

Counterfactuals and Modality

Gabriel Greenberg, September 15, 2020

This essay calls attention to a set of linguistic interactions between counterfactual conditionals, on one hand, and possibility modals like *could have* and *might have*, on the other. These data present a challenge to the popular variably strict semantics for counterfactual conditionals. Instead, they support a version of the strict conditional semantics in which counterfactuals and possibility modals share a unified quantificational domain. I'll argue that pragmatic explanations of this evidence are not available to the variable analysis. And putative counterexamples to the unified strict analysis, on careful inspection, in fact support it. Ultimately, the semantics of conditionals and modals must be linked together more closely than has sometimes been recognized, and a unified strict semantics for conditionals and modals is the only way to fully achieve this.

Section 1 first outlines the rival strict and variably strict semantics for conditionals, then sketches the options for each with respect to the semantics of possibility modals. Sections 2 and 3 describe two patterns of modal-conditional interaction and examine their semantic implications; I argue that the unified strict analysis explains both patterns directly, but no such semantic explanation is available to the variably strict analysis. Section 4 considers possible pragmatic explanations of the data on behalf the variably strict analysis, but finds that they fall short. Section 5 discusses variations of the interaction data which at first appear to present counterexamples to the unified strict analysis, but ultimately end up supporting it. Section 6 takes stock of the essay's conclusions.

1 Counterfactuals and Modals

I begin with the semantics of subjunctive or **counterfactual conditionals** (for short, **conditionals**). I represent conditionals symbolically as $\phi > \psi$, corresponding to sentences of roughly the following form:¹

- (1) If it had been/were the case that ϕ , it would have been/would be the case that ψ .

While a number of theories of conditionals have been proposed over the last fifty years, two semantic theories in particular have distinguished themselves. Today, the orthodox view is the **variably strict analysis** (Stalnaker 1968; D. Lewis 1973; Kratzer 1981), while the chief alternative is the **strict conditional analysis** (Lowe 1995; von Stechow 2001; A. Gillies 2007).² I'll refer to these

Acknowledgements: This essay has benefited from years of conversations with teachers and classmates at Rutgers University and students and colleagues at UCLA, as well as the comments of anonymous reviewers, an audience at the 2009 CEU Summer School on Conditionals, and the METH Philosophy of Language Group. Special thanks to Maria Bittner, Sam Cumming, Jeff King, Karen Lewis, Jessica Rett, Jason Stanley, Will Starr, Una Stojnić, and Matthew Stone.

¹I assume that there is some relatively systematic way of identifying counterfactual conditionals, and that, at some legitimate level of abstraction, counterfactual conditionals can be translated by a sentence of propositional logic containing two sub-clauses and a single logical connective. For discussion, see e.g. Kratzer 1986 and Lycan 2001, ch. 1.

²The strict view has ancient roots, perhaps originating with the Stoic philosopher Chrysippus in the third century B.C.E.

as the **variable** and **strict** analyses respectively.³

The variable analysis begins by positing a “similarity” or “closeness” \geq_i over worlds, defined relative to the world of evaluation i . To state the analysis, I’ll adopt an extension of Stalnaker’s (1968, p. 46) idea of a contextually determined **selection function**; here, the selection function f takes as arguments a sentence ϕ , and an evaluation world i , and returns the set of worlds closest to i , relative to \geq_i , at which ϕ is true. Then the variable analysis claims that the conditional $\phi > \psi$ is true at i iff all worlds in $f(\phi, i)$, the set of closest ϕ -worlds, are ψ -worlds. On behalf of the variable theorist, I’ll assume that the selection function normally returns a set containing more than one world (perhaps infinitely many); in other words *Uniqueness* is false. I’ll also make the *Limit Assumption*: for any sentence ϕ , and closeness ordering \geq_i , if there are ϕ -worlds ordered by \geq_i , then there is at least one closest ϕ -world according to \geq_i .⁴

Meanwhile, the strict analysis employs the familiar notion of a set of **accessible** worlds. Here I use a contextually-determined accessibility function s , which maps worlds to sets of accessible worlds. For both analyses, I’ll assume that the relevant contextual elements— closeness ordering or accessible worlds— are determined by a context c relative to which truth is evaluated. I’ll use $\llbracket \cdot \rrbracket$ as an interpretation function, from sentences to sets of worlds. To simplify for present purposes, I’ll assume that this function only applies to non-modal sentences, is not relativized to worlds or contexts.⁵

Strict analysis

$$\lceil \phi > \psi \rceil \text{ is true}_{i,c} \quad \text{iff} \quad \forall w : w \in s(i) \cap \llbracket \phi \rrbracket \rightarrow w \in \llbracket \psi \rrbracket$$

Variable analysis

$$\lceil \phi > \psi \rceil \text{ is true}_{i,c} \quad \text{iff} \quad \forall w : w \in f(\phi, i) \rightarrow w \in \llbracket \psi \rrbracket$$

The two analyses differ because the set of accessible antecedent-worlds comes apart from the set of *closest* antecedent-worlds. Even if we restrict the ordering of worlds to accessible worlds,

(Sanford 2003); it was first defended in the 20th century by C.I. Lewis (1914) and C.S. Pierce (1933). Contemporary, context-sensitive versions of the theory have been developed by Warmbröd (1981a,b), Lycan (1984, 2001), Lowe (1983, 1995), von Fintel (2001), and A. Gillies (2007).

³For an overview of the debate, see Moss 2012, §1, §2, W. Starr 2019, §2. Besides the variable and strict analyses of counterfactual conditionals, there are a variety of other competing accounts, including probabilistic analyses like those of Edgington 2003, Kvart 1986, Loewer 2007, or Leitgeb 2012; cotenability theories such as Goodman 1947, Daniels and Freeman 1980; or the arbitrary-selection approach of Schulz 2014. Unfortunately, discussion of these alternatives is beyond the scope of this essay.

⁴See W. Starr (2019, §2.3, §B) for discussion and references.

⁵The definitions of the strict and variable analyses offered here constitute the semantic core of each view, but neither is limited to this core. A variety of additional constraints are commonly imposed. For example, besides Uniqueness and the Limit Assumption, discussed above, various constraints are normally added to validate *Modus Ponens*. See W. Starr (2019, §2.3, §B). Further, by articulating the semantics in terms of truth-conditions here and throughout, I do not mean to exclude presuppositional or dynamic aspects of modal semantics. I’ll discuss these, along with additional pragmatic and discourse-based amendments as the need arises.

the strict conditional always quantifies over the full set of accessible antecedent-worlds, while the variable analysis would only quantify over the closest antecedent-worlds among the accessible ones. It is this flexible departure from the set of accessible worlds which is distinctive of the variable analysis, but also the feature I will contest in the discussion to come.

Counterfactual conditionals belong to a broader linguistic class of counterfactual (or subjunctive) modals, including might-counterfactuals, as well as “bare” (unembedded) possibility and necessity modals. My focus in what follows is a set of interactions between counterfactual conditionals and bare possibility modals. For present purposes, **counterfactual bare possibility modals** (for short, **modals**) will be represented symbolically as $\Diamond\phi$, corresponding to sentences of the form:

- (2) It could/might have been the case that ϕ .

I’ll refer to ϕ as the **prejacent** of the sentence $\Diamond\phi$. Note that the intended counterfactual reading of (2) can be distinguished from an alternative past-epistemic reading, because only the counterfactual reading is compatible with denial of the prejacent: “It could have happened, but it didn’t” (W. B. Starr 2014a, p. 1023).⁶

As a baseline, I’ll assume a widely accepted approach to the semantic analysis of modals that understands them to contribute a quantifier over possible worlds with existential force (Kripke 1963; Kratzer 1977, 2012). The simplest version of this account holds that $\Diamond\phi$ is true at i if and only if there is some ϕ -world among the set of accessible worlds relative to i . Here I’ll let k be a contextually-selected function from worlds of evaluation to sets of accessible worlds.

k -Semantics

$$\ulcorner \Diamond\phi \urcorner \text{ is true}_{i,c} \text{ iff } \exists w : w \in k(i) \cap \llbracket \phi \rrbracket$$

The k -semantics on its own does not enforce any particular relation between the domain of modal quantification and the domain of conditional quantification, so predicts no substantive logical interactions between modals and conditionals, on either theory of the conditional. Although this may be the null hypothesis, we’ll soon consider data which suggests a rich interface between the two classes of expression, and invites a more integrated approach to modal and conditionals semantics. There are any number of ways this might be achieved in principle, but here I will focus on some of the most natural alternatives discussed by advocates of the variable and strict approaches respectively.

I begin with the variable analysis. The first approach corresponds to the semantic analyses of \Diamond proposed by Stalnaker (1968, p. 47), Stalnaker and Thomason (1970, p. 25), and D. Lewis

⁶Another diagnostic is that, on the epistemic reading, synonymy is preserved if *could have* / *might have* is replaced by *may have*, but not so on the counterfactual reading.

