Croatian Journal of Philosophy
Vol. XXV, No. 75, 2025
https://doi.org/10.52685/cjp.25.75.6
Received: September 11, 2024
Accepted: November 8, 2024

Sobel-esque Sequences and Felicity Judgments in Philosophy of Language

ADAM MICHAEL SENNET
University of California, Davis, USA

TYRUS FISHER
Independent scholar

This paper considers reverse Sobel sequences, NPI licensing, and related speaker judgments as they bear on von Fintel-style dynamic approaches to the semantics of subjunctive conditionals. We argue that neither reverse Sobel sequences nor von Fintel's explanation of NPI licensing speak in favor of von Fintel's semantics and that the alternative Lewisian approach, augmented by Moss-style pragmatic considerations, can accommodate and predict the relevant data at least as well as, and in some cases better than, von Fintel's view. Our arguments include counterexamples to von Fintel's semantics with respect to both its treatment of reverse Sobel sequences and NPI licensing.

Keywords: Subjunctive Conditional; Sobel sequence; pragmatics; dynamic semantics; counterfactual

1. Introduction

Work in philosophical semantics typically proceeds without a clear-cut answer to the question of how infelicity data should guide semantic analyses. It would be great if we could reliably infer constraints on positing semantic values from utterance or sentence infelicity. Sadly, we should know better.

A case in point concerns Sobel Sequences (hereafter SSs) and their reverse counterparts, Reverse Sobel Sequence (RSSs). These provide a nice example of unclarity about how felicity judgments should guide our theorizing. For, if the semantic import of the data here were clear, much of the controversy surrounding RSSs would evaporate. It's agreed

by all that the following sequence is felicitous given normal assumptions about parades and the presence of Pedro Martinez at the relevant parade:

(1) a. If Sophie had gone to the parade she would have seen Pedro.
b. But if Sophie had gone to the parade and been stuck behind a tall person, she wouldn't have seen Pedro.

Even Al Hájek (ms.), who argues that counterfactuals such as (1a) and (1b) are both false (regardless of order) will agree that the above sentences uttered in this sequence are felicitous. On the other hand, the reverse counterparts of such sequences are often clearly infelicitous:

(2) a. If Sophie had gone to the parade and been stuck behind a tall person, she wouldn't have seen Pedro.
b. # But If Sophie had gone to the parade she would have seen Pedro.

No one denies the infelicity of (2b) when uttered after (2a), and it is tempting to explain this infelicity via a story involving the claim that (2b) is false when uttered after (2a). If we could be sure this was the correct way to interpret the infelicity data here, we would be sure of an interesting and substantive constraint on the semantics of subjunctive conditionals. Indeed, the infelicity of (2b) here has been taken as motivation and even evidence for a range of theories that predict its falsity. Kai von Fintel (2001) provides what we take to be the paradigm case of an account that predicts the truth of (1a) and (1b) but the falsity of (2b) when uttered in the order given by (2) (Cf. von Fintel 2001: 132, 146, endnote 8).[1]

On the other hand, there is work that explains infelicity in terms of some type of conversational incoherence. On such a view, the infelicity of (2b) is to be explained by appeal to an assertability failure, while preserving the possibility that (2b) is true in some contexts even when uttered after (2a). We will take Sarah Moss' (2012) augmentation of a Lewis-style semantics with her pragmatic principle (EI) as our paradigm case here.

In any case, a first point to appreciate is that the infelicity data doesn't give an obvious advantage to either view, but it does make questions about the right theoretical treatment here pressing and interestingly hard to settle.

In this paper, we defend the following claims:

(a) *Against If to Might* Von Fintel touts a proposed entailment of sentences like (2a) as evidence for his theory's explanation of the falsity of sentences like (2b) (von Fintel 2001, Gillies 2007, von Fintel & Gillies 2012). But there are counterexamples to the alleged entailment.

[1] Von Fintel's 2001 and Gillies' 2007, and Moss' 2012 are the best known treatments of RSSs. In this paper we do not consider Gillies' treatment. This is primarily because Gillies' machinery is more complex. Nevertheless, our points in this paper apply for the most part to Gillies' treatment.

