \neg A \urcorner$ are both assertable. Suppose I say:

(23) If it had been made of gold, it would have turned green.

(23) strongly suggests that the ring is not made of gold and that it did not turn green. But then it would be strange to continue with

(24) If it hadn't turned green, it wouldn't have been made of gold.

since (24) strongly suggests that the ring *did* turn green.

There are, however, cases where counterfactuals do not carry this implicature, and in many of these cases Contraposition does seem like a good inference. Consider future-less-vivid conditionals. I think my ring is not made of gold, but I'm not sure. I dip the ring in the solution. I say:

(25) If it were made of gold, it would turn green by tomorrow.

You infer:

(26) If it were to turn red, it couldn't be made of gold.

This seems like a good inference. And note that (27) sounds incoherent:

(27) # If it were made of gold, it would turn green by tomorrow. But if it were to turn red, it could (still) be made of gold.

Finally, Antecedent Strengthening. On first blush, this principle does not seem as intuitively compelling as the others. Consider this **Sobel Sequence**.

(28) If Alice comes to the party, she will have a great time.

(29) But of course, if Alice and David come to the party, Alice won't have a great time.

This sequence of conditionals is unremarkable: both (28) and (29) can be true, it seems. But if Antecedent Strengthening were valid, (28) and (29) would be inconsistent.

Nevertheless, I have been convinced by strict theorists that there are strong reasons to accept Antecedent Strengthening. I will mention two.

The first reason is that Antecedent Strengthening follows from each of Simplification and Contraposition, given minimal background assumptions. I leave the proofs to a footnote.[8]

[8] *Proof that Simplification entails Antecedent Strengthening* Suppose $\ulcorner A \mathrel{\Box\!\!\rightarrow} B \urcorner$ is true. It follows from Substitution of Logical Equivalents that $\ulcorner (A \wedge C \vee A \wedge \neg C) \mathrel{\Box\!\!\rightarrow} B \urcorner$ is true. By Simplification, it follows that

The second is that, on closer examination, it becomes clear that Sobel sequences are not convincing counterexamples to Antecedent Strengthening.

Why not? That we judge (28) and (29) true threatens Antecedent Strengthening only if we judge them true in the same context. But—as von Fintel (2001) argues—we have reason to doubt that (28) and (29) are judged true in the same context. To see why, consider what happens when we reverse the order of the sentences.

(29) If Alice and David come to the party, Alice won't have a great time.

(28) # But of course, if Alice comes to the party, she will have a great time.

This **reverse Sobel sequence** sounds much worse than the original (forward) Sobel sequence. Once (29) has been asserted, it is no longer acceptable to continue with (28). (We're tempted to ask: what if David had come to the party?)

This suggests that the premises in the original (forward) Sobel sequence are not evaluated in the same context—they are not evaluated relative to the same accessibility relation.

Let me explain. As I said in §2, conditionals carry a compatibility presupposition: they presuppose that there are accessible antecedent-worlds. Hearers accommodate this presupposition. Suppose the presupposition $\ulcorner A \rightarrow B \urcorner$ is not satisfied in a given context: there are no accessible A-worlds, according to the contextually-supplied accessibility relation. Then hearers will choose a different accessibility relation to evaluate the conditional—one that yields a set of accessible worlds that is compatible with A.

This presupposition accommodation is asymmetric: it often demands expansion of the set of accessible worlds, but it never demands contraction. If there are no accessible A-worlds, hearers will accommodate by expanding the set of accessible worlds. If there are accessible A-worlds, the conditional's presupposition is satisfied—nothing needs changing.

This is what gives rise to the asymmetry between forward and reverse Sobel sequences. Consider the forward Sobel sequence. When we first encounter (28), we're ignoring the possibility that David comes to the party: we evaluate this sentence relative to a set of worlds where only Alice goes. This means that the presupposition of (29) is not satisfied. So, when (29) is asserted, we accommodate its presupposition: we evaluate (29) relative to a larger set of accessible worlds—one that includes some where both Alice and David come to the party. If Alice does not have a great time in any of these worlds, (29) comes out true in our new context—that is, relative to our new accessibility relation.

Now consider the reverse Sobel sequence. I assert (29) first. If my assertion is accepted, then (29)'s presupposition is satisfied—there are accessible worlds where Alice and David both come to the party. This, in turn, means that (28)'s presupposition is satisfied—there are accessible worlds where Alice comes to the party. We have no reason to choose a different accessibility

---

$\ulcorner A \wedge C \boxrightarrow B \urcorner$ is true.

Proof that Contraposition entails Antecedent Strengthening. Suppose $\ulcorner A \boxrightarrow B \urcorner$ is true. By Contraposition it follows that $\ulcorner \neg B \boxrightarrow \neg A \urcorner$ is true. That means that $\ulcorner \neg B \boxrightarrow \neg(A \wedge C) \urcorner$ is true. By another application of Contraposition it follows that $\ulcorner A \wedge C \boxrightarrow B \urcorner$ is true.

relation to evaluate (28): the context does not change. But if Antecedent Strengthening is valid, (28) and (29) cannot both be true in the same context.

(Note that, as I have presented things, reverse Sobel sequences figure in a purely defensive argument—a response to the claim that (forward) Sobel sequences are counterexamples to Antecedent Strengthening. But—as strict theorists point out—they can also figure in a powerful offensive argument in favor of Antecedent Strengthening. The argument is simple. Reverse Sobel sequences sound terrible. This is to be expected if Antecedent Strengthening is valid. This is not to be expected if Antecedent Strengthening is invalid: if (28) and (29) are consistent, why can't I assert (29) after (28)?)

3.2 Apparent Counterexamples

We have already discussed apparent counterexamples to Antecedent Strengthening. In this section, I turn to apparent counterexamples to Transitivity, Simplification, and Contraposition. I argue that the apparent counterexamples are merely apparent.

Let's start with an alleged counterexample to Transitivity from Stalnaker (1968).

(30) If Hoover had been a communist, he would have been a traitor.

(31) If Hoover had been Russian, he would have been a communist.

(30) and (31) seem true. But (32) does not.

(32) ? If Hoover had been Russian, he would have been a traitor.

If all three sentences are evaluated in the same context—that is, relative to the same accessibility relation—we have a counterexample to Transitivity. We have *Russian* $\Box\!\!\rightarrow$ *Communist*—that's (31). We have *Communist* $\Box\!\!\rightarrow$ *Traitor*—that's (30). Yet we do not have *Russian* $\Box\!\!\rightarrow$ *Communist*.

But I doubt that (30) and (31) are evaluated in the same context. To see why, consider what happens when we reverse the order of the premises.[9]

(31) If Hoover had been Russian, he would have been a communist.

(30) ? If Hoover had been a communist, he would have been a traitor.

When we encounter (31) first, we find it harder to accept (30). (We're tempted to ask: What if Hoover had been born Russian?)

As we saw with Antecedent Strengthening, this suggests that the premises in Stalnaker's example, when presented in their original order, are not evaluated in the same context—they are not evaluated relative to the same accessibility relation. But if they are not evaluated in the same context, we do not have a counterexample to Transitivity.

Here's an apparent counterexample to Simplification from McKay and van Imwagen (1977).

[9] This observation is due to von Fintel (2001).

(33) If Spain had joined the Axis or the Allies, it would have joined the Axis.

(33) is true. But now consider:

(34) If Spain had joined the Allies, it would have joined the Axis.

(34) does not sound true. It sounds like a contradiction.

I say (34) is true—or, more precisely, true in any context in which (33) is true and accepted.