(1973, pp. 22–23).⁷ This analysis holds that a possibility modal is true at i just in case there is at least one closest prejacent-world relative to i :

$$(3) \quad \ulcorner \Diamond \phi \urcorner \text{ is true}_{i,c} \text{ iff } \exists w : w \in f(\phi, i)$$

Since this condition only requires that there be *some* closest ϕ -world, but puts no constraints on *how* close, (3) is equivalent to saying that among the worlds ordered by \geq_i , there is at least one ϕ -world (D. Lewis 1973, p. 22). I'll use the expression $\max(\geq_i)$ to denote the maximal sphere of ordered worlds, understood here as the set of all worlds ordered by \geq_i . Then we can state the proposed semantics as the condition that there is at least one prejacent-world in $\max(\geq_i)$. This analysis, in turn, is equivalent to the k -semantics plus the constraint that, for any i, c : $k(i) = \max(\geq_i)$.

max-Semantics

- (i) $\ulcorner \Diamond \phi \urcorner \text{ is true}_{i,c} \text{ iff } \exists w : w \in k(i) \cap \llbracket \phi \rrbracket$
- (ii) $k(i) = \max(\geq_i)$

Beyond the k - and \max -semantics, Kratzer (1981, 1991) describes a range of different possibility modals which vary in strength, each defined by a different relation between the background ordering and the quantified domain of the modal. Of particular relevance here is Kratzer's analysis of the modality she terms "possibility," and associates with *might* in English (1991, p. 644).⁸ Whereas the \max -semantics aligns the modal with maximal set of worlds, Kratzer's proposes to define bare possibility modals by focusing on the *minimal* set of worlds, the worlds closest to world of evaluation.⁹ Such a possibility modal is true at i if and only if there are prejacent-worlds among the accessible worlds which are closest *simpliciter* to the world of evaluation i . I'll refer to the set of worlds which are closest to i as $\min(\geq_i)$, that is, the minimal set of worlds relative to \geq_i .¹⁰ For commensurability, I likewise define the *min*-semantics as a constraint on a k -semantics:

⁷It is not entirely clear whether Stalnaker and Lewis intended their \Diamond , whose semantics they define, to be the translation of specific bare possibility modals in natural language, as the latter are not explicitly discussed in their original presentations. I will treat the various analyses of possibility modals as a theoretical option for these authors, rather than firm commitments.

For the analysis in question, Stalnaker (1968, p. 47) defines $\Diamond \phi$ as a translation of $\neg(\phi > \neg\phi)$, and adopt a variable analysis of the conditional. Thus $\Diamond \phi$ is true at i iff it is not the case that the world in $f(\phi, i)$ is a $\neg\phi$ -worlds. But worlds in $f(\phi, i)$ could only be $\neg\phi$ -worlds if there are no ordered ϕ -worlds, in which case $f(\phi, i)$ returns the absurd world, which verifies every sentence (p. 46). Thus $\Diamond \phi$ is true iff $f(\phi, i)$ returns a non-absurd world. Transposed into the present framework, this would mean that $\Diamond \phi$ is true iff $f(\phi, i)$ is not empty. D. Lewis (1973, p. 22) reaches essentially the same analysis by similar means.

Note finally that Stalnaker and Thomason (1970, pp. 27–29) do invoke an accessibility relation to impose a background constraint on ordering, but they do not directly define the semantics of possibility modals in terms of accessibility, so I have omitted the complication here. D. Lewis (1973, pp. 7–10) conceptualizes his system of spheres as a system of *spheres of accessibility*, so the conceptual elision is built-in.

⁸Kratzer (1981, §3) call this modality "human possibility;" Kratzer (1991, p. 644) and Kratzer (2012, p. 40) refer to it simply as "possibility."

⁹D. Lewis (1973, pp. 30–31) anticipates this analysis, characterizing it as an "inner" modality, in contrast with the "outer" modality of the \max -analysis. Lewis (p. 31) credits Sobel with the observation of the two strengths of modal are definable in terms of the counterfactual.

¹⁰In Kratzer's (1981, 1991) system modals are defined by more than one parameter, including both a "modal base"

***min*-Semantics**

- (i) $\lceil \Diamond \phi \rceil$ is $\text{true}_{i,c}$ iff $\exists w : w \in k(i) \cap \llbracket \phi \rrbracket$
- (ii) $k(i) = \min(\geq_i)$

Although the distinction between *k*, *max*, and *min* semantics is hardly exhaustive, it will serve as an instructive choice point, capturing centrally divergent visions for how possibility modals and conditionals semantically interact in the variabilist framework.

In the context of the strict analysis, there is no semantically specified ordering over worlds, so neither *max* nor *min* analyses are directly applicable. Instead, following von Fintel (2001, fn. 15) and A. Gillies (2007, §6), it is natural to suppose that modals and strict conditionals quantify over the same domain of counterfactual possibility. Let us call this the **unified strict analysis**; it is this version of the strict analysis which I'll be arguing for here. Formally, the domain for modals, $k(i)$, is simply identified with the conditional domain $s(i)$. Thus conditionals express restricted universal quantifiers over the set of accessible worlds, while possibility modals express existential quantifiers over the same set of accessible worlds.

Unified strict analysis

- $\lceil \phi > \psi \rceil$ is $\text{true}_{i,c}$ iff $\forall w : w \in k(i) \cap \llbracket \phi \rrbracket \rightarrow w \in \llbracket \psi \rrbracket$
- $\lceil \Diamond \phi \rceil$ is $\text{true}_{i,c}$ iff $\exists w : w \in k(i) \cap \llbracket \phi \rrbracket$

In the unified strict analysis, conditionals and modals are conceived as devices of quantification that are keyed to a common domain. This idea complements the already prominent view that indicative conditionals and epistemic modals form a unified semantic class in a similar way.¹¹ And it follows the pattern of explicit quantifiers in English, like nominal quantifiers (*a(n)*, *some*, *no*, *all*), temporal quantifiers (*always*, *sometimes*, *never*), or personal quantifiers (*everybody*, *somebody*). Each of these involves a class of quantificational expressions, each of which is semantically tuned to a common, contextually-selected domain of quantification. According to the unified strict analysis, counterfactual modals are no different.

In defending the unified strict analysis, I assume that modals and conditionals are embedded in a system of contextual update in discourse, along the lines of von Fintel (2001) and A. Gillies (2007). As von Fintel and Gillies have shown, such a system allows a strict analysis to accommodate much of the challenging data which originally motivated the variable analysis over the strict

and an “ordering source” (in the present nomenclature, an accessibility function and an ordering on worlds). Although the presentation here does contain two parameters, they are simply identified, so do not allow for the range of relations allowed by Kratzer. While that additional expressive flexibility may be important for defining the range of modal expressions, it won't play a central role in the argument to come.

¹¹See, e.g. Yalcin 2007; A. Gillies 2010; W. B. Starr 2014b; Willer 2017. Nearly all of the data presented here in Sections 2 and 3 can be recapitulated for indicative conditionals. It is a separate question whether this data militates in the same way for the unified strict approach to indicatives. I discuss this issue in footnote 24.

analysis. For present purposes it doesn't matter if this system is built into the modal semantics, pragmatics, or some other means of discourse management.

2 *Impossibility Above*

I begin with a class of cases where the quantificational domain of modals appears to put constraints on that of conditionals. After reviewing the data, I'll show how both the unified and variable analyses have compelling semantic explanations of the evidence available to them. In the next section, I turn to more challenging phenomena.

2.1 *Impossibility Above: the evidence*

Consider the inference exhibited below.¹²

- (4) a. Though she stayed home, Isabel could have gone sailing Tuesday afternoon.
 b. But she couldn't have gone sailing and not brought a hat.
 c. So if she had gone sailing, she would have brought a hat.

Line (a) introduces the scene and ensures that the subsequent modals receive a counterfactual, rather than epistemic, reading. The object of interest is the transition from (b) to (c), that is, from a sentence of the form $\neg\Diamond(sail \wedge \neg hat)$ to one of the form $sail > hat$. It seems that (c) can be legitimately inferred from (b), and this impression is reinforced by the felicitous use of the inferential marker *so* at the opening of (c). Other inferential markers like *therefore*, *thus*, *hence*, and *it follows that* work just as well.

I propose that this example conforms with a general pattern of linguistic inference, which I call *Impossibility Above* (IA), displayed below:¹³

$$\text{Impossibility Above (IA)} \quad \frac{\neg\Diamond(\phi \wedge \neg\psi)}{\phi > \psi}$$

My claim for now is that IA is simply an **inference pattern**, a schema for sequences in which the second sentence seems to “follow from” the first. By this I don't mean, necessarily, that IA is a *logical entailment*, though I will ultimately come to that conclusion. The inferential force in question may be deductive, ampliative, some type of pragmatic inference, or a purely epistemic inferential relation. What the linguistic evidence most immediately reveals, I believe, is that a speaker who asserts the premise is thereby *committed* (in some sense) to the truth of the conclusion, and one who

¹²Unless otherwise noted, the reader should assume that all examples in this essay are spoken by the same individual, and that this individual is speaking in some kind of deductive context – either reasoning to herself or trying to convince an interlocutor of her conclusions.

¹³Williamson (2007, p. 156) discusses this inference pattern, considered as a logical principle, under the name “NECESSITY”, formulated in terms of a necessity modal scoped over a material conditional: $\Box(\phi \rightarrow \psi) \rightarrow \phi > \psi$.

denies the conclusion is similarly committed to the denial of the premise. I will say that, according to IA, the premise **implies** the conclusion, in a similarly neutral sense of *imply*.