(b) *NPIs and SDE* Von Fintel touts his explanation of negative-polarity-item licensing in conditional antecedents as evidence for his view. But von Fintel's explanation overgenerates by predicting NPI licensing in environments that don't license NPIs, so it doesn't seem like his explanation works.

(c) *Need Pragmatics* Von Fintel's theory utilizes only a limited range of tools to explain the predicted falsity of (2b). To substantially improve the empirical coverage of theories like his, one would need pragmatic elements of the same sort that Moss already appeals to directly in her treatment.

(d) *True RSSs* Some RSSs are comprised of sentences that are jointly true regardless of their order of utterance. But von Fintel's theory entails that the second sentence of any RSS is inconsistent with the first.

(a), (b), (c), and (d) are considerations that count against accepting von Fintel's view. Interestingly, each of (a) through (d) concern features of von Fintel's theory that others have thought count in favor of von Fintel's view over the Moss-Lewis view, but we show they do not count as such.

Section 2 describes von Fintel's theory and the Moss-Lewis view. Section 3 presents and defends (a) and (b), and Section 4 does the same for (c) and (d).

All in all, we think that RSSs and their surrounding phenomena end up supporting views like Moss' (2012) better than views like von Fintel's. This is surprising given how RSS's and the related felicity-judgment data have been used to motivate and support von Fintel's theory over Lewisian alternatives.[2]

2. Background: von Fintel's theory and the Moss-Lewis view

Crediting J. Howard Sobel for calling his attention to such sequences, David Lewis (1973b: 10) observed that the sentences of SSs are quite often jointly assertable. But the joint truth of such a pair would constitute a counterexample to the validity of *Antecedent Strengthening* (AS):

(AS) $\phi > \psi \; F \; (\phi \wedge \chi) > \psi$[3]

Since (AS) is predicted to fail on a Lewis-Stalnaker theory, the felicity of Sobel sequences was thought by Lewis and others to speak in favour of such theories. But von Fintel (2001) and Gillies (2007),

[2] Cory Nichols (2017) has recently presented argumentation in a similar vein to some of our argumentation in this paper. We note that the present paper differs from Nichols' in that we consider von Fintel's argumentation concerning NPI licensing, and we consider a number of cases not considered by Nichols (and he considers some cases we do not).

[3] We use '*F*' here to denote natural-language entailment. We will let context disambiguate when we use this symbol to denote the entailment relation of some formal theory.

following an observation of Irene Heim's, present a problem for Lewis-Stalnaker theories: reversing the order of utterance for the sentences of a felicitous Sobel sequence makes the last sentence unassertable.[4]

2.1 Von Fintel's theory

Von Fintel (2001) takes the infelicity of RSSs as a data point and a mark of the falsity of their second coordinate. The basic idea of von Fintel's theory and his explanation of the infelicity of RSSs is that a context carries a *modal horizon* with it. This modal horizon comprises the possibilities that are available for interpreting subjunctive conditionals. If we consider a discourse-initial subjunctive conditional, then on von Fintel's view a counterfactual is to be evaluated just as in Lewis' theory of counterfactuals (assuming Limit):[5] a counterfactual $\phi > \psi$ is true just in case all the nearest ϕ worlds are ψ worlds.

But, if a more distant possibility, χ, is mentioned in the antecedent of a subjunctive conditional, then the modal horizon expands just enough to let in all the nearest χ worlds plus all the worlds at least as close to the actual world as the nearest χ worlds. Von Fintel motivates this expansion by positing a presupposition carried by subjunctive conditionals. The presupposition is that the modal horizon contains a world satisfying the antecedent. Thus, if the modal horizon doesn't contain such a world, it expands to include one. In the formal semantics, the modal horizon is modeled as a set of worlds and any expansion comprises a superset, as above, of the worlds previously in the horizon.

For simplicity of presentation, we consider only bare conditionals (conditionals with no embedded conditionals or other modals). We describe the theory more precisely for this special case with a few definitions:

> *The nearness relation* $\leq$ is a function from worlds to orderings of worlds corresponding to some metric of similarity assumed to be given by context.[6]

[4] As noted by Moss (2012 footnote 5), von Fintel credits Heim with noticing the infelicity of RSSs (von Fintel 2001: 130).