Why, then, doesn't it seem true? When we evaluate and accept (33), we use a set of accessible worlds that does not include any where Spain joins the Allies: we assume that Spain couldn't have joined the Allies. But if Spain couldn't have joined the Allies, then the presupposition of (34) is not satisfied: there are no accessible worlds where its antecedent is true. (34) does not seem true, then, because it suffers from presupposition failure.[10]

(More generally, take any conditional of the form $\ulcorner A \rightarrow \neg A \urcorner$. Sometimes, this conditional is true. But it is only ever vacuously true—true because are no accessible A-worlds. If there are no accessible A-worlds, it suffers from presupposition failure, and so it never seems true.)

Similar things can be said about apparent counterexamples to Contraposition. Here's an example from Adams (1988). Consider:

(35) If it rains, it won't pour.

We can easily imagine situations in which I would accept (35). But now consider:

(36) If it pours, it won't rain.

(36) sounds like a contradiction.

I say that (36) is true—or, more precisely, true in any context in which (35) is true and known.

Why, then, doesn't it seem true? Consider any context in which we know (35). In any such context, there are no epistemically possible worlds where it pours—we know that it won't pour. But if we know that it won't pour, then the presupposition of (36) is not satisfied: there are no accessible worlds where its antecedent is true. (36) does not seem true, then, because it suffers from presupposition failure.

4 Indicative Conditionals

We have seen that Transitivity, Simplification, and Contraposition are intuitively compelling. Although Antecedent Strengthening may seem less attractive at first, close attention to the full range of data reveals that it too has considerable appeal.

An adequate theory of conditionals should account for these facts. The strict theory has a simple explanation: it says the strict inferences are valid. The variably strict theory should say that the strict inferences have some validity-*like* status—some property that makes arguments seem valid even when they aren't.

[10] For defenses of this explanation, see Warmbrod (1981), Fine (2012), and Starr (2014).

For indicative conditionals, there is a good candidate for what that property might be: Stalnaker's **reasonable inference**.[11] To say that an inference is reasonable is to say, roughly, that whenever the premises can be asserted, it is impossible for you to come to know those premises without also coming to know the conclusion of the inference. Here is a precise definition, where '$\Box_E$' stands for epistemic necessity.

> **Reasonable Inference**
> The inference from $\varphi_1, \varphi_2, \ldots, \varphi_n$ to ψ is a reasonable inference if and only if: for any context c, if $\varphi_1, \varphi_2, \ldots, \varphi_n$ can be felicitously asserted in c, and $\Box_E\varphi_1, \Box_E\varphi_2, \ldots,$ and $\Box_E\varphi_n$ are all true in c, then $\Box_E\psi$ is true in c.

I assume that the epistemic necessity operator $\Box_E$ is the operator denoted by the English epistemic necessity modal 'must'. Thus, to say the inference from φ_1, $\varphi_2, \ldots,$ and φ_n to ψ is a reasonable inference is to say that whenever the premises can be felicitously asserted in a given context c, and ⌜Must φ_1⌝, ⌜Must φ_2⌝, … and ⌜Must φ_n⌝ are true in c, ⌜Must ψ⌝is also true in c.

In §4.1, I show that, given two compelling, widely acceptable principles governing indicative conditionals, it follows that the strict inferences are reasonable for indicatives. In §4.2, I show that Stalnaker's variably strict theory secures these principles, and therefore predicts that the strict inferences are reasonable.

4.1 If-to-Or and Or-to-If

The two principles are:

> **If-to-Or**
> $A > B \vDash \neg A \vee B$

> **Boxy Or-to-If**
> $\Box_E(A \vee B) \vDash \Box_E(\neg A > B)$

If-to-Or is equivalent to Modus Ponens: to say that ⌜A > B⌝ entails ⌜¬A or B⌝ is to say that ⌜A > B⌝ and A jointly entail B. It goes without saying that Modus Ponens is a compelling principle. Consider:

(37) If the butler didn't do it, it was the gardener.

(38) The butler didn't do it.

(39) Therefore, it was the gardener.

If I accept (37) and (38) I have no choice but to accept (39).

Or-to-If is equally compelling. Consider:

[11] See Stalnaker (1975). Stalnaker himself observes that Contraposition and Transitivity are reasonable inferences for indicatives on his theory. In a similar spirit, Dorr & Hawthorne (ms) observe that the strict inferences are what they call *quasi-valid* for indicatives. These authors do not discuss counterfactuals.

(40) Either the butler or the gardener did it.

(41) Therefore, if the butler didn't do it, it was the gardener.

This inference seems excellent.

(One might wonder why I do not endorse a *non-Boxy* Or-to-If principle that says that $\ulcorner A \vee B \urcorner$ entails $\ulcorner \neg A > B \urcorner$. The reason is that if this stronger principle were valid, indicative conditionals would be equivalent to material conditionals. But the overwhelming consensus of contemporary philosophers and linguists is that indicative conditionals are not equivalent to material conditionals.)

Any theory that validates both If-to-Or and Boxy Or-to-If will predict that Transitivity, Simplification, Contraposition, and Antecedent Strengthening are reasonable inferences.

Here is the proof for Transitivity. Suppose that:

1. $\ulcorner \Box_E(A > B) \urcorner$ is true.
2. $\ulcorner \Box_E(B > C) \urcorner$ is true.

Given If-to-Or, 1 entails 3 and 2 entails 4:

3. $\ulcorner \Box_E(\neg A \text{ or } B) \urcorner$ is true.
4. $\ulcorner \Box_E(\neg B \text{ or } C) \urcorner$ is true.

It follows that:

5. $\ulcorner \Box_E(\neg A \text{ or } C) \urcorner$ is true.

Finally by Boxy Or-to-If, 5 entails 6:

6. $\ulcorner \Box_E(A > C) \urcorner$ is true.

The proofs for Simplification, Contraposition, and Antecedent Strengthening are similar.[12]

That the strict inferences are reasonable inferences explains why we find them so compelling: when we come to know the premises of these inferences on the basis of a successful assertion of those premises, we come to know their conclusions too.

[12] In the proof that Transitivity is a reasonable inference for indicatives, I do not need to assume that the premises are assertable. Transitivity is therefore *informationally valid* for indicatives: whenever the premises are known, the conclusion is known, too. The same goes for Simplification, Contraposition, and Antecedent Strengthening. Why, then, am I working with reasonable inference rather than informational validity? The reason is that the assertability condition is needed for counterfactuals: on the theory of counterfactuals that I develop in §6, the strict inference patterns are not informationally valid, but they are reasonable inferences.

4.2 A Stalnakerian Variably Strict Theory

I will now present a version of Stalnaker's variably strict theory of indicatives.[13]

Begin with a contextually-supplied epistemic accessibility relation E: $\mathsf{E}(w)$ is the set of worlds consistent with what's known, in w, by the conversational participants at the time of utterance. I will assume that E is reflexive, transitive, and symmetric.

A **Stalnakerian indicative selection function** f_{E} is a contextually-supplied function that takes a world and a proposition to a set containing at most one world. Then:

> **Stalnaker Semantics for Indicatives**
> $[\![\mathrm{A} > \mathrm{B}]\!]^{\mathsf{E},w} = 1$ if and only if $f_{\mathsf{E}}(w, [\![\mathrm{A}]\!]^{\mathsf{E}}) \subseteq [\![\mathrm{B}]\!]^{\mathsf{E}}$

We make four assumptions about f_{E}.