A second set of data provides further evidence that IA corresponds to a genuine inference pattern. These are cases which involve a stark *infelicity* when IA is violated. When reading these examples, it is important to imagine the speaker using a relatively “flat” affect, and to resist the temptation to substitute the contrastive discourse marker *but* for the discourse conjunctions *and* or *what’s more* used in the examples below. The significance of this constraint is discussed below and in Section 5; the point, for now, is to avoid unintentional shifts to the relevant modal context.

- (5) a. Norbert could have gone home over the holidays.
b. But he couldn’t have gone home and met the deadline.
c. * What’s more, if he had gone home, he would have met the deadline.
- (6) a. Norbert could have gone home over the holidays.
b. If he had gone home, he would have met the deadline.
c. * And he couldn’t have gone home and met the deadline.

This type of example suggests that it is infelicitous to utter sentences of the form $\neg\Diamond(\phi \wedge \psi)$ and $\phi > \psi$ in conjunction. This is directly explained by IA: for by IA, $\neg\Diamond(\phi \wedge \psi)$ implies $\phi > \neg\psi$; further, assuming the antecedent is possible in context, then on either semantics of the conditional, $\phi > \neg\psi$ entails $\neg(\phi > \psi)$; so $\neg\Diamond(\phi \wedge \psi)$ implies $\neg(\phi > \psi)$. But (5) and (6) above involve asserting the premise and the negation of the conclusion in conjunction, which we should expect to yield infelicity.

2.2 *Impossibility Above: semantic explanations*

What explains the inferential force of IA? The unified strict analysis has a straightforward explanation, for according to the unified strict analysis, IA is simply a logical entailment. For if it is true at a world i that $\neg\Diamond(\phi \wedge \neg\psi)$, then there are no accessible $\phi \wedge \neg\psi$ -worlds. Thus all accessible ϕ -worlds (if any) are ψ -worlds; so, by the unified semantics of conditionals, $\phi > \psi$ must be true at i as well.¹⁴ The patterns of felicitous inference highlighted above reflect logical entailment; the

¹⁴A qualification: recent advocates of the strict analysis have proposed that conditionals presuppose the compatibility of their antecedents with the possibilities made available by the context of evaluation (von Stechow 1998, p. 2, 2001, pp. 15–20; A. Gillies 2007, pp. 333–34). In the framework of this essay, this means that conditionals presuppose that their antecedents are compatible with the set of accessible worlds. If this principle is correct, and assuming presupposition failure leads to change in semantic value, then only a slightly weaker principle than IA is viable, namely:

$$\text{Weak Impossibility Above} \quad \frac{\neg\Diamond(\phi \wedge \neg\psi) \quad \Diamond\phi}{\phi > \psi}$$

But this does not disrupt the argument of the essay. The data which I cited as conforming with IA supports Weak IA just as well, and the unified strict analysis explains the inferential force of both by making both semantically valid.

patterns of infelicitous assertion reflect that assumption that it is generally infelicitous to assert contradictions.

What of the variable analysis? As I presented the *k*-semantics for possibility modals, there is no built-in alignment between the semantics for possibility modals and the variable semantics for conditionals. The absence of constraint here gives rise to models which straightforwardly violate IA. The *min*-semantics does introduce a constraint, but it does not validate IA. For, in a model where there are no ϕ -worlds among $\min(\geq_i)$, then the premise can be true (there are no $\phi \wedge \neg\psi$ -worlds among the closest worlds), but the conclusion false (among the closest ϕ -worlds, not all are ψ -worlds). Only the *max*-semantics validates IA in the variabilist framework.

The *max*-semantics works because it sets the quantified domain of the modal as wide as possible relative to that of the conditional. The lesson generalizes: IA requires that the domain of worlds quantified by the possibility modal must be equally wide or wider than the domain of worlds quantified by the conditional. Such a requirement is a natural extension of the variabilist framework, reflecting the intuitively plausible idea that all the worlds relevant to the evaluation of a conditional must be considered possible in the context in which it is expressed. Formally, any set of commitments which imply the following constraint will validate IA for the variabilist:

Accessibility

For any $i, c : \max(\geq_i) \subseteq k(i)$.

However, this assumption makes trouble when combined with the *min*-semantics. The *min*-semantics requires that the set of accessible worlds is the minimal set. Accessibility requires that the maximal set be a subset of the accessible worlds. The two constraints can only be satisfied together when the minimal set *is* the maximal set.¹⁵ And this, only when the ordering on worlds is “flat”: all worlds are either maximally close or unordered altogether. The result is a notational variant of the unified strict analysis, since it predicts that modals and conditionals draw from the same unordered domain of worlds. The non-monotonic behavior of the variably strict conditional is lost (cf. A. Gillies 2007, pp. 341–42).

In what follows, I will assume that variabilists wish to validate IA. The inference itself is compelling, and the philosophical rationale for Accessibility is credible. Going forward, this means setting aside the *min*-semantics, at least for present purposes, and adopting either (i) a combination of Accessibility with the *k*-semantics, or (ii) the *max*-semantics.

¹⁵The *min*-semantics assumes that $k(i) = \min(\geq_i)$; by Accessibility, then, $\max(\geq_i) \subseteq \min(\geq_i)$; but since, by definition, $\min(\geq_i) \subseteq \max(\geq_i)$, it follows that $\max(\geq_i) = \min(\geq_i)$.

3 *Impossibility Below*

I turn next to linguistic evidence which cannot be so easily accommodated by the variable analysis. Whereas, in the IA-cases, the domain of the modal appears to constrain that of the conditional, the next set of cases suggest the inverse effect. As before, I begin with the apparent inference pattern exhibited in the data, then turn to possible semantic explanations.

3.1 *Impossibility Below: the evidence*

Consider the following inferential discourses:

- (7) a. Amira didn't play in the match yesterday, but she could have.
 b. If she had played, she would have won.
 c. So she couldn't have played and lost.
- (8) a. The season finale aired Saturday when Katie was busy, but it could have aired Tuesday.
 b. If the finale had aired Tuesday, Katie would have watched.
 c. Thus the finale couldn't have aired Tuesday without Katie watching.

Focusing on (7) for illustration, it seems to be appropriate to infer from (b) to (c), that is, from a sentence of the form $play > win$ to one of the form $\neg\Diamond(play \wedge \neg win)$. And again, this impression is reinforced by the felicitous use of *so* (or *thus*, *therefore*, etc.) at the beginning of (c).

These observations suggest the existence of a second inference pattern, the mirror image of *Impossibility Above*:¹⁶

$$\text{Impossibility Below (IB)} \quad \frac{\phi > \psi}{\neg\Diamond(\phi \wedge \neg\psi)}$$

A range of further data supports the inferential force of IB. Consider the following dialogues involving *denial*:

- (9) a. A: If Talia had played, she would have won.
 b. B: No, she might have played and lost.
- (10) a. A: Talia might have played and lost.
 b. B: No, if she had played, she would have won.

The felicitous use of the B-sentences to deny the A-sentences in each case suggest that sentences of the form $\phi > \psi$ and $\Diamond(\phi \wedge \neg\psi)$ are considered incompatible in context. And this is explained directly by IB, since by IB, $\phi > \psi$ implies $\neg\Diamond(\phi \wedge \neg\psi)$.

¹⁶IB came to me via John Hawthorne's 2008 metaphysics seminar at Rutgers University; he presented the inference in its inverted form, reasoning from $\Diamond(\phi \wedge \neg\psi)$ to $\neg(\phi > \psi)$.

In a similar manner, IB can be used to explain the following *infelicities* as the assertion of incompatible propositions. In evaluating these cases, I remind readers that it is crucial to maintain a “flat” or “neutral” tone, since there is a natural tendency to put focus on the modal, which tends to shift the interpretation in a way that eliminates the proffered infelicity. I return to this effect below, and in Section 5.

- (11) a. Yesterday Ivan decided to drive to work.
- b. If he had stopped for coffee, he would have arrived late.
- c. * And he might have stopped for coffee and arrived on time.
- (12) a. Talia missed the ping-pong match last Tuesday.
- b. But she might have played and won.
- c. * What’s more, if she had played, she would have lost.

The same explanation seems to apply here: if IB is a genuine inference pattern, and $\phi > \psi$ implies $\neg\Diamond(\phi \wedge \neg\psi)$ then it should be infelicitous to assert $\phi > \psi$ and $\Diamond(\phi \wedge \neg\psi)$ in conjunction, just as we see in the examples above.

Vivid variants of these cases involve infelicitous dialogues, where the second interlocutor explicitly signals agreement with the first through the use of the affirmative *yes*.

- (13) a. A: If Talia had played, she would have won.
- b. * B: Yes, and she might have played and lost.
- (14) a. A: Talia might have played and lost.
- b. * B: Yes, and if she played, she would have won.

Another type of example relies on the inversion of IB, derived by reasoning from the negation of its conclusion to the negation of its premise, from $\Diamond(\phi \wedge \neg\psi)$ to $\neg(\phi > \psi)$. This inference is exhibited directly in the following script:

- (15) a. John wanted to take Organic Chemistry. But in all fairness, it’s a very difficult class.
- b. He could have taken it and failed.
- c. So it isn’t true that if he had taken it, he would have passed.