[5] Limit being the thesis that for any proposition ϕ and any world w, there is a closest ϕ world to w, provided there are some ϕ worlds at all. Assuming Limit typically allows one to simplify one's preferred statement of truth conditions for counterfactuals. The Limit assumption makes a difference to which counterfactuals come out true at a world (Cf. Lewis 1973b: 20). The affect of Limit on the logic generated by a semantics is more subtle. For a system of spheres (SOS) semantics, Limit has no characteristic axiom (Cf. Lewis 1973a: 121). However, Nute shows that adding Limit may change the theory's deduction principles. In particular, every SOS semantics satisfying Limit validates the deduction principle GCP (Nute 1980: 72).

For more on discoveries about the content and consequences of Limit Assumption(s) in formal frameworks, see William Starr's (2019) SEP entry, "Counterfactuals" (in particular, the supplement, "Formal Constraints on Similarity") as well as Stefan Kaufmann's (2017) "The Limit Assumption."

[6] We make no assumptions here about the properties of these orderings. However, following Lewis, it is typical to hold that the orderings should be total, reflexive,

Accessibility functions An accessibility function, f^n, is a function given by context from worlds to sets of worlds (modal horizons). We may think of each accessibility function as mapping a world of utterance to the contextually-live counterfactual possibilities. Letting f^0 denote a discourse-initial accessibility function, we assume for simplicity that $f^0(w) = \{w\}$. We assume a given discourse and stipulate that the counting numbers indexing the accessibility functions track the number of counterfactuals uttered in this discourse.

Then, on von Fintel's theory we can define the *context change potential* (CCP) of a counterfactual $\phi > \psi$ as a function from accessibility functions to accessibility functions such that:

CCP $f^n|\phi > \psi|^{\leq} = \lambda w.\ f^n(w) \cap^{\cup} \{w^l : \forall w^{ll}[w^{ll} \in [\![\phi]\!] \supset w^l \leq_w w^{ll}]\} = \lambda w.\ f^{n+1}(w)$. And the truth conditions:

vF Truth conditions $[\![\phi > \psi]\!]^{f^{n\leq}}(w) = 1$ iff $\forall w^l \in f^n|\phi > \psi|^{\leq}(w)$: $w^l \in [\![\phi]\!] \supset w^l \in [\![\psi]\!]$,

equivalently:

iff $\forall w^l \in f^{n+1}(w)$ $[w^l \in [\![\phi]\!] \supset w^l \in [\![\psi]\!]]$.[7]

In effect, we have an analysis of the counterfactual as a strict conditional $\Box(\phi \supset \psi)$, where the necessity operator is interpreted by the quantified expression: $\forall w^l \in f^{n+1}(w)$, and the uttered counterfactual determines the operative set of accessible worlds, $f^{n+1}(w)$ (the modal horizon), as a function of the set $f^n(w)$ and the antecedent ϕ. Hence, the range of quantification continually expands so as to track the contextually-live subjunctive possibilities. As von Fintel (2001) highlights at the start of his paper, on this analysis the meaning of a counterfactual has two aspects: "it alters the initial context c [corresponding above to $f^n(w)$] to a new context c^l [corresponding above to $f^{n+1}(w)$]. . . and maps c^l to the proposition p [the proposition expressed by the counterfactual] in a systematic. . . way" (von Fintel 2001: 123; our brackets). Thus we

and transitive (i.e., they are total preorders). But see Pollock (1976: 43–44, 1981), Hiddleston (2005) and Briggs (2012) for presentation and/or motivation of semantic theories that violate totality.