> **Success**
> $f_{\mathsf{E}}(w, [\![\mathrm{A}]\!]^{\mathsf{E}}) \subseteq [\![\mathrm{A}]\!]^{\mathsf{E}}$
>
> **Non-Vacuity**
> If $\mathsf{E}(w) \cap [\![\mathrm{A}]\!]^{\mathsf{E}} \neq \varnothing$, then $f_{\mathsf{E}}(w, [\![\mathrm{A}]\!]^{\mathsf{E}}) \neq \varnothing$
>
> **Minimality**
> If $w \in [\![\mathrm{A}]\!]^{\mathsf{E}}$, then $f_{\mathsf{E}}(w, [\![\mathrm{A}]\!]^{\mathsf{E}}) \subseteq \{w\}$.
>
> **Epistemic Accessibility Constraint**
> $f_{\mathsf{E}}(w, [\![\mathrm{A}]\!]^{\mathsf{E}}) \subseteq \mathsf{E}(w)$

Success secures the validity of Identity, the principle that $\ulcorner \mathrm{A} > \mathrm{A} \urcorner$ is always true. Non-Vacuity secures a form of Conditional Non-Contradiction: if A is epistemically live, $\ulcorner \mathrm{A} > \mathrm{B} \urcorner$ and $\ulcorner \mathrm{A} > \neg \mathrm{B} \urcorner$ cannot both be true.

Minimality secures the validity of If-to-Or. To see this, suppose $\ulcorner \mathrm{A} > \mathrm{B} \urcorner$ is true at w. There are two cases. If A is false, then $\ulcorner \neg \mathrm{A}$ or $\mathrm{B} \urcorner$ is true. If A is true, then the set of selected A-worlds is $\{w\}$. It follows from Stalnaker's Semantics that B is true at w, and so the disjunction $\ulcorner \neg \mathrm{A}$ or $\mathrm{B} \urcorner$ is again true.

What about the Epistemic Accessibility Constraint? Together with Success, it secures the validity of Boxy Or-to-If. To see this, suppose $\Box_{\mathsf{E}}(\mathrm{A} \vee \mathrm{B})$ is true. Then $[\![\neg \mathrm{A}]\!]^{\mathsf{E}} \cap \mathsf{E}(w) \subseteq [\![\mathrm{B}]\!]^{\mathsf{E}}$. Consider an arbitrary w' in $\mathsf{E}(w)$. Since E is reflexive, transitive, and symmetric, it follows that $\mathsf{E}(w) = \mathsf{E}(w')$. So, $[\![\neg \mathrm{A}]\!]^{\mathsf{E}} \cap \mathsf{E}(w) = [\![\neg \mathrm{A}]\!]^{\mathsf{E}} \cap \mathsf{E}(w') \subseteq [\![\mathrm{B}]\!]^{\mathsf{E}}$. By Success and the Epistemic Accessibility Constraint, we know that $f_{\mathsf{E}}([\![\neg \mathrm{A}]\!]^{\mathsf{E}}, w') \subseteq [\![\neg \mathrm{A}]\!]^{\mathsf{E}} \cap \mathsf{E}(w')$. It follows that $f_{\mathsf{E}}(w', [\![\neg \mathrm{A}]\!]^{\mathsf{E}}) \subseteq [\![\mathrm{B}]\!]^{\mathsf{E}}$. By Stalnaker's Semantics, it follows that $\ulcorner \neg \mathrm{A} > \mathrm{B} \urcorner$ is true in w'. Since w' was chosen arbitrarily, it follows that $\Box_{\mathsf{E}}(\neg \mathrm{A} > \mathrm{B})$ is true in w.

[13] This version of Stalnaker's theory is not original to me. It is very similar to Bacon (2015)'s theory of indicatives, as well to a more recent theory due to Dorr & Hawthorne (ms).

5 Counterfactuals

As we have seen, that the strict inferences are reasonable for indicative conditionals follows from two plausible principles governing indicatives: Boxy Or-to-If and If-to-Or. Likewise, that the strict inferences are reasonable for counterfactuals is derivable from two plausible principles governing counterfactuals: a counterfactual counterpart of Or-to-If and a counterfactual counterpart of If-to-Or. Letting $\Box_H$ stand for *It had to have been that…*, and '$\rightsquigarrow$' for reasonable inference, I state the principles below.

Counterfactual Or-to-If
$\Box_H(A \vee B) \vDash \neg A \mathrel{\Box\!\!\rightarrow} B$

Counterfactual If-to-Or
$A \mathrel{\Box\!\!\rightarrow} B \rightsquigarrow \Box_H(\neg A \vee B)$

In this section, I defend the principles, and I show that they together entail that the strict inferences are reasonable. In §6, I sketch a theory of counterfactuals and show that it secures Counterfactual Or-to-If and Counterfactual If-to-Or.

5.1 Counterfactual Or-to-If

Counterfactual Or-to-If says that if it had to have been that A or B, then if it hadn't been that A, it would have been that B. I assume that 'could have' is the dual of 'had to'. Thus,

(42) It had to have been in the attic.

is true if and only if

(43) It couldn't have not been in the attic.

is true. This means that we can also state Counterfactual Or-to-If using 'couldn't have': if it couldn't have been that A and ¬B, then if it had been that A, it would have been that B. I will refer to 'had to' and 'could have' as **counterfactual modals**.[14] (Note: I will assume that counterfactual modals obey an S5 modal logic.)

Counterfactual Or-to-If looks just as plausible as its indicative counterpart. Take an example adapted from Edgington (2008). We're hunting for a treasure. The organizer gives me a hint. He tells me it's either in the attic or the garden. I trust him. So I go to the attic and tell my partner to search the garden. I discover the treasure. "Why did you tell me to search the garden?" my partner asks. I reply:

(44) The treasure had to have been either in the attic or the garden. (The organizer told me it was in one of those places.)

My partner concludes:

[14] Dorr & Hawthorne (ms) independently observe that there is a close connection between counterfactuals and the modal 'could have' and defend a version of Counterfactual Or-to-If.

(45) If the treasure hadn't been in the attic, it would have been in the garden.

This inference seems excellent.

One more example. Suppose Matt arrives on time to a dinner at six. We're told he caught the bus at five. Doubting that leaving at five left him enough time to get here by six, you say:

(46) Matt couldn't have caught the bus at five and made it to dinner by six.

Trusting you, I conclude:

(47) If Matt had caught the bus at five, he wouldn't have made to the dinner by six.

Once again, this inference is impeccable.

5.2 Counterfactual If-to-Or

Counterfactual If-to-Or says that if a counterfactual $\ulcorner A \mathrel{\Box\!\!\rightarrow} B\urcorner$ is assertable in a given context and $\ulcorner \Box_E(A \mathrel{\Box\!\!\rightarrow} B)\urcorner$ is true in the context, then $\ulcorner \Box_E \Box_H (\neg A \vee B)\urcorner$ is also true in the context. That is to say, if $\ulcorner A \mathrel{\Box\!\!\rightarrow} B\urcorner$ is assertable and known by the conversational participants in a given context, then $\ulcorner \Box_H(\neg A \vee B)\urcorner$ is also known.[15] For example, suppose that (48) is assertable in our context.

(48) If Matt hadn't bought the first house, he would have bought the second.

Suppose further that we know (48). Then counterfactual Or-to-If says that we are also in a position to know:

(49) Matt had to have either bought the first house or the second.