A final case is aimed at those who, like me, accept the duality of *would*-conditionals and *might*-conditionals, represented as $\phi \Diamond\rightarrow \psi$ (D. Lewis 1973, p. 21; Bennett 2003, §73). That is:

- (16) $\phi > \psi \Leftrightarrow \neg(\phi \Diamond\rightarrow \neg\psi)$

To add to this, the following inference seems generally impeccable:¹⁷

$$\text{Might Inference} \quad \frac{\Diamond(\phi \wedge \psi)}{\phi \Diamond \rightarrow \psi}$$

- (17) a. Isabel worked at the library all day yesterday.
 b. But she might have gone into town and seen her friends.
 c. So if she had gone into town, she might have seen her friends.

But if one assumes duality and *Might Inference*, then IB follows directly. For one can infer from $\phi > \psi$ to $\neg(\phi \Diamond \rightarrow \neg\psi)$, by duality; and one can infer from $\neg(\phi \Diamond \rightarrow \neg\psi)$ to $\neg\Diamond(\phi \wedge \neg\psi)$, by *Might Inference*. So this is additional evidence for the inferential force of IB.

At this point in the argument, readers will naturally be drawn to consider whether there are counterexamples to IB. These tend to take the form of felicitous conjunctions like the following:

- (18) a. If Kiara had played, she would have won.
 b. But of course, she *might* have played and lost.

I'll discuss this and similar examples in greater detail in Section 5. To anticipate, cases like (18) seem at first blush to contradict IB. But they notably include a cluster of distinctive linguistic markers, including focal stress (on *might*), discourse intensifiers (*of course*, *in fact*), and discourse contrast markers (*but*, *however*). I'll argue that, collectively, these devices have the effect of shifting the modal context between the first and second sentences. As a consequence, discourses like (18) do not represent counterexamples to the classical validity of IB, which assumes the fixity of context. Importantly, the infelicitous conjunctions presented above in support of IB, (11) and (12), did not invoke this cluster of markers.

What's more, when these markers are stripped from (18), the result is again infelicitous, as in (19). Such observations lend credence to the hypothesis that the markers in question work as context-shifters.

- (19) a. If Kiara had played, she would have won.
 b. * And she might have played and lost.

I mention these points here to forestall the impression that there are *obvious* counterexamples to IB. Support for IB comes from sentences which do not change context. So assessment of IB must carefully control for linguistic triggers of context change such as discourse markers and intonation. To navigate the nuances here, I will treat separately the types of case which *support* IB and the types

¹⁷A. Gillies (2007, pp. 348–49) anticipates the truth-conditional equivalence of the premise and conclusion of *Might Inference* for counterfactual conditionals; A. Gillies (2010, 4:16) observes a bi-directional version of the rule for indicatives, under the name “IF/MIGHT”.

of case which *challenge* IB. In this section I've presented data of the first type, and I will consider semantic explanations of this data in what follows. I'll take up data of the second type in Section 5.

Stepping back from the evidence, my experience is that speakers vary in the degree to which they feel the strength of the judgments reported here. Some, like myself, find the judgments in support of IB just as strong as those which support IA; others find the IB judgments of inference and infelicity in some way weaker, though nevertheless present. One potential source of this hesitation is the contextual confound just noted.

On behalf of the unified strict analysis, an additional suggestion may further explain why some readers experience the judgements supporting IB as weaker than those for IA. A plausible hypothesis is that subjunctive possibility modals have (at least) two standard, or default readings. On one reading, the domain of the modal is the same as that as the counterfactual; this explains the judgments I report here. On another reading, it naturally assumes a wider domain than that of the counterfactual; this would explain the instinct to adjudicate in the opposite direction. Consider the infelicitous cases like (11) and (12). The unified-domain reading predicts both to be contradictory, but the wider-domain reading predicts both to be consistent. My conjecture is that readers who experience the weak IB judgments are divided between these two readings. The same facts would not weaken the IA-judgments, thus explaining the alleged asymmetry between the evidence for IA and for IB. (Crucially, this ambiguity hypothesis is perfectly consistent with the unified strict analysis, since the theory requires only that *one* reading of the modal be unified with the conditional.)

In any case, even the weaker judgments will do for my purposes. What is important to the present argument is not that the inferences intuitively feel like *validities*, or that the infelicities feel like *contradictions*— validity and contradiction are a matter of theory. Instead, what is important is that there are indeed judgments of some kind which conform with IB, even if they are less vivid. As I'll ultimately argue, on semantic grounds alone the variable semantics can only explain the judgments that do not conform with IB, but none of the judgments that support IB. Meanwhile, I'll argue that the unified strict analysis can explain both.

3.2 The logic of the unified strict analysis

The semantic significance of IB is brought out by the logical consequences of conjoining IA and IB, when it is assumed that both are validities. In short, if both are valid, then the unified strict analysis must be true, and the variable analysis false.

To begin, suppose that both IA and IB are logically valid:

$$(20) \text{ IA: } \neg\Diamond(\phi \wedge \neg\psi) \models \phi > \psi$$

$$(21) \text{ IB: } \phi > \psi \models \neg\Diamond(\phi \wedge \neg\psi)$$

From which it follows that the two formulae are equivalent:

$$(22) \quad \phi > \psi \Leftrightarrow \neg\Diamond(\phi \wedge \neg\psi)$$

But by a few short steps, $\neg\Diamond(\phi \wedge \neg\psi)$ is equivalent to $\Box(\phi \supset \psi)$:

$$\begin{aligned} & \neg\Diamond(\phi \wedge \neg\psi) \\ \Leftrightarrow & \neg\Diamond\neg(\neg\phi \vee \psi) && [\text{DeMorgan's Law}] \\ \Leftrightarrow & \neg\Diamond\neg(\phi \supset \psi) && [\text{Definition of } \supset] \\ \Leftrightarrow & \Box(\phi \supset \psi) && [\text{Definition of } \Box] \end{aligned}$$

It follows that:

$$(23) \quad \phi > \psi \Leftrightarrow \Box(\phi \supset \psi)$$

In other words, the counterfactual is equivalent to the strict conditional. And since the \Box in question is defined as the dual of \Diamond , it follows that this strict conditional is also coordinated with the \Diamond in the manner of the unified strict analysis.

The preceding argument shows that the unified strict analysis follows from the validity of IA and IB. Intuitively, the argument works because IA and IB each restricts the quantificational domain of the conditional in opposite directions. On one hand, IA implies that the range of worlds quantified by the conditional are limited to those quantified by the modal. On the other, IB holds that the worlds quantified by the modal are limited to those quantified by the conditional. Putting these together, it follows that in any context, the domains quantified by the conditional and modal are the same, just as the unified strict analysis requires.

The significance of this result is that any approach that makes sense of IA and IB by making them both semantic entailments will result in a theory which is truth-conditionally equivalent to the unified strict analysis. As a consequence, semantic explanations of both IA and IB are simply not available to the variabilist. Assuming the variabilist wishes to hold on to the validity of IA, they must propose an account of IB which does not render it logically valid.

3.3 *Impossibility Below: semantic explanations*

As the preceding argument shows, the inferential force of IB is directly explained by the semantics of the unified strict analysis. In explicitly semantical terms: the premise of IB, $\phi > \psi$, is only true at a world i if every accessible ϕ -world is a ψ -world. Hence no accessible ϕ -world is a $\neg\psi$ -world. And so, by the unified semantics for modals, the conclusion, $\neg\Diamond(\phi \wedge \neg\psi)$ must also be true at i . With little elaboration, this explains the full range of data, from felicitous inferences, to denials, infelicitous conjunctions, and interaction with might-conditionals.

Meanwhile, the semantic options for the variable analysis are the mirror image of those generated by IA: (i) the k -semantics on its own does not validate IB; (ii) the *min*-semantics *does* validate

IB; (iii) the *max*-semantics does not.¹⁸ The *min*-semantics validates IB, because it makes the quantified domain of the modal as narrow as that of the conditional. Yet as we saw in Section 2.2, the *min*-semantics is not a viable option for the variable analysis in the present context. Still, the same semantic effects can be achieved for the *k*-semantics and *max*-semantics by adopting the corresponding constraint:

Proximity

For any $i, c : k(i) \subseteq \min(\geq_i)$.

But Proximity, together with the *max*-semantics once again renders the variable analysis a notational variant of the unified strict analysis. The *max*-semantics makes the set of accessible worlds equal to the maximal set of worlds; Proximity makes the accessible worlds a subset of the minimal set of worlds. These constraints can only be satisfied together if the ordering on worlds is such that all worlds are either in the closest sphere, or not ordered at all. The result is truth-conditionally equivalent to unified strict analysis.¹⁹ For parallel reasons, the *k*-semantics, when combined with both Accessibility and Proximity, results in the same flat ordering. Either way, the conditional would lose the semantically non-monotonic character distinctive of the variable analysis.