[7] Given the definitions above, we can present Lewisian truth conditions (assuming Limit) as follows:

Weird Lewis Truth Conditions

$[\![\phi > \psi]\!]^{f^n}(w) = 1$ iff $\forall w \in f^0(w) \cap^{\cup} \{w^l : \forall w^{ll}[w^{ll} \in [\![\phi]\!] \supset w^l \leq w, w^{ll}]\}$: $w^l \in [\![\phi]\!] \supset w^l \in [\![\psi]\!]$

The above is equivalent to more typical presentations of Lewisian truth conditions to the effect that a counter- factual $\phi > \psi$ is true just in case all the nearest ϕ worlds are ψ worlds. Of course, referencing the indexed accessibility functions we used to characterize von Fintel's theory is unmotivated beyond the purpose of comparison with the von Fintel semantics. This is because, on a Lewisian view, the relevant accessibility function does not change across utterances of counterfactuals (at least not as a matter of semantical rules). That is, $\phi > \psi$ is true just in case all the ϕ worlds nearest to the world of utterance according to $\leq w$ are ψ worlds.

have the two-step procedure of semantic interpretation displayed by von Fintel (2001) at the outset of his paper.

(1) $c\,|\,\alpha\,|\, = c^{l}$
$[\![\alpha]\!]^{c}_{l} = p$

Applying von Fintel's account to Sobel and reverse Sobel sequences is straightforward. Suppose (1a) is uttered at a context in which what it expresses comes out true. Then, given standard assumptions, all the nearest worlds where Sophie went to the parade are worlds where she saw Pedro. None of those worlds are worlds where she got stuck behind a tall person. However, on von Fintel's theory, an utterance of (1b) forces the modal horizon to change in such a way as to include all the nearest worlds where Sophie is stuck behind a tall person because the initial set contained no such worlds. And, presumably, at all of those worlds, Sophie's view of Pedro is obscured. So both (1a) and (1b) are true.

On the other hand, if the order of utterance is reversed, so that (2a/1b) is uttered first, the modal horizon is updated to include some worlds where Sophie went to the parade and got stuck behind a tall person. The resulting modal horizon contains parade worlds where Sophie didn't see Pedro. Consequently, when (2b) is uttered, there are worlds in the modal horizon that satisfy its antecedent while falsifying its consequent.

Von Fintel's proposed semantics has the following two interesting features, both of which he explicitly considers advantages. First, his semantics validates the inference from $(\phi \wedge \psi) > \chi$ to $\phi > \psi$.[8] This is easy to see: if $(\phi \wedge \psi) > \chi$ is uttered truly, this will have the effect of expanding the modal horizon so that it includes a world that satisfies $\phi \wedge \psi$. And that is just what it takes to satisfy $\phi > \psi$.[9] Second, von Fintel's semantics, in virtue of treating subjunctive conditionals as strict, renders the antecedents of subjunctive conditionals as downward entailing contexts (DECs). This, von Fintel claims, helps explain why subjunctive conditionals license Negative Polarity Items (NPIs) such as 'any' and 'even'. DECs have long been thought to hold the key to understanding NPI licensing. We discuss this apparent advantage in section 3.2 below.

2.2 Moss' pragmatic account

As noted, RSSs look problematic for Lewisian theories of counterfactuals. These theories predict that the coordinates of an RSS will be true

[8] Here $\phi > \psi$ is the dual of $\phi > \psi$ in the typical (if controversial) sense that the latter is equivalent to $\neg(\phi > \neg\psi)$. In pseudo-English: the inference from 'If it were that ϕ and ψ, then it would be that χ' to 'If it were that ϕ, then might ψ'.

[9] That is, unless there are no $\phi \wedge \psi$ worlds, but then the context can't satisfy the utterance due to presupposition failure. We discuss this point below.

together just in case the coordinates of its corresponding SS are true together. But that is just what Heim's RSSs put pressure on.[10]

Here is a standard statement of Lewisian truth conditions (with Limit):

Lewis truth conditions $[\![\phi > \psi]\!]\ (w) = 1$ iff $\{w^l : w^l \in [\![>\phi]\!] \wedge \forall w^{ll} \in [\phi],[w^l \leq_w w^{ll}]\} \subseteq [\![\psi]\!]$

Moss (2012) defends the adequacy of Lewisian truth conditions from the challenge posed by RSSs by augmenting them with a pragmatic principle, EI. Moss' formulation of EI is:

(EI) It is epistemically irresponsible to utter sentence S in context C if there is some proposition ϕ and possibility μ such that when the speaker utters S:
(i) S expresses ϕ in C
(ii) ϕ is incompatible with μ
(iii) μ is a salient possibility
(iv) The speaker of S cannot rule out μ (Moss 2012: 568).[11]