To flesh out what Counterfactual If-to-Or amounts to, we need to say what the assertability conditions are for counterfactuals. A tempting first thought is that a counterfactual is assertable only if its antecedent is known to be false. But there are well known counterexamples to this generalization. Here's an example due to Anderson (1951). A patient enters the emergency room displaying symptoms of what the doctor suspects is arsenic poisoning. The doctor says:

(50) If the patient had taken arsenic, he would have been showing exactly these symptoms.

The doctor does not believe that the patient did not take arsenic—indeed, (50) is most naturally interpreted as evidence that the patient did take arsenic.

In light of examples like this, we should not say that a counterfactual is assertable only if its antecedent is known to be false.[16] But a weaker generalization is plausible—that a counterfactual is assertable only if its antecedent is not known to be true.[17] Notice that if the doctor and his

[15] One might wonder why I do not say that Counterfactual If-to-Or is valid. The reason is that doing so would make indicative conditionals strict conditionals. If both Counterfactual Or-to-If and Counterfactual If-to-Or were classically valid, we would have: $\ulcorner A \mathrel{\Box\!\!\rightarrow} B\urcorner$ is true if and only if $\ulcorner \Box_H(\neg A \vee B)\urcorner$. Those who reject the strict theory must reject the classical validity of Counterfactual If-to-Or.

[16] See Zakkou (2021) for a dissenting view.

[17] See von Prince (2019).

interlocutors know that the patient took arsenic, (50) is no longer acceptable. Consider:

(51) #He took arsenic. If he had taken arsenic, he would have been showing exactly these symptoms.

Similar things can be said about other examples:

(52) David is at the party.

(53) #If David had come to the party, he would have given a speech.

In the remainder of the paper, I assume for simplicity that this is the only licensing condition for counterfactuals. This will allow me to significantly streamline the exposition of the theory: if we say that a counterfactual is licensed if and only if its antecedent is not known, then we can treat Counterfactual If-to-Or as equivalent to the principle that whenever the counterfactual $A \boxrightarrow B$ is known, and its antecedent is not, $\ulcorner \Box_H(\neg A \text{ or } B)\urcorner$ is known. But I stress that the assumption is made purely for ease of exposition: none of the results will depend on it.

Why accept Counterfactual If-to-Or? Two arguments.

The first is that is never acceptable to assert the *if*-claim—that is, the counterfactual $\ulcorner A \boxrightarrow B\urcorner$—while denying the necessity of the *or*-claim—that is, while asserting that it could have been that A and $\neg$B. Take Edgington's treasure case. I say:

(45) If the treasure hadn't been in the garden, it would have been in the attic.

If you trust me and accept (45), you must also accept:

(54) The treasure couldn't have been hidden in the kitchen.

The conjunction (55) is completely unacceptable.

(55) #If the treasure hadn't been in the garden, it would have been in the attic. But it could have been in the kitchen.

The second argument is that Counterfactual If-to-Or follows from Duality, stated below. (Remember that '$\rightsquigarrow$' stands for reasonable inference.)

Duality
$A \boxrightarrow B \rightsquigarrow \neg(A \boxrightarrow \Diamond_H \neg B)$

The case for Duality is straightforward: conjunctions of the form $\ulcorner A \boxrightarrow B$ and $A \boxrightarrow \Diamond_H \neg B\urcorner$ are invariably defective.

(56) #If I had gotten an A on the exam, I would have passed the course. But if I had gotten an A on the exam, I could have failed the course.

(57) #If the treasure hadn't been in the garden, it wouldn't have been in the attic. But if it hadn't been in the garden, it could have been in the attic.

Duality and Counterfactual If-to-Or are closely related. Given Counterfactual Or-to-If, Duality entails Counterfactual If-to-Or. To see this, suppose:

1. $\ulcorner A \mathrel{\Box\!\!\rightarrow} B \urcorner$ is assertable.

2. $\ulcorner \Box_E (A \mathrel{\Box\!\!\rightarrow} B) \urcorner$ is true.

Suppose, for contradiction, that:

3. $\ulcorner \Diamond_E \Diamond_H (A \wedge \neg B) \urcorner$ is true.

Since counterfactual modals obey a logic of S5, it follows that:

4. $\ulcorner \Diamond_E \Box_H \Diamond_H (A \wedge \neg B) \urcorner$ is true.

(4) entails:

5. $\ulcorner \Diamond_E \Box_H (\Diamond_H \neg B) \urcorner$ is true.

And (5) entails:

6. $\ulcorner \Diamond_E \Box_H (\neg A \vee \Diamond_H \neg B) \urcorner$ is true.

By Counterfactual Or-to-If, it follows that:

7. $\ulcorner \Diamond_E (A \mathrel{\Box\!\!\rightarrow} \Diamond_H \neg B) \urcorner$ is true.

But by Duality, 1 and 2 entail that:

8. $\ulcorner \Box_E \neg (A \mathrel{\Box\!\!\rightarrow} \Diamond_H \neg B) \urcorner$ is true.

Contradiction.

5.3 The Strict Inferences

So I think we have strong reasons to accept both Counterfactual Or-to-If and Counterfactual If-to-Or. I'll now show that any theory that secures both inferences predicts that the strict inferences are reasonable for counterfactuals.

Here is the proof for Transitivity. Suppose that:

1. $A \mathrel{\Box\!\!\rightarrow} B$ is assertable.

2. $B \mathrel{\Box\!\!\rightarrow} C$ is assertable.

3. $\Box_E (A \mathrel{\Box\!\!\rightarrow} B)$ is true.

4. $\Box_E (B \mathrel{\Box\!\!\rightarrow} C)$ is true

It follows from the fact that Counterfactual If-to-Or is a reasonable inference that:

5. $\Box_E \Box_H$ (¬A or B)

6. $\Box_E \Box_H$ (¬B or C)

5 and 6 together entail:

7. $\Box_E \Box_H$ (¬A or C)

And given Counterfactual Or-to-If, 8 entails:

8. $\Box_E$(A $\boxright$ C)

That completes the proof that Transitivity is a reasonable inference. The proofs for Simplification, Contraposition, and Antecedent Strengthening are similar.

That the strict inferences are reasonable for counterfactuals explains why we find them so compelling: when we come to know the premises of these inferences on the basis of a successful assertion of those premises, we come to know their conclusions too.

6 A Sketch of A Theory

I have defended Counterfactual Or-to-If and Counterfactual If-to-Or, and I have shown that they jointly entail that the strict inferences are reasonable for counterfactuals. In this section, I sketch a theory of counterfactuals and show that it secures Counterfactual Or-to-If and Counterfactual If-to-Or.

I start in §6.1 by presenting a simplified version of the theory—a version that is not quite right. Although the simple theory validates Counterfactual Or-to-If, I show in §6.2 that it does not secure Counterfactual If-to-Or. In §6.3, I diagnose the problem, offer a more sophisticated **sequence semantics** for conditionals, and show that the resulting theory predicts that the strict inferences are reasonable for counterfactuals.

6.1 A Simple Theory

I assume that all of the differences between indicatives and counterfactuals are derived from differences in what is held fixed when we evaluate the conditional. When we evaluate indicative conditionals we hold fixed all of what we know. When we evaluate counterfactuals we hold fixed only some of what we know.

Let us make this more precise. Say that one accessibility relation R_1 is less informed than another accessibility relation R_2 if and only if $R_2(w) \subset R_1(w)$ for all w. That is to say, R_1 is less informed than R_2 if and only if, for any world w, the set of accessible worlds according to R_1 is larger than the set of accessible worlds according to R_2.