These results are of course inevitable given the argument in the preceding subsection. Any set of constraints that guarantees the validity of both IA (i.e. Accessibility) and IB (i.e. Proximity) yields a theory equivalent to the unified strict analysis. Since any purely semantic explanation of IA and IB will, by design, validate them, such semantic explanations are foreclosed for the variable analysis. Assuming that the variable theorist is committed to validating IA, they must reject Proximity, along with the *min*-semantics, and adopt either the *k*-semantics with Accessibility or the *max*-semantics alone. But then they must account for the evidence supporting IB via some other mechanism. In the next section, I'll consider alternative pragmatic and discourse-based explanations of IB on behalf of the variable analysis.

4 Counter-explanations?

In this section I ask whether the variable analysis might explain the inferential force of IB with the help of pragmatic or discourse-based resources. I'll consider two apparently plausible accounts culled from recent work on conditionals in discourse, but argue that, in each case, the proposed explanations are not supported by the linguistic evidence. The prospects for a pragmatic account of IB, I conclude, are dim.

¹⁸Thanks to an anonymous reviewer for highlighting this fact pattern.

¹⁹According to the *max*-semantics, $k(i) = \max(\geq_i)$. But according to Proximity, $k(i) \subseteq \min(\geq_i)$, hence $\max(\geq_i) \subseteq \min(\geq_i)$. By definition, $\min(\geq_i) \subseteq \max(\geq_i)$. So, in the ordering that results, $\min(\geq_i) = \max(\geq_i)$.

I wish to begin by recalling the most straightforward positive evidence for IB. These included felicitous inferences as well as infelicitous monologues. (There was other evidence, which I don't focus on here, including cases of denial, and interactions with might-conditionals.) I reproduce the inference and infelicity cases in compact form below.

- (24) a. If Amira had played, she would have won.
b. So she couldn't have played and lost.
- (25) a. If Amira had played, she would have won.
b. * And she might have played and lost.
- (26) a. Amira might have played and lost.
b. * And if she had played, she would have won.

How might the variabilist explain this sort of data? Inspiration comes from recent work by Moss (2012), W. B. Starr (2014a), K. S. Lewis (2016, 2017), and Stojnić (2017, 2016) on the role of conditionals in discourse, particularly Sobel and reverse Sobel sequences. Adopting the broad strategy of these authors, one might explain the inferential case (24) by claiming that sentence (a) induces some kind of systematic shift of context (either linguistic or epistemic) so that (b) is either entailed or naturally inferred from the new context. The inferential force of IB is explained as pragmatic, discourse-bound, or epistemic, rather than logical. I'll consider two ways of implementing this idea, ultimately arguing that neither will work.

In defending a variable analysis treatment of reverse Sobel sequences, K. S. Lewis (2017, pp. 18–21) proposes that asserting a conditional tends to raise the set of nearest antecedent-worlds to salience. As a consequence, context shifts in such a way that these antecedent-worlds are now included among the closest worlds for subsequent conditionals. Adapted to the present case, we may imagine that modals too have the force of raising their prejacent-worlds to salience, such that subsequent conditionals treat those as among the closest worlds. This could explain the judgments about (26): the modal in (a) raises *play* \wedge *lose*-worlds to salience; these are treated as closest for the conditional in (b), which must thereby be false, since some closest *play*-worlds are *lose*-worlds.

But the same strategy cannot be extended to the other examples. Focusing on (24) for illustration, suppose that the conditional in (a) has the effect of raising the closest *play*-worlds to salience, all of which will also be *win*-worlds. Then, in the spirit of Lewis' suggestion, we can assume these must be included in the modal domain of (b). But this is not enough to ensure that (b) is entailed by the new context: the presence of closest *play* \wedge *win*-worlds among the accessible worlds doesn't settle the question of whether there are accessible *play* \wedge *lose*-worlds. Thus the general strategy fails to capture the full range of data that support IB. The stumbling block for this account, and others like it, is that it predicts strong order effects. For it only works when the modal precedes the conditional. By contrast, as (25) and (26) demonstrate, the immediate evidence for IB is order

invariant. This is part of why it is natural to explain the full range of evidence in terms of semantic entailment, a relation which is insensitive to the order of discourse presentation.

A second strategy, which circumvents this concern, is inspired by Starr’s (2014a, pp. 1051–55) treatment of reverse Sobel sequences and Stojnić’s (2016, §1.4, §2, 2017, §4) analyses of discourses involving indicative conditionals. These analyses focus on the role of **modal subordination**, the phenomenon in which one conditional or modal functions to provide comment on a possibility introduced into the discourse by an antecedent modal expression (Roberts 1989, pp. 683–92). Thus, in (27) below, (a) introduces a possibility, that a wolf might walk in, and (b) is modally subordinated to (a) since it says *of that possibility*, that the wolf would eat you. It does not say, for example, that a wolf *will* eat you, come what may.

- (27) a. A wolf might walk in.
b. It would eat you.

For concreteness, I’ll focus on adapting Stojnić’s detailed proposal to the data at hand. Following Stone (1997, 1999), Stojnić (2016, pp. 177–82) holds that modals and conditionals are covertly anaphoric, their quantificational domain restricted by an anaphorically selected proposition (set of worlds), just as in (27b) above. Stojnić makes two additions to this picture: first, the idea that modals and conditionals also promote the possibilities they describe to contextual prominence (2016, pp. 189–93); and second, that in the right discourse context, subsequent modals are anaphorically linked to these prominent possibilities by default (2016, pp. 196–98). The result is that, in the right discourse context, modals and conditionals tend to restrict the quantificational domain of subsequent modal expressions. Discourse context, in turn, is understood in terms of the available **discourse coherence relations**, posited by linguists to help explain the informational relations between clauses in discourse (see e.g. Hobbs 1985; Kehler 2002; Asher and Lascarides 2003). Stojnić hypothesizes that some relations, like *Elaboration*, exhibited in (27), impose modal subordination, while others, like *Contrast*, do not.²⁰

If this strategy could be sustained here, it would explain several key pieces of data. Here’s how the story might go: in (24) the conditional in (a) shifts to prominence the set of closest *play*-worlds which are also *win*-worlds. If (b) is modally subordinated to this set, then of necessity (b) will be true, since no world in the set is a *play* \wedge *lose*-world. A parallel explanation applies to (25). As for (26): the modal in (a) raises to prominence the set of accessible *play* \wedge *lose*-worlds. The domain of the selection function in (b) is then restricted to this set, with the result that the conditional is necessarily false.

While this account offers an in-principle explanation of the available data, additional evidence tells against it. The problem is that the proposed explanation goes through only if the discourses

²⁰To be clear, in Stojnić’s analysis, all modals are anaphoric; what I am calling “modal subordination” is, in Stojnić’s terms, modal anaphora to the most prominent possibility in the immediately preceding discourse.

in question (quite different from those originally discussed by Stojnić) are such that they in fact do involve modal subordination. Stojnić focuses primarily on the relation of *Elaboration* as a trigger of modal subordination, but there are many other discourse relations which do not have this effect.

The chief obstacle for the modal subordination account arises for felicitous *inferences* like (24). Here a relation of inference or conclusion is expressed by the discourse particle *so*. Yet the use of *so* actually seems to block modal subordination in many contexts.²¹ Evidence comes from the following discourse, where the (b) and (c) sentences are close semantic analogues of those from (24).²² I change “played” to “played using a new bat” here so as to allow for the possibility of playing (without a new bat) and losing.

- (28) a. Amira didn’t play in the match yesterday, but she could have.
- b. If she had played using a new bat, she would have won.
- c. * So she couldn’t have lost.

Suppose (c) were subordinated to (b). Then the *play bat* \wedge *win*-worlds introduced by (b) would constrain the modal in (c). Since these worlds include no *lose*-worlds, (c) should be an obvious entailment, and the use of *so* should be felicitous as a result. Yet it is not, so (28) is not an instance of modal subordination. This shows that, at the very least, *so* is not a marker of subordination, and further suggests that it may even block subordination. By extension, there is reason to doubt that (24) itself is an example of modal subordination.

The same lesson is illustrated with other combinations of modals and conditionals. In the following discourses, modal subordination would predict that the conclusion is contextually entailed by the assertion of the premise, hence felicitous. Yet the use of *so* is once again infelicitous.

- (29) a. Amira might have played and won.
- b. * So if she had played, she would have won.
- (30) a. If Amira had played, she might have won.
- b. * So if she had played, she would have won.

As for the infelicitous conjunctions (25) and (26), the modal subordination account again requires that these cases be governed by *Elaboration* or some other modally subordinating relation. But the evidence here is mixed at best. In a comprehensive discussion of the discourse particle *and*, Txurruka (2003) argues that one function of *and* is to signal the class of *coordinating* discourse relations, which tend to introduce a new discourse topic, rather than offer further information on

²¹The particle *so* blocks modal subordination in many contexts, but its overall behavior is complex. The following discourse is felicitous: “A wolf might walk in. It would be hungry. So it would eat you.” As it this: “If a wolf walked in, it would be hungry. So it would eat you.” But this variant is not: “A hungry wolf might walk in. So it would eat you.” It seems that the modally anaphoric behavior of *so* may be modulated both by discourse context and lexical encoding. In the text I focus on cases that are relevantly similar to the target examples.

²²Thanks to Una Stojnić (p.c.) for this example.

an existing one. While there are a variety of coordinating relations, including relations of narration and causation, *Elaboration* is specifically excluded. Whether some other relation, compatible with modal subordination is at work in these infelicitous conjunctions is a separate question. But absent specific evidence to that effect, the proposed explanation falls short.