(EI) is independently plausible. Further, anyone who likes a broadly Gricean story about a cooperative principle—in particular the maxim of quality—or is sympathetic to knowledge as a norm of assertion should find it plausible that (EI) explains the infelicity of utterances that appear to violate it. Moreover, (EI)'s power to make plausible in-

[10] We have been writing throughout of "Lewisian semantics" rather than "Stalnaker-Lewis semantics" because we are restricting our attention to theories involving accessibility functions from worlds to *sets* of worlds, rather than from worlds to worlds as in Stalnaker (1968). Though of course, for (nearly?) all purposes, special cases of Lewisian semantic theories involving functions from worlds to singleton sets are equivalent to the Stalnakerian machinery.

[11] Notice that (EI) concerns the speaker's epistemic condition and, correspondingly, the felicity of the utterance.

However, Moss claims the following in her endnote 12:

> For simplicity, I talk as if infelicity is a property of utterances. Strictly speaking, infelicity is audience- relative: an utterance sounds infelicitous to an agent insofar as she takes the resulting assertion to be epistemically irresponsible (Moss 2012: 585).

A few points worth taking seriously: First, Moss conflates infelicity with judgments of infelicity. This puts into interesting relief the question of what is our target of explanation—is it infelicity judgments or utterance infelicity? These are not obviously the same.

Second, and relatedly, this raises a meta-puzzle about RSSs: if the infelicity depends on the epistemic status of the speaker, why is it that when we encounter cases like (2) we detect any infelicity at all? Notice that (2a) and (2b) aren't being asserted. (Moss isn't really telling us about Sophie and some parade.) As such, there is no irresponsibility to judge. It's tempting to respond by saying that in considering (2a) and (2b) we imagine someone asserting them and then make the judgment. This may be right, but why do we take our imaginary speaker to have an impoverished epistemic state? It's strange that we aren't more charitable towards our own imaginary conversants. It's worth noting that von Fintel's view has an answer to our question: the dynamics of the system predict the falsity of the sentence at the context, irrespective of the speaker's epistemic state.

felicity predictions applies to utterances in general rather than to the particular case of conditionals. Consider the following example.

(3) a. We're going to have a good time skiing chair 10 at Kirkwood today.
b. But Chair 10 has been closed a lot this week because of high winds.

On the other hand, the following sounds bad

(4) a. Chair 10 has been closed a lot this week because of high winds.
b. # But we're going to have a good time skiing chair 10 at Kirkwood today.

EI facilitates an explanation of what is going on here: When one speaker asserts that Chair 10 has been closing a lot due to high winds, an effect of this is that the possibility that Chair 10 will be closed today becomes salient. We are not in an epistemic position to rule out that Chair 10 is on wind hold today and this possibility is incompatible with having fun skiing lines off of Chair 10. Hence, the infelicity is predicted.

Once we have EI as a plausible constraint on epistemically-responsible assertion, we can utilize it to explain the infelicity of RSSs. Here is the initial RSS again:

(2) (a) If Sophie had gone to the parade and been stuck behind a tall person, she wouldn't have seen Pedro.
(b) # But If Sophie had gone to the parade she would have seen Pedro.

The idea is that asserting (2a) raises to salience the possibility that if Sophie had gone to the parade she might have been stuck behind a tall person. And on Lewis' semantics, this possibility is inconsistent with 2b.[12] Hence, assuming the newly salient possibility can't be ruled out, EI predicts the infelicity. Notice, however, that infelicity is not falsity and Lewis' semantics tells us that (2a) and (2b) are perfectly consistent.