Let E be a contextually supplied epistemic accessibility relation used to evaluate indicative conditionals. Let *exp* be a contextually-supplied function that takes an epistemic accessibility relation and returns a less informed accessibility relation $exp(E)$ that we use to evaluate counterfactuals. (I will call $exp(E)$ a **counterfactual accessibility relation**.)

Then we will say that a counterfactual $\ulcorner A \mathrel{\Box\!\!\rightarrow} B \urcorner$ is true at a world w, relative to our information state E, just in case the corresponding indicative conditional $\ulcorner A > B \urcorner$ is true at w, relative to a contextually-determined counterfactual accessibility relation $exp(\mathsf{E})$.[18]

We have the following semantic entry.

Stalnakerian Semantics for Counterfactuals
$[\![A \mathrel{\Box\!\!\rightarrow} B]\!]^{\mathsf{E},w} = 1$ if and only if $f_{exp(\mathsf{E})}(w, [\![A]\!]^{exp(\mathsf{E})}) \subseteq [\![B]\!]^{exp(\mathsf{E})}$

To get a feel for how this works, take an example from earlier. We know Matt made it on time to the dinner at six. We're told he caught the bus at five. Doubting that leaving at five left him enough time to get here by six, you say:

(58) If Matt had left at five, he wouldn't have made it to the dinner on time.

To evaluate (58), we suspend some of our knowledge—our knowledge of the fact that Matt made it to the dinner on time, among other things. But we hold much of what we know fixed. In particular, we hold fixed much of our knowledge about what happened before the time of the events described in the antecedent. We hold fixed when Matt started to get dressed, which buses were running at that time, and so forth. What we do and do not hold fixed is represented by the accessibility relation $exp(\mathsf{E})$. If we're holding fixed facts about when Matt started to get dressed, then $exp(\mathsf{E})$ takes each world w to a set of worlds consistent with what we know, in w, about when he started getting dressed. If we're holding fixed facts about the bus schedules, then $exp(\mathsf{E})$ takes each world w to a set of worlds that is consistent with what we know, in w, about the bus schedules.

When we introduced Stalnaker's semantics for indicatives in §4, we stated four constraints on the selection function. When we're evaluating an indicative conditional, we use a selection function that is indexed to E, the accessibility relation representing what the conversational participants know. When we're evaluating a counterfactual, we use a selection function that is indexed to $exp(\mathsf{E})$, the counterfactual accessibility relation. Here are the four constraints stated in terms of $exp(\mathsf{E})$.

Success
$f_{exp(\mathsf{E})}(w, [\![A]\!]^{exp(\mathsf{E})}) \subseteq [\![A]\!]^{exp(\mathsf{E})}$

Non-Vacuity
If $exp(\mathsf{E})(w) \cap [\![A]\!]^{exp(\mathsf{E})} \neq \varnothing$, then $f_{exp(\mathsf{E})}(w, [\![A]\!]^{exp(\mathsf{E})}) \neq \varnothing$

Minimality
If $w \in [\![A]\!]^{exp(\mathsf{E})}$, then $f_{exp(\mathsf{E})}(w, [\![A]\!]^{exp(\mathsf{E})}) \subseteq \{w\}$.

[18] I am not the first to develop a theory along these lines. See Heim (1992) for the suggestion that counterfactuals are evaluated relative to (something like) an expansion of the epistemically possible worlds. See Schulz (2014) and Mackay (2019), and Schultheis (2023) for semantic entries that are very close to the theory presented in the main text.

Counterfactual Accessibility Constraint
$f_{exp(\mathsf{E})}(w, [\![\mathrm{A}]\!]^{exp(\mathsf{E})}) \subseteq exp(\mathsf{E})(w)$

Success secures the validity of Identity for counterfactuals: ⌜A □→ A⌝ is always true. Non-Vacuity secures a form of Counterfactual Non-Contradiction: if there are counterfactually accessible A-worlds, ⌜A □→ B⌝ and ⌜A □→ ¬B⌝ cannot both be true. Minimality ensures that a counterfactual ⌜A □→ B⌝ entails the material conditional ⌜¬A or B⌝, and therefore secures the validity of Modus Ponens.

The Counterfactual Accessibility Constraint says that the selected antecedent-world must be counterfactually accessible—it must be consistent with what we're holding fixed for the purpose of evaluating the counterfactual. Return to the dinner case. You say:

(58) If Matt had left at five, he wouldn't have made it to the dinner on time.

If we're holding fixed facts about when Matt started to get dressed, then, as I said earlier, $exp(\mathsf{E})$ takes each world w to a set of worlds consistent with what we know, in w, about when he started getting dressed. In that case, the Counterfactual Accessibility Constraint says that the selected antecedent-world must be consistent with what we know, in w, about when he started to get dressed. If we're holding fixed facts about the bus schedules, then the Accessibility Constraint says, for any world w, the selected antecedent-world must be consistent with what we know, in w, about the bus schedules.

I will assume a semantics for counterfactual modals that parallels our semantics for counterfactuals. Specifically, I say that the counterfactual modal claim ⌜It had to have been that A⌝ is true at a world w, relative to our information state E, just in case the epistemic modal claim ⌜It has to be that A⌝ is true at w, relative to $exp(\mathsf{E})$. (I assume that 'has to' has an epistemic interpretation on which it is synonymous with epistemic 'must'.)

Counterfactual Modals
$[\![\text{It had to have been that A}]\!]^{\mathsf{E},w} = 1$ if and only if $exp(\mathsf{E})(w) \subseteq [\![\mathrm{A}]\!]^{exp(\mathsf{E})}$

Given our assumption that counterfactuals and counterfactual modals are interpreted uniformly, we can show that the Counterfactual Accessibility Constraint, together with Success, secures the validity of Counterfactual Or-to-If. To see this, suppose $\Box_{\mathsf{H}}(\text{A or B})$ is true in w. By the semantics for counterfactual modals, it follows that $exp(\mathsf{E})(w) \subseteq [\![\text{A or B}]\!]^{exp(\mathsf{E})}$. So $[\![\neg\mathrm{A}]\!]^{\mathsf{E}} \cap exp(\mathsf{E})(w) \subseteq [\![\mathrm{B}]\!]^{exp(\mathsf{E})}$. By Success and the Accessibility Constraint, $f_{exp(\mathsf{E})}([\![\neg\mathrm{A}]\!]^{exp(\mathsf{E})}, w) \subseteq [\![\neg\mathrm{A}]\!]^{exp(\mathsf{E})} \cap exp(\mathsf{E})(w)$, and so $f_{exp(\mathsf{E})}([\![\neg\mathrm{A}]\!]^{exp(\mathsf{E})}, w') \subseteq [\![\mathrm{B}]\!]^{exp(\mathsf{E})}$. By the Stalnakerian Semantics fo Counterfactuals, it follows that ⌜¬A > B⌝ is true in w.

6.2 Counterfactual If-to-Or and Fine-Grained Contents

Let's recap. I have presented a simple theory of counterfactuals. I have shown that the theory secures Counterfactual Or-to-If. In this section, I will show that the simple theory does

not secure Counterfactual If-to-Or. I will diagnose the problem and, in §6.3, offer a more sophisticated theory that does secure the principle.

Counterfactual If-to-Or says that which counterfactuals we count as knowing in a given context depends, in part, on which 'could have' (or 'had to') claims we know. If, for all we know, Matt did not flip the coin, and for all we know, the coin could have landed heads, then it follows that for all we know it would have landed heads if it had been flipped.