In sum, the modal subordination strategy fails for the inferential data, and is inconclusive for the infelicity data. Meanwhile, the unified strict analysis provides a clear explanation of both, with little elaboration.

At this point, I've considered two possible pragmatic explanations of the phenomena at hand, both based on flexible, contemporary models of modal discourse. But neither successfully explains the available evidence. Of course there may be some further pragmatic account compatible with the variabilist theory that I have not considered.²³ But I know of no such proposal in the current literature. At the very least, the explanatory burden has been moved to the variabilist's court.

While there is no proof that all variabilist pragmatic approaches fail, such accounts face serious in-principle challenges. The evidence for IB covers a heterogeneous set of linguistic data, including inferences, denials, and infelicities. A semantic account of IB, like the unified strict analysis, provides the same kind of logic-based explanation across discursive contexts. But it is the nature of pragmatic and discourse-based phenomena to act differently in different discursive contexts. As a result, pragmatic accounts have trouble evoking stable results in these varied settings. We saw this with Lewis-based account, which could only predict IB-like effects in one discourse order, but not the other. And the Stojnić-based account could only predict IB-like effects with some discourse relations, but not others. Any pragmatic account must overcome this basic challenge.

The difficulties for the variabilist are not over. In the next section, we'll encounter a range of additional discourse phenomena which the unified strict analysis smoothly accounts for, but the variable analysis does not. These must also be counted as part of the total evidential burden facing each theory.²⁴

²³One might, for example, reconsider the data at hand in light of Schulz's arbitrary selection theory of counterfactuals, an analysis more aligned with the variable than the strict analysis (Schulz 2014, §1). (Thanks to an anonymous reviewer for this idea.) Schulz suggests that might-counterfactuals and their dual would-counterfactuals are semantically compatible, but epistemically exclusive (2014, §3.3)– and as a consequence, one can't felicitously assert both at the same time. The same strategy might be attempted for the IB-data discussed here, but this approach faces immediate problems. Schulz's analysis crucially turns on the fact that might- and would-counterfactuals have the same quantified domain, the set of closest antecedent-worlds. But it is just this assumption that is challenged by the data I've presented here.

²⁴The dialectic here plays out somewhat differently for indicative conditionals. I believe the data favoring IB is just as robust for indicatives. But in many cases, the variabilist has a further pragmatic response at their disposal. In short: if indicative conditionals are variably strict, asserting $\phi > \psi$ will, on plausible assumptions, have the pragmatic effect of eliminating all $\phi \wedge \neg\psi$ -worlds from the context set. As a result, on an analysis in which indicative (or epistemic) modals quantify over the context set, $\neg\Diamond\phi \wedge \neg\psi$ will be true. These facts can be used to explain many forms of IB-supporting data for indicatives; whether it can be extended to *all* IB-supporting data requires further investigation.

5 Counter-examples?

I turn finally to a family of examples which seem to challenge the inferential force of IB, thus presenting apparent counterexamples to the unified strict analysis itself. I'll argue that these examples not only fail as counterexamples, they turn out to be part of a larger body of data which collectively supports the unified strict analysis.

The examples in question have the following form, where the words written in *italics* are pronounced with emphasis:

- (31) a. If Kiara had played, she would have won.
b. But of course, she *might* have played and lost.
- (32) a. Kiara might have played and lost.
b. But in fact, if she *had* played, she *would* have won.

In each case, the (b) sentence is clearly felicitous. And this, on its face, is a challenge for the unified strict analysis, because according to the unified strict analysis the sentence with the logical form of (a) should be inconsistent with a sentence with the form of (b). Meanwhile, according to the variable semantics, the propositions expressed by the (a) sentences and (b) sentences are perfectly compatible, so no infelicity is predicted.²⁵

What is notable about these and similar cases is the presence of three distinctive linguistic signals: (i) focal stress; (ii) discourse intensifiers; and (iii) discourse contrast markers. **Focal stress** is the phenomenon in which a speaker puts special emphasis on a particular term as a means of modulating its meaning. **Discourse intensifiers** (as I call them) are clausal prefixes such as *of course*, *really*, or *in fact*. **Discourse contrast markers** are clausal prefixes such as *but*, *however*, and *by contrast*. (It may be that *of course* belongs to this latter class as well.) Collectively, I shall call these **shift markers**, for reasons detailed below.

I'll return to the interpretation of shift markers shortly. What is striking is that, when they are removed, and the discourse particle *and* (or *what's more*) is introduced in their stead, the sense of contradiction and infelicity vividly reemerges:

- (33) a. If Kiara had played, she would have won.
b. * And she might have played and lost.
- (34) a. Kiara might have played and lost.
b. * What's more, if she had played, she would have won.

²⁵I assume that the variable semantics follows either the *k*-semantics (+ Accessibility) or the *max*-semantics; see Section 3.3.

I'll call the alleged counterexamples, (31) and (32), **marked** cases, and the revised conjunctive examples, (33) and (34), **unmarked** cases. On semantic grounds alone (together with the assumption that contradictions are infelicitous), the variable analysis can explain the felicity of the marked cases, but not the infelicity of the unmarked cases. The reverse is true for the unified strict analysis. So each party must offer something like a pragmatic or discourse-based explanation of the opposing case.²⁶ But this leaves the variabilist in a doubly bad position. First, as I argued in the last section, there seems to be no readily available variabilist explanation of the unmarked cases. But even if there were, the proffered explanation (whatever it may be) would seem to be of the wrong kind. It is the marked cases which we would normally expect to involve a shift in context and the unmarked cases where context should remain inert. I turn now to outline a discourse-based explanation of the marked cases on behalf of the unified strict analysis.

In the dynamic semantics outlined by von Fintel (2001) and A. Gillies (2007), context includes a **modal domain**— more precisely, a function from worlds to domains— which is the quantificational domain of conditionals.²⁷ By default, the modal domain remains fixed as discourse progresses, so that each successive sentence in a discourse expresses its content relative to the same contextual parameter.²⁸ Only specific linguistic signals can trigger an expansion or contraction of the modal domain.

For von Fintel (2001) and A. Gillies (2007) a primary case is domain expansion triggered by a **possibility presupposition** associated with the antecedent of a conditional.²⁹ That is, an assertion of $\phi > \psi$ presupposes the compatibility of ϕ with the modal domain. If the presupposition of the conditional is already met in a context, the modal domain is unchanged by the antecedent. But if it is not, then the modal domain expands to accommodate the presupposition.

In the account of von Fintel (2001, pp. 7, 19–22) and A. Gillies (2007, pp. 334–337), expansions of the modal domain are defined relative to a background ordering on worlds. The idea is that accommodation-driven expansions are *conservative*, not arbitrary. When the possibility presupposition is accommodated, the modal domain expands as little as possible, given the ordering, so as to fit the presupposition. The original variabilist idea of an ordering on worlds as relevant to counterfactual domain is here recast as a feature of discourse structure, rather than a parameter in the semantics of the conditional.

In addition, A. Gillies (2007, §6) offers an analysis of possibility modals, which likewise have the power to trigger domain expansion. Gillies' possibility modals quantify over the same modal

²⁶It is a remarkable fact that nearly the entire corpus of linguistic data surrounding the variable/strict debate (including, for example, Sobel and reverse Sobel sequences) has failed to control for the presence of discourse markers or test for their effects. It would be instructive to revisit these debates with this dimension of variation in mind.

²⁷Cf. the "modal horizon" of von Fintel (2001, p. 7) and the "hyperdomains" of A. Gillies (2007, p. 334).

²⁸To be clear, when I say here and below that the *modal domain* persists through discourse, I mean that, *relative to* a world of evaluation, the modal domain persists. What actually is carried forward in discourse is an accessibility function, or its equivalent.

²⁹Cf. the "compatibility presupposition" of von Fintel (2001, p. 16), and the "entertainability presupposition" of A. Gillies (2007, p. 330).

domain as conditionals, so that what I have been calling the set of *accessible* worlds (for possibility modals) is simply identified with the modal domain. Although Gillies' focus is on possibility modals in the consequent of conditionals, the same analysis is straightforwardly extended to bare possibility modals, and the overall framework is one of a unified strict analysis.

Following a proposal from D. Lewis (1979, ex. 6), Gillies characterizes the domain shifting effect of modals on context in terms of **accommodation**. When a sentence $\Diamond\phi$ is asserted, if there are no ϕ -worlds in the modal domain, then, under suitable circumstances, the context shifts to accommodate the assertion: the set of accessible worlds expands, so that, relative to the new context, $\Diamond\phi$ expresses a true proposition which is compatible with preceding discourse. Gillies further analyzes such accommodation in terms of a possibility presupposition associated with the prejacent of possibility modals themselves. I'll remain uncommitted here about the specific semantic mechanisms at work in the accommodation of possibility modals, but, like D. Lewis (1979, pp. 346–47), I will allow accommodation of asserted content, as well as presupposed content.

In sum, I propose to adopt this general framework– the shifting modal domain, the background ordering of worlds, and the accommodation-driven effects of modals and conditionals– if not all the specifics of each formulation. But I take a more variegated view of the role of accommodation in shifting the modal domain than von Fintel or Gillies. While they recognize standardized domain shifts driven by fixed semantic mechanisms, I propose that the modal domain is liable to shift in more than one way, and in response to a range of distinct semantic and discourse-level constraints.