2.3 Scorekeeping

One might think that as far as RSSs go, the score looks tied for von Fintel vs Moss-Lewis. However, von Fintel's view retains some apparent advantages. First, on Moss' account it is somewhat mysterious why uttering "If Sophie had gone to the parade and been stuck behind a tall person, she wouldn't have seen Pedro" should make "If Sophie had gone to the parade, she might have been stuck behind a tall person" a salient possibility. Certainly, EI tells us that when such a possibility is raised to salience and we are not in an epistemic position to rule it out, it is irresponsible to utter something incompatible with this possibility. But

[12] On Lewis' semantics, '$P > R$' is true iff there is an R world among the nearest P worlds. But if all R worlds are $\neg Q$ worlds, '$P > Q$ is false. (Where, again, $\phi > \psi$ is the dual of $\phi > \psi$ in the typical (if controversial) sense that the latter is equivalent to $\neg(\phi > \neg\psi)$.) Moss provides derivations illustrating this for a system corresponding to Lewis' favored semantics (Moss 2012: 569–572).

it doesn't tell us anything about how or under what conditions such a possibility is raised to salience.[13]

On the other hand, von Fintel's semantics gives us a story about how the possibility at issue is relevant that is as explicit as one could want—on von Fintel's theory, a subjunctive conditional of the form $(\phi \wedge \psi) > \chi$ entails $\phi > \psi$.

The second apparent advantage for von Fintel's view is that, on account of treating subjunctive conditionals as strict conditionals, the view can explain the fact that the antecedents of subjunctive conditionals licence negative polarity items. Given that being downward monotone gives rise to the licensing of NPIs, his dynamic approach explains the relevant NPI licensing, since strict conditionals are downward monotone in their antecedents.

The first supposed advantage is actually a disadvantage. Below we offer counterexamples to the entailment at issue. The second advantage is merely apparent as well: Building on work by Anastasia Giannakidou (2006, 2011), we show that being a downward entailing environment is neither necessary nor sufficient for licensing NPIs.

3. Against if to might, and NPIs and SDE

3.1 Against if to might

On von Fintel's semantics $(\phi \wedge \psi) > \chi$ Strawson entails $\phi > \psi$, where *ϕ Strawson entails ψ* iff for every context such that ϕ and ψ are both defined—which requires having their presuppositions satisfied—the truth of ϕ guarantees the truth of ψ.[14] (We'll say that an inference is *Strawson valid* iff its premises Strawson entail its conclusion.) The inference from $(\phi \wedge \psi) > \chi$ to $\phi > \psi$ is Strawson valid because an utterance of $(\phi \wedge \psi) > \chi$ will have the effect of expanding the modal horizon so that it includes a world that satisfies $\phi \wedge \psi$. And that is just what it takes to satisfy $\phi > \psi$. This observation might be thought to weigh heavily in favor of von Fintel's theory and against the Moss-Lewis view.

Unfortunately for the von Fintel-style theorist, there are counterexamples to the entailment claim at issue. Consider the following example:

(5) Context: We are discussing (the fictional) Gandalf the White's awesome power in *The Lord of the Rings*.

[13] This was first brought to our attention by I-Sen Chen. Von Fintel made this point as well in a 2012 talk at the conference *What If* at the University of Konstanz (von Fintel and Gillies 2012).

[14] Strawson gets the honour on account of his application of the concept to sentences such as 'Every F is G'. Strawson argued that while 'Every *F* is *G*' doesn't entail 'Some F is G', it does when the presupposition that there is at least one F is satisfied (Strawson 1950: 343–344). Von Fintel (other paper) defends the view that a Strawson Entailing context licenses NPIs even if it is weaker than the requirement of downward entailment.

a. If Sophie had gone to the parade and had her eyes cursed by Gandalf, then she wouldn't have seen Pedro.
b. # So if Sophie had gone to the parade, she might have had her eyes cursed by Gandalf.[15]

Surely (5a) is true. But is a subsequent assertion of (5b) true or felicitous? We feel certain that (5b) is false since Gandalf doesn't exist. It seems wild to accept (5b) on the basis of an acceptance of (5a).

Of course, there are some philosophical moves available to von Fintel but the quick fixes don't seem to offer the desired succor. An obvious idea is to deny worlds that satisfy the antecedent of (5a) a place within the modal horizon. The modal horizon is there to track possibilities and Gandalf showing up is not a possibility to take seriously. This, in effect, is to mimic part of Moss' explanation though implemented in the semantics. But this seems like a bad move—if the modal horizon comprises possibilities being tracked, then pretty clearly it should sometimes admit worlds where Gandalf curses eyeballs. This can be illustrated by considering cases involving modal subordination. The following sounds fine and seems to be a case of elaborating on possibilities:

(6) a. If Sophie had gone to the parade and had her eyes cursed by Gandalf, then she wouldn't have seen Pedro.
b. Yeah, she would have wandered around terrified and bumping into people. It would have been a disaster.