But nothing we have said so far guarantees that this is so. Consider a simple model. There are three worlds: w_1, w_2, and w_3. In w_1 and w_2, Matt flips a fair coin. In w_1, it lands heads. In w_2, it lands tails. In w_3, he does not flip the coin. Suppose I know, in w_3, that he does not flip the coin: $\mathsf{E}(w_3) = \{w_3\}$. And finally suppose that all three worlds are counterfactually possible: $exp(\mathsf{E})(w_3) = \{w_1, w_2, w_3\}$.

This model is a counterexample to Counterfactual If-to-Or. To see why, first observe that our semantics for counterfactuals validates Conditional Excluded Middle.[19]

Conditional Excluded Middle
$\vDash A \boxright B \vee A \boxright \neg B$

Conditional Excluded Middle entails that one of the following counterfactuals is true in w_3.

(59) If Matt had flipped the coin, it would have landed heads.

(60) If Matt had flipped the coin, it would have landed tails.

Suppose that it is (59) that is true in w_3. Then, in w_3, I know (59): w_3 is the only world epistemically accessible from w_3. We may suppose that the counterfactual is assertable in our context: in w_3, I do not know that Matt flipped the coin, and we can assume that any other licensing conditions are satisfied. And yet, since w_2 is counterfactually accessible from w_3, (61) is true in w_3.

(61) The coin could have landed tails.

It follows that I do not know that the coin couldn't have landed tails. In summary, if (59) is true in w_3, we have:

1. ⌜Flip $\boxright$ Heads⌝ is assertable in w_3.
2. ⌜$\Box_{\mathsf{E}}$(Flip $\boxright$ Heads)⌝ is true in w_3.
3. ⌜$\Box_{\mathsf{E}}\,\Box_{\mathsf{H}}$ (¬Flip or Tails)⌝ is false in w_3.

Similarly, if it is (60) that is true in w_3, we have:

[19] My claim that the simple theory presented in §6.1 does not secure Counterfactual If-to-Or does not depend on Conditional Excluded Middle. Any variably strict theory says that $f(w_3, \text{Flip})$ excludes some counterfactually accessible worlds. Then—without the constraint I am about to propose—we should be able to construct a model in which I know the counterfactual Flip $\boxright$ Heads, the counterfactual is assertable, and yet I do not know that the coin couldn't have landed tails.

4. ⌜Flip $\Box\!\!\rightarrow$ Tails⌝ is assertable in w_3.

5. ⌜$\Box_E$(Flip $\Box\!\!\rightarrow$ Tails)⌝ is true in w_3.

6. ⌜$\Box_E$ $\Box_H$ (¬Flip or Tails)⌝ is false in w_3.

Either way, then, we have a counterexample to Counterfactual If-to-Or.

How do we rule out this model? We need a *plenitude* assumption: specifically, that if ¬A is epistemically possible, then every counterfactually possible A-world w is such that it is epistemically possible that if it had been that A, it would have been that w.

In the three-world model, the plenitude assumption fails: there are two few epistemic possibilities. In w_3, there is a counterfactually possible world where the coin lands heads (w_1), and a counterfactually possible world where the coin lands tails (w_2). But there is only *one* epistemic possibility. Either

(62) If the coin had been flipped, it would have landed heads.

is epistemically possible, or

(63) If the coin had been flipped, it would have landed tails.

is epistemically possible. But not both.

The problem of having too few epistemic possibilities is familiar in the literature on the probabilities of conditionals. Here is a simple way of seeing the problem that is due to Bacon (2015).[20] Suppose I roll a six-sided die, but I have not seen how it landed. You might have thought we can model my epistemic state with exactly six equiprobable worlds, one for each outcome of the roll. But, as Bacon explains, if we accept Conditional Excluded Middle, this model is inadequate. To see this, suppose the die lands on six. By Conditional Excluded Middle, one of the following conditionals is true.

(64) If it landed on four or five, it landed on four.

(65) If it landed on four or five, it landed on five.

But clearly I am in no position to know *which* of (64) or (65) is true. Bacon concludes that we must expand our simple six-world model. Specifically, we need to split the world in which the die lands on six into at least two worlds—one where (64) is true, and one where (65) is true. It is easy to see that we can generate many more epistemic possibilities by considering other antecedents that are false when the die lands on six. In short, we must countenance many more epistemic possibilities than our original six.

There is a similar flaw in our model of the coin flip. I said that there were three counterfactual possibilities: w_1 (where the coin lands heads), w_2 (where the coin lands tails) and w_3 (where the coin is not tossed). And I said that, in w_3, only w_3 is epistemically possible. (I know the

[20] See also Hájek (1989), and Khoo & Santorio (2018). My presentation follows Bacon (2015).

coin is not tossed.) But this can't be right. For as we have seen, Conditional Excluded Middle entails that one of (59) or (60), repeated below, is true in w_3.

(59) If Matt had flipped the coin, it would have landed heads.

(60) If Matt had flipped the coin, it would have landed tails.

But clearly I am in no position to know which of (59) or (60) is true. And so w_3 needs to be split into two epistemic possibilities—one where (59) is true, and one where (60) is true.

It is common for philosophers to model these more fine-grained possibilities with **sequences** of worlds, following van Fraassen (1976).[21] To see how these models work, return to the case of the die. Consider the possibility that the die lands on six. We have seen that there are many ways of settling the conditional facts that are compatible with the die landing on six—it could be that if the die didn't land on six, it landed on three, or it could be that if the die didn't land on six, it landed on two, and so forth. Each of these epistemic possibilities is modeled as a sequence of worlds. For example, consider the sequence:

$$\langle w_6, w_2, w_3, w_4, w_5, w_1 \rangle$$

This sequence represents one way of the settling all of the facts—both the non-conditional facts and the conditional facts. The first world tells us how the non-conditional facts have been settled—in this case, it tells us that the die landed on six. The other worlds in the sequence tell us how the conditional facts have been settled. For example, the second world tells us what is true if we are not in the first world. This sequence tells us that if the die didn't land on six, it landed on two. The third world tells us what is true if we are not in the first or the second world. This sequence tells us that if it didn't land on six or two, it landed on three. And so on.

In the next section, I suggest a generalization of van Fraassen's sequence semantics—one that provides a simple, uniform semantics for indicatives and counterfactuals. Roughly, I will say that an indicative conditional is true at a sequence just in case the first epistemically possible antecedent-world in the sequence is a consequent-world, and that a counterfactual is true at a sequence just in case the first counterfactually possible antecedent-world is a consequent-world. (Note: I will only consider simple conditionals—conditionals whose antecedents and consequents do not themselves contain conditionals.)

6.3 Sequence Semantics

Begin with a finite set of 'factual' worlds W. Let $\underline{S}_W$ be the set of all permutations of W. Thus, where the elements of W represent all possible ways of settling the non-conditional facts, the elements of $\underline{S}_W$ represent all possible ways of settling all of the facts—the non-conditional facts and the conditional facts. Where s is a sequence in $\underline{S}_W$ we will write w_s for the first world in s.

The semantics for non-conditional sentences is simple. A non-conditional sentence A is

[21] See, among others, Bacon (2014), and Goldstein & Santorio (2021).

true at a sequence s just in case A is true at w_s.

To state the semantics for modals and conditionals, I need to introduce some terminology. For any set of worlds A (any subset of W), we can **lift** A to a set of sequences $\underline{A}$ (a subset of $\underline{S}_W$) in the following way. (I use underlined uppercase letters for sets of sequences.)