For the particular case at hand, I wish to propose two modifications to the von Fintel-Gillies picture. First, I hold that, at least in the kinds of cases at hand, accommodation does not occur automatically; it must be triggered by an appropriate linguistic signal. In particular, the use of the contrastive discourse marker, in conjunction with the other shift markers, is sufficient to trigger an accommodation-driven change to the modal domain.³⁰ Meanwhile, the use of *and*, in the absence of shift markers, tends to signal that the modal domain stays fixed.³¹ Second, whereas (D.) D. Lewis (1979, pp. 354–55), as well as von Fintel (2001, §8) and A. Gillies (2007, §9), claim that only domain *expansions* are easy to achieve in discourse, I join (K.) Lewis (2017, pp. 7–9) in concluding that both expansions and contractions of the modal domain occur smoothly in normal conversation as a result of accommodation, albeit only when accompanied by the appropriate linguistic signals.

³⁰A potential simplification of this analysis might hold that the presence of the *Contrast* discourse relation alone is sufficient to trigger modal accommodation. Yet it seems to me that infelicitous readings of the marked cases are (more) available when only the contrast marker is used (which is sufficient to incur *Contrast*), but not available when all three shift markers are employed. Thus the discourse relation of *Contrast* is necessary for triggering a domain shift, but that is not the whole story about the discourse mechanisms at play here.

³¹Cf. A. Gillies (2007, p. 354), conjectures that the inferential marker *so* has a similar shift-blocking effect. In addition, von Fintel and A. S. Gillies (2018, pp. 11–13), suggest that context shift is blocked between conjoined clauses prefixed by *although*, or embedded under *suppose* or in the antecedent of a conditional.

Consider again the roster of cases I began with. In the unmarked cases, the absence of any discourse marker except *and* signals the default persistence of the modal domain. Thus, in each case, the two sentences are evaluated at the same modal domain. By the unified strict analysis, the propositions expressed are logically incompatible, hence the second assertion in each case is infelicitous.

By contrast, in the marked cases (31) and (32), the cluster of shift markers in the second sentence signal an adjustment to the modal domain, which thereby expands or contracts to accommodate the main clause. In (31), accommodation of the modal results in an *expansion* of the modal domain. The conditional in (a) asserts that all *play*-worlds in the domain are *win*-worlds; then (b) triggers an expansion of the modal domain to include some *play* \wedge *lose*-worlds, so that the modal is true. In (32), accommodation of the conditional results in a *contraction* of the modal domain. The modal in (a) asserts that there are some *play* \wedge *lose*-worlds in the modal domain; (b) triggers a contraction of the domain, restricting to *play*-worlds which are all *win*-worlds, thereby making the conditional true.

I maintain that this account of the linguistic evidence data is not *ad hoc*, but well motivated by a range of parallel phenomena. Here it is helpful to focus on the role of the contrastive discourse marker *but*, since it seems to be the one shift marker which is essential to trigger a shift. (Nevertheless, felicitous readings of the marked cases are clearer with the addition of both intensifiers and focal stress.)

In the study of discourse structure, contrast markers are thought to signal a contrast, or violated expectation, between a preceding segment of discourse and the clause to follow. (Hobbs 1985, p. 32; Knott 1996) Yet the notion of contrast at work here is quite open-ended; the relevant contrast may be an inconsistency or difference in embedded clauses, predicates, presuppositions, implicatures, or merely associated expectations (Hobbs 1985, pp. 21–23; Asher and Lascarides 2003, §1.1, §4.8.4). Plausibly, a contrastive particle could be used to signal a shift to a new context which differs from the previous context in a manner relevant for the discourse. In the alleged counterexamples, I contend, *but* is used to signal a shift from a wide modal domain that satisfies the possibility modal to a narrow modal domain that satisfies the conditional, or *visa versa*. On the other hand, *and* is thought to be a marker of coordination; it signals that the adjoined segments of discourse are in some sense parallel or equally situated in the information structure of the discourse (Hobbs 1985; Txurruka 2003). And it is further plausible that *and* has the opposite effect of *but*, actually requiring that the contextual domain to remain in its default fixed state.

As corroboration of these ideas, we find that the very same shift markers can be used to trigger a contextual shift in the domain of quantification for nominal quantifiers, where the semantics of the relevant expressions are relatively well established. For example, the standard analysis of English quantifiers predicts that “there is (only) an α in β ” and “there is nothing in β ” are, in the same context, logically incompatible, hence infelicitous to assert in conjunction. Yet the combined

use of contrast markers, intensifiers, and focal stress render these discourses acceptable:

- (35) a. My friend Thomas never goes shopping.
b. There is only a ketchup bottle and a can of beer in the fridge.
c. But really, there is *nothing* in the fridge.
- (36) a. My friend Thomas never goes shopping.
b. There is nothing in the fridge.
c. Though of course, there *is* a ketchup bottle and a can of beer in the fridge.

The context shifting mechanisms at work here appear to be precisely the nominal counterpart of that proposed for the modal case. Thus the use of shift markers in (35) appears to have the effect of admitting a narrowing of the domain of quantification, while those in (36) seem to have the effect of admitting a widening of the domain. Notably, the same pattern of infelicity that characterized the modal cases emerges when the shift markers are removed:

- (37) a. There is only a ketchup bottle and a can of beer in the fridge.
b. * And there is nothing in the fridge.
- (38) a. There is nothing in the fridge.
b. * And there is a ketchup bottle and a can of beer in the fridge.

Such observations lend credence to the account of signal-driven accommodation I've sketched on behalf of the unified strict analysis. They suggest that shift markers are a general way to signal accommodation of changes to the relevant quantificational domain, and discourse conjunctions a general way of blocking such domain change.

Still, the account of modal accommodation outlined here isn't meant to be comprehensive. The use of shift markers to signal accommodation appears to be a partially pragmatic and defeasible process, potentially superseded by other linguistically encoded cues of accommodation. In the systems of von Stechow (2001, pp. 15–20) and A. Gillies (2007, pp. 333–34), for example, the possibility presuppositions associated with conditional antecedents are automatically accommodated by dynamic semantic mechanisms.

Such semantically-driven accommodation of the antecedent possibility appears to be an especially explicit and powerful way to push a possibility into the modal domain. Indeed, it may force an expansion of the modal domain even in the face of contrary discourse signals. This can be seen in the relative felicity of Sobel sequences conjoined by *and*.³²

- (39) a. Mary didn't go to work today, but she could have.

³²See Moss (2012, §1-2) for a review of Sobel and reverse Sobel sequences.

- b. If Mary had driven to work she would have arrived on time.
- c. And if she had driven to work and got a flat tire, she would have arrived late.

Here the modal domain in (c) is expanded from that of (b), thereby making both (b) and (c) true and felicitous in their respective contexts. It is allowed to shift despite the presence of *and*, which normally signals the fixity of modal context, because the driver of accommodation is the conditional antecedent in (c). Still, it is notable that a contrastive version with *but* is clearly preferable, and telling that the motivating cases of antecedent strengthening in D. Lewis (1973, p. 10) all make use of the contrastive particle.

Once the antecedent of a conditional is used to embed a possibility into the modal domain, it is difficult to eliminate it merely via signal-driven accommodation. This seems to be illustrated by the case of reverse Sobel sequences that employ the contrastive particle *but* as a shift marker.³³

- (40)
- a. Mary didn't go to work today, but she could have.
 - b. If she had driven to work and got a flat tire, she would have arrived late.
 - c. * But if Mary had driven to work, she would have arrived on time.

In this case, (c) is infelicitous, suggesting a failure of accommodation, despite the presence of a shift marker. It seems that the presuppositions of conditional antecedents make a kind of indelible mark on the discourse record that can only be overcome through equally explicit means. This dominance of semantic presupposition over signal-driven accommodation is to be expected. It's one thing to assume an open ended change in context when the speaker invites you to do so. It's another to assume a change in context in the face of an explicit commitment on behalf of the speaker to the contrary.

Still, when all three shift markers—contrast markers, intensifiers, and focal stress—are brought to bear on reverse Sobel sequences, the sense of infelicity is weakened (at least to my ears). At any rate, a coherent narrowed interpretation makes itself available, if somewhat awkwardly.

- (41)
- a. Mary didn't go to work today, but she could have.
 - b. If she had driven to work and got a flat tire, she would have arrived late.
 - c. ? But in fact, if Mary *had* driven to work, she *would* have arrived on time.

In such cases we can see a tension emerge between the explicit semantic signaling of conditional antecedents and the pressure to accommodate in the presence of shift markers.³⁴ Further work is

³³Moss (2012, pp. 574–75) and K. S. Lewis (2017, pp. 7–8) note that reverse Sobel sequences (with the contrastive particle) may be felicitous when the first conditional antecedent contains a bizarre possibility as a conjunct, and the second does not.

³⁴A. Gillies (2007, p. 354) discusses a parallel case of (attempted) contraction with the marker of violated expectation *still*, in lieu of *but*, also noting its awkwardness. For me, variants with *still* are even easier to accommodate than those with *but*. Another violated expectation marker, *nevertheless*, seems to issue in a similarly effortless contraction of the modal domain. Moss (2012, p. 576) makes parallel observations about explicit signals like the clausal prefix “Oh *come on*.” These are subtle judgements that deserve further analysis.

called for in quantifying and systematizing these sometimes opposed conversational forces.

I began this section with examples which appeared, at first glance, to be direct counterexamples to the unified strict analysis. On closer inspection, I've argued, just the opposite seems to be true: the unified strict analysis has at its disposal a simple, independently motivated account of a wide range of relevant data, while the variable analysis appears ill-suited to offer equal explanatory coverage.

6 Semantics and Modality

This essay has concerned itself with the interactions between different linguistic expressions of modality— counterfactual conditionals and counterfactual possibility modals. I've presented evidence for two inference patterns, IA and IB, that establish inferential links between modals and conditionals. As I've shown, the unified strict analysis directly explains both as semantic validities, and any theory which makes both valid is equivalent to the unified strict analysis. Although the variable analysis easily validates IA, I've argued that it faces serious difficulties in explaining away IB as a merely pragmatic or discourse-based inference. Further, alleged counterexamples to the unified strict analysis only bolster its evidential position as compared to the variable analysis. Together, these considerations constitute my case for the unified strict analysis of modals and conditionals, and by extension, for the strict analysis of conditionals, and against the variable analysis.

These conclusions offer a new perspective on the enduring debate between strict and variably strict theories of the conditional.³⁵ With plausible accounts of the the linguistic evidence on both sides, as well as unanswered questions, it can look like the dispute has reached an impasse. In this essay, I hoped to have tipped the balance in favor of the strict analysis by drawing attention to a new source of linguistic evidence, and a new set of arguments.

Beyond the particular debate at hand, the deeper lesson of this discussion is that modals and conditionals are not independent semantic devices, expressing merely loosely related modal facts. Like universal and existential nominal quantifiers, like indicative conditionals and epistemic modals, counterfactual conditionals and counterfactual modals constitute a unified suite of quantificational tools. Their specialized task is the navigation of a common counterfactual geography.

³⁵The variable analysis was originally motivated by Stalnaker (1968) and D. Lewis (1973) with a range of non-monotonic data, such as Sobel sequences. Discourse-based treatments of strict conditionals by von Fintel (2001) and A. Gillies (2007) subsequently offered plausible answers to these challenges, and replied with their own set of data— reverse Sobel sequences and their variants. And this data has, in turn, been met with counter-explanations on behalf of the variable analysis (Moss 2012; W. B. Starr 2014a; K. S. Lewis 2017), and new challenges to the strict analysis (K. S. Lewis 2017; Nichols 2017).

References

- Asher, N. and A. Lascarides (2003). *Logics of Conversation*. Cambridge University Press.
- Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press, USA.
- Daniels, C. and J. Freeman (1980). "An analysis of the subjunctive conditional." In: *Notre Dame Journal of Formal Logic* 21.4.
- Edgington, D. (2003). "Counterfactuals and the benefit of hindsight". In: *Causation and Counterfactuals*. London: Routledge, pp. 12–27.
- Fintel, K. von and A. S. Gillies (2018). "Still going strong". In: *Manuscript, MIT and Rutgers University*.
- Gillies, A. (2007). "Counterfactual scorekeeping". In: *Linguistics and Philosophy* 30, pp. 329–360.
- (2010). "Iffiness". In: *Semantics and Pragmatics* 3.0, p. 4. ISSN: 1937-8912.
- Goodman, N. (1947). "The Problem of Counterfactual Conditionals". In: *Journal of Philosophy* 44.5, pp. 113–128.
- Hobbs, J. R. (1985). *On the Coherence and Structure of Discourse*. Center for the Study of Language and Information.
- Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Center for the Study of Language and Information.
- Knott, A. (1996). "A Data-Driven Methodology for Motivating a Set of Coherence Relations". PhD thesis. University of Edinburgh.
- Kratzer, A. (1977). "What 'Must' and 'Can' Must and Can Mean". In: *Linguistics and Philosophy* 1, pp. 337–55.
- (1981). "The notional category of modality". In: *Words, worlds, and contexts: : New approaches in word semantics*. Ed. by H. J. Eikmeyer and H. Rieser. de Gruyter, Berlin, pp. 38–74.
- (1986). "Conditionals". In: *Chicago Linguistics Society* 22.2, pp. 1–15.
- (1991). "Modality". In: *Semantics: An International Handbook of Contemporary Research*, pp. 639–650.
- (2012). *Modals and conditionals: New and revised perspectives*. Vol. 36. Oxford University Press.
- Kripke, S. (1963). "Semantical considerations on modal logic". In: *Acta philosophica fennica* 16.1963, pp. 83–94.
- Kvart, I. (1986). *A theory of counterfactuals*. Hackett Pub. Co. ISBN: 0915145634.
- Leitgeb, H. (2012). "A probabilistic semantics for counterfactuals, Part A." In: *Review of Symbolic Logic* 5.1, pp. 26–84.
- Lewis, C. (1914). "The Calculus of Strict Implication". In: *Mind* 23.1, p. 240.
- Lewis, D. (1973). *Counterfactuals*. Blackwell Publishers.
- (1979). "Scorekeeping in a Language Game". In: *Journal of Philosophical Logic* 8, pp. 339–59.
- Lewis, K. S. (2016). "Elusive counterfactuals". In: *Noûs* 50.2, pp. 286–313.
- (2017). "Counterfactual Discourse in Context". In: *Noûs*.
- Loewer, B. (2007). "Counterfactuals and the second law". In: *Causality, Physics, and the Constitution of Reality: Russell's Republic Revisited*.
- Lowe, E. (1983). "A Simplification of the Logic of Conditionals". In: *Notre Dame Journal of Formal Logic* 24.3.
- (1995). "The truth about counterfactuals". In: *The Philosophical Quarterly*, pp. 41–59.
- Lycan, W. (1984). "A syntactically motivated theory of conditionals". In: *Midwest Studies in Philosophy* 9.1, pp. 437–455.
- (2001). *Real Conditionals*. Oxford University Press, USA.

- Moss, S. (2012). "On the pragmatics of counterfactuals". In: *Noûs* 46.3, pp. 561–586.
- Nichols, C. (2017). "Strict conditional accounts of counterfactuals". In: *Linguistics and Philosophy*, pp. 1–25.
- Peirce, C. (1933). *Collected Papers of Charles Sanders Peirce, Vol. III: Exact Logic*. Ed. by C. Hartshorne and P. Weiss. Cambridge, MA: Harvard University Press.
- Roberts, C. (1989). "Modal subordination and pronominal anaphora in discourse". In: *Linguistics and philosophy* 12.6, pp. 683–721.
- Sanford, D. (2003). *If P, then Q: Conditionals and the Foundations of Reasoning*. Routledge.
- Schulz, M. (2014). "Counterfactuals and arbitrariness". In: *Mind* 123.492, pp. 1021–1055.
- Stalnaker, R. (1968). "A theory of conditionals". In: *Ifs. Conditionals, Belief, Decision, Chance, and Time*. Dordrecht, D. Reidel Publishing Company, pp. 41–55.
- Stalnaker, R. and R. Thomason (1970). "A semantic analysis of conditional logic". In: *Theoria* 36.1, pp. 23–42.
- Starr, W. (2019). "Counterfactuals". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Metaphysics Research Lab, Stanford University.
- Starr, W. B. (2014a). "A uniform theory of conditionals". In: *Journal of Philosophical Logic* 43.6, pp. 1019–1064.
- (2014b). "Indicative Conditionals, Strictly". In: *Manuscript, Cornell*.
- Stojnić, U. (2016). "Content in a Dynamic Context". In: *Noûs*.
- (2017). "One's Modus Ponens: Modality, Coherence and Logic". In: *Philosophy and Phenomenological Research* 95.1, pp. 167–214.
- Stone, M. (1997). "The Anaphoric Parallel between Modality and Tense". In: *IRCS Report* 97.
- (1999). "Reference to Possible Worlds". In: *RuCCS Report* 49.
- Txurruka, I. G. (2003). "The natural language conjunction and". In: *Linguistics and Philosophy* 26.3, pp. 255–285.
- von Fintel, K. (1998). "The Presupposition of Subjunctive Conditionals". In: *The Interpretive Tract, MITWPL* 25, pp. 29–44.
- (2001). "Counterfactuals in a Dynamic Context". In: *Ken Hale: A Life in Language*. Cambridge, MA: MIT Press, pp. 123–52.
- Warmbröd, K. (1981a). "An Indexical Theory of Conditionals". In: *Dialogue: Canadian Philosophical Review* 20.4, pp. 644–664.
- (1981b). "Counterfactuals and substitution of equivalent antecedents". In: *Journal of Philosophical Logic* 10.2, pp. 267–289.
- Willer, M. (2017). "Lessons from Sobel Sequences". In: *Semantics and Pragmatics* 10.4, pp. 1–57.
- Williamson, T. (2007). *The philosophy of philosophy*. John Wiley & Sons.
- Yalcin, S. (2007). "Epistemic modals". In: *Mind* 116.464, pp. 983–1026.