Lifting for Sets
$\uparrow A = \{s \in S : w_s \in A\}$

This says that the $\uparrow A$ is the set of all sequences whose first world is in A. We can also **flatten** a set of sequences $\underline{A}$ (a subset of $\underline{S}_W$) to a set of worlds (a subset of W) as follows.

Flattening for Sets
$\downarrow \underline{A} = \{w_s \in W : s \in \underline{A}\}$

This says that $\downarrow \underline{A}$ is the set of first worlds of the sequences in $\underline{S}_W$/

To state the semantics for epistemic modals, let E be a contextually-supplied epistemic accessibility relation over W. Then:

Epistemic Modals
$[\![\text{Must A}]\!]^{E,s} = 1$ if and only if $\uparrow E(w_s) \subseteq [\![A]\!]^E$

$E(w_s)$ is the is the set of worlds epistemically accessible from w_s, the first world in s. $\uparrow E(w_s)$ is the set of sequences that begin with a world epistemically accessible from w_s. The semantics for epistemic modals says that ⌜Must A⌝ is true at s if and only if A is true at all of the sequences in $\uparrow E(w_s)$.

We turn now to conditionals. We need to redefine the selection function. Let R be a reflexive, transitive, and symmetric accessibility relation over W. Then:

$$f_R(s, \underline{A}) = \begin{cases} \{\text{the first } [R(w_s) \cap \downarrow \underline{A}]\text{-world in s}\} & \text{if } [R(w_s) \cap \downarrow \underline{A}] \neq \varnothing \\ \varnothing & \text{otherwise} \end{cases}$$

The selection function f_R takes a sequence s and a set of sequences $\underline{A}$ to a set containing at most one world: the singleton containing the first R-accessible world where $\downarrow \underline{A}$ is true if there are any such worlds, and the empty set otherwise.[22]

We can now state our Stalnakerian semantics for indicative conditionals. Let s be any sequence. Let E be a contextually-supplied epistemic accessibility relation over W. Then:

Stalnakerian Semantics for Indicatives
$[\![\text{A} > \text{B}]\!]^{E,s} = 1$ if and only if $f_E(s, [\![A]\!]^E) \subseteq \downarrow [\![B]\!]^E$

[22] Here I adopt a version of what Bacon (2014) calls *Harper's Constraint* on selection functions: $f_R(s, \underline{A}) = f_R(s, \underline{A} \cap R(s))$.

Putting this Stalnakerian Semantics together with our definition of the selection function tells us that ⌜A > B⌝ is true at a sequence s if and only if either there are no A-worlds that are E-accessible from w_s or the first A-world that is E-accessible from w_s is a B-world.

Let us check that all of the principles governing the logic of indicative conditionals discussed in §4 continue to hold.

Identity is valid: ⌜A > A⌝ is always true. This is because the selection function satisfies Success. Conditional Non-Contradiction is also valid: if A is epistemically live, then ⌜A > B⌝ and ⌜A > ¬B⌝ are not both true. This is because the selection function satisfies Non-Vacuity.

If-to-Or and Boxy Or-to-If are also valid, since the selection function satisfies versions of Minimality and the Epistemic Accessibility Constraint. This means that our sequence semantics predicts that the strict inferences are reasonable for indicative conditionals.

Turn now to counterfactuals and counterfactual modals. Where $exp(\mathsf{E})$ is a contextually-determined counterfactual accessibility relation over W, we state the semantics for counterfactual modals as follows.

Counterfactual Modals
$[\![\text{Had to have been that A}]\!]^{\mathsf{E},s} = 1$ if and only if $\uparrow exp(\mathsf{E})(w_s) \subseteq [\![\text{A}]\!]^{exp(\mathsf{E})}$

Remember, $exp(\mathsf{E})(w_s)$ is the is the set of worlds counterfactually accessible from w_s, the first world in s. And $\uparrow exp(\mathsf{E})(w_s)$ is the set of sequences that begin with a world counterfactually accessible from w_s. Our semantics for counterfactual modals says that ⌜Had to have been that A⌝ is true at s if and only if A is true at all of the sequences in $\uparrow exp(\mathsf{E})(w_s)$.

Finally, we have the following entry for counterfactuals.

Stalnakerian Semantics for Counterfactuals
$[\![\text{A } \Box\!\!\rightarrow \text{ B}]\!]^{\mathsf{E},\mathsf{s}} = 1$ if and only if $f_{exp(\mathsf{E})}(s, [\![\text{A}]\!]^{exp(\mathsf{E})}) \subseteq \downarrow [\![\text{B}]\!]^{exp(\mathsf{E})}$

This says that ⌜A $\Box\!\!\rightarrow$ B⌝ is true at a sequence s if and only if either there are no A-worlds that are $exp(\mathsf{E})$-accessible from w_s or the first A-world that is $exp(\mathsf{E})$-accessible from w_s is a B-world.

Let us now check that all of the principles governing the logic discussed in §5–6 hold.

It is easy to see that Identity and Counterfactual Non-Contradiction are both valid: the selection function satisfies Success and Non-Vacuity.

Counterfactual Or-to-If is also valid because the selection function satisfies the Counterfactual Accessibility Constraint.

The final order of business is to check that Counterfactual If-to-Or is a reasonable inference. I will give an informal, intuitive explanation in the main text, leaving the full proof of Counterfactual If-to-Or to a footnote.[23]

[23] *Proof of Counterfactual If-to-Or*. Let A and B be any two non-conditional sentences. Suppose (1) $[\![\Diamond_{\mathsf{E}} \neg \text{A}]\!]^{\mathsf{E},s} = 1$ and (2) $[\![\Box_{\mathsf{E}}(\text{A } \Box\!\!\rightarrow \text{ B})]\!]^{\mathsf{E},s} = 1$. Suppose, for contradiction, that (3) $[\![\Diamond_{\mathsf{H}}(\text{A and} \neg \text{B})]\!]^{s,\mathsf{E}} = 1$. (3) entails, by our semantics for counterfactual modals, that (4) there's an $s_1 \in \uparrow exp(\mathsf{E})(w_s)$ such that $[\![\text{A and} \neg \text{B}]\!]^{exp(\mathsf{E}),s_1} = 1$. Since A and B are non-conditional it follows that (5) $w_{s_1} \in exp(\mathsf{E})(w_s) \cap \downarrow [\![\text{A}]\!]^{exp(\mathsf{E})}$, and (6) $w_{s_1} \notin \downarrow [\![\text{B}]\!]^{exp(\mathsf{E})}$. Remember

Earlier I said that, in order to secure Counterfactual If-to-Or, we need a certain plenitude assumption: if it is epistemically possible that ¬A, then every counterfactually possible A-world *w* is such that it is epistemically possible that if it had been that A, it would have been that *w*. In the simple three-world model of the coin flip, this plenitude assumption failed: there were too few epistemic possibilities.

Our new sequence models respect the plenitude assumption. Why? Suppose that ¬A is epistemically possible. And suppose that *w* is a counterfactually possible A-world. Remember, the set of epistemically possible sequences is the set of all permutations of W beginning with an epistemically possible world. This means that, among the epistemically possible A-sequences, there will be some whose first A-world is *w*. And so, by the Stalnakerian semantics for counterfactuals, it will follow that it is epistemically possible that $\ulcorner \mathrm{A} \boxrightarrow w \urcorner$ is true.

Consider the coin case. I know the coin is also fair. There are counterfactually possible worlds where the coin is flipped and lands heads, and counterfactually possible worlds where it is flipped and lands tails. I also know that the coin was not tossed. The set of epistemically possible sequences is the set of all permutations of W beginning with an epistemically possible world. This means that there will be some epistemically possible sequences whose first world where the coin is flipped is one there it lands heads; and there will be some sequences whose first world where the coin is flipped is one where it lands tails. And so we predict—in accordance with Counterfactual If-to-Or—that I do not know that the coin would have landed tails had it been tossed, and I do not know that it would have landed heads, had it been tossed.

7 Conclusion

In the first part of the paper, we saw that, given two compelling, widely acceptable principles—Or-to-If and If-to-Or—it follows that the strict inferences—Transitivity, Simplification, Contraposition, and Antecedent Strengthening—are reasonable for indicatives. A variable strict theory of the indicative, like Stalnaker's, that secures these principles therefore predicts that the strict inferences are reasonable. In the second half of the paper, I turned my attention to counterfactuals. I showed that given two plausible principles—counterfactual analogues of Or-to-If and If-to-Or—it follows that the strict principles are reasonable for counterfactuals. I sketched a variably strict theory of counterfactuals that secures these two principles, and therefore predicts that the strict inferences are reasonable for counterfactuals.

---

that $\uparrow \mathsf{E}(w_s)$ is the set of all permutations of W beginning with a world in $\mathsf{E}(w_s)$. Given (1), it follows that (7) for some $s_2 \in \uparrow \mathsf{E}(w_s)$, w_{s_1} is the first $[exp(\mathsf{E})(w_s) \cap \downarrow [\![\mathrm{A}]\!]^{exp(\mathsf{E})}]$-world in s_2. Since $exp(\mathsf{E})$ is an equivalence relation, it follows that (8) w_{s_1} is the first $[exp(\mathsf{E})(w_{s_2}) \cap \downarrow [\![\mathrm{A}]\!]^{exp(\mathsf{E})}]$-world in s_2. It follows from (8) that (9) $f_{exp(\mathsf{E})}(s_2, [\![\mathrm{A}]\!]^{exp(\mathsf{E})}) = \{w_{s_1}\}$. By the Stalnakerian semantics for counterfactuals, it follows from (9) and (6) that (10) $[\![\mathrm{A} \boxrightarrow \mathrm{B}]\!]^{\mathsf{E},s_2} = 0$ And since $s_2 \in \uparrow \mathsf{E}(w_s)$, it follows that (11) $[\![\Box_{\mathsf{E}}(\mathrm{A} \boxrightarrow \mathrm{B})]\!]^{\mathsf{E},s} = 0$. But that contradicts (2). Therefore (3) must be false and so we conclude that (12) $[\![\Box_{\mathsf{H}}(\neg \mathrm{A} \text{ or } \mathrm{B})]\!]^{s,\mathsf{E}} = 1$. Since $\Box_{\mathsf{H}}$ obeys a logic of $\mathrm{S5}$ it follows that (13) $[\![\Box_{\mathsf{H}} \Box_{\mathsf{H}} (\neg \mathrm{A} \text{ or } \mathrm{B})]\!]^{\mathsf{E},s} = 1$. And since 'had to' entails 'has to' it follows that (14) $[\![\Box_{\mathsf{E}} \Box_{\mathsf{H}} (\neg \mathrm{A} \text{ or } \mathrm{B})]\!]^{\mathsf{E},s} = 1$.

8 References

Adams, Ernest (1975). *The Logic of Conditionals.* Dordrecht, Holland: D. Reidel.

Bacon, Andrew (2015). "Stalnaker's thesis in context." *Review of Symbolic Logic*, 8 (1):131-163.

Bledin, Justin (2014). "Logic Informed." *Mind*, 123 (490): 277-316.

Dorr, Cian & John Hawthorne (2018). *If...: A theory of conditionals.* Manuscript.

Edgington, Dorothy (1995). "On conditionals." *Mind*, 104 (414):235–329.

Gillies, Thony (2007). "Counterfactual Scorekeeping." *Linguistics and Philosophy*, 30 (3): 329–360.

Hájek, Alan (1989). "Probabilities of conditionals—revisited." *Journal of Philosophical Logic*, 18 (4):423 - 428.

Heim, Irene (1992). "Presupposition projection and the semantics of attitude verbs." *Journal of Semantics*, 9(3): 183–221

Heim, Irene (1994). "Comments on Abusch's Theory of Tense." In Hans Kamp (ed.), *Ellipses, Tense, and Questions*: 143-170

Herzberger, Hans (1973). "Dimensions of truth." *Journal of Philosophical Logic*, 2 (4): 535 - 556.

Iatridou, Sabine (2000). 'The Grammatical Ingredients of Counterfactuality.' *Linguistic Inquiry* 31(2): 231–270.

Ippolito, Michela (2013). 'Subjunctive Conditionals: A Linguistic Analysis.' *Linguistic Inquiry Monograph* (Series 65), Cambridge: MIT Press.

Karttunen, Lauri & Stanley Peters (1979). "Conventional implicature." *Syntax and Semantics*, 11: 1–56.

Kaufmann, Stefan (2009). 'Conditionals right and left: Probabilities for the whole family.' *Journal of Philosophical Logic* 38, pp. 1-53.

Khoo, Justin (2015). 'On Indicative and Subjunctive Conditionals.' *Philosopher's Imprint* 15, pp. 1-40.

Khoo, Justin (2022). 'The meaning of *If*.' Oxford University Press.

Kolodny, Niko & MacFarlane, John (2010). 'Ifs and Oughts.' *Journal of Philosophy* 107 (3):115-143.

Lewis, David (1973). *Counterfactuals*. Basil Blackwell Ltd., Malden, MA.

Lowe, E. J. (1995). 'The Truth about Counterfactuals.' *Philosophical Quarterly* 45 (178):41-59.

Mackay, John (2019). "Modal Interpretation of Tense in Subjunctive Conditionals." *Semantics and Pragmatics*, 12 (2)

Mandelkern, Matthew (2018). "Talking about worlds." *Philosophical Perspectives*, 32 (1): 298-325.

McGee, Vann (1989). 'Conditional probabilities and compounds of conditionals.' *Philosophical Review* 98, 485-541.

Partee, Barbara (1973). "Some structural analogies between tenses and pronouns in English." *Journal of Philosophy*, 70 (18): 601-609.

Schulz, Katrin (2014). 'Fake tense in conditional sentences: A modal approach.' *Natural Language Semantics*

Schulz, Moritz (2017). *Counterfactuals and Probability*. Oxford: Oxford University Press.

Stalnaker, Robert (1968). 'A Theory of Conditionals.' In Nicholas Rescher (ed.) *Studies in Logical Theory*(American Philosophical Quarterly Monographs 2), Oxford: Blackwell. pp. 98-112.

Stalnaker, Robert (1975). 'Indicative conditionals.' *Philosophia* 5, pp. 269-286.

von Fintel, Kai (2001). "Counterfactuals in a Dynamic Context." In Ken Hale: A Life in Language, Michael Kenstowicz, editor. MIT Press, Cambridge.

van Fraassen, Bas C. (1976). 'Probabilities of conditionals.' In *Foundations of probability theory, statistical inference, and statistical theories of science*, Springer, pp. 261 - 308.

von Prince, Kilu (2019). 'Counterfactuality and past.' *Linguistics and Philosophy* 42 (6):577-615.

Veltman, Frank (1985). "Logics for Conditionals." Doctoral Dissertation, University of Amsterdam.

Warmbrod, Ken (1981). "Counterfactuals and Substitution of Equivalent Antecedents." *Journal of Philosophical Logic*, 10: 267–289.