To make matters worse, there are infelicitous RSSs involving similar possibilities:

(7) a. If Sophie had her eyes cursed by Gandalf but then drank the magical antidote, she would have seen Pedro.
b. # But If Sophie had her eyes cursed by Gandalf, she wouldn't have seen Pedro.

[15] Regarding the Gandalf case, one commentator reports that since the initial context of (5) is underspecified, he wonders if in fact (5b) might sound felicitous when uttered after (5a) in any adequately specified context such that (5a) is felicitous. Some have also worried that contexts involving talk of fictional characters are atypical and might muddy our intuitions. Here is a similar case involving real people and a more detailed context:

(i) Context: Tim lives in India, and that's where he is today. Tim is a funny guy so whenever the authors of this paper see him he makes jokes that we laugh at. There was a time when Tim and the authors of this paper would often meet for beers at a local brewery. The authors of this paper almost met for a beer at that brewery today but decided to stay home and work on this paper via Zoom instead. The following conditionals were uttered in the Zoom meeting.
a. If we had grabbed a beer at the brewery today and Tim had joined us, we would have laughed.
b. # So if we had grabbed a beer at the brewery today, Tim might have joined us.
(ib) sounds bad to us, since we know Tim is in India and couldn't have visited Fieldwork today regardless of whether we had gone there.

(7b) is pretty clearly infelicitous, and, as such, assuming the infelicity stems from the same source as the original RSS, the von Fintelian can't deny Gandalf-worlds places within modal horizons. Since a place within the modal horizon brings along the requisite entailment, a view that explains the infelicity of RSSs in this way inherits the problems of over-generation. There are less silly counterexamples to the entailment claim at issue. Some concern descriptions of empirical hypotheses, their test conditions, and what inferences are warranted.

Consider:

(8) Context: Boyle meant to put a candle in his air pump on Monday but didn't get around to it. On Tuesday he uttered the following:

a. If a vacuum had been created around the candle and it had continued to burn, then my hypothesis would have been confirmed.[16]

b. # So if a vacuum had been created around the candle, it might have continued to burn.

On the relevant (non-epistemic) reading of 'might' (as the dual of 'would'), the argument has a true premise and a false conclusion.[17]

But if von Fintel is correct, the argument is sound. Notice that Moss's EI principle is suggestive of an explanation of what's going on in (5). Take, for example, an utterance of (5b). The hearer of this assertion is in a position to rule out that Gandalf will arrive at any parades attended by Sophie. So, it is natural to expect the infelicity judgment since it is often epistemically irresponsible to make assertions that are inconsistent with a salient possibility. Gandalf's absence from the parade regardless of Sophie's presence is not merely possible, but known. (However, it remains true that *if* Gandalf were to attend a parade and curse someone's eyes, that person wouldn't see Pedro.) Keep in mind, of course, that EI is meant to articulate necessary and not sufficient conditions for infelicity, so while the explanation sits well with the principle, it is no entailment of it.

But what of (8)? Notice that Boyle and his contemporaries had not yet ruled out that the candle would keep burning. Nevertheless the relevant notion of ontic modality expressed by 'might' in (8) seems to make that sentence false as uttered. If we consider that what is important to the truth value of (8) is the set of actual physical and chemical

[16] We gather that Boyle endorsed the phlogiston theory of combustion, and in his time it was an open question whether a flame would continue to burn in a vacuum. At the time some theorists, including Boyle apparently, suspected that in the absence of gas surrounding the candle, Phlogiston particles would be able to escape more easily thus aiding combustion (Cf, e.g., Myers 2003: 20).

[17] Whether a non-epistemic reading of 'might' is the most natural reading or not, it is available. In any case, the 'might' that von Fintel's theory concerns for subjunctive might conditionals is non-epistemic.

laws or regularities, then perhaps Karen Lewis' (2018) account (which we discuss below in 4.1.1) explains this case—simply take what we just called "importance" to be relevance. However, we aren't certain this is a case involving Lewis' intended notion of relevance.[18]

3.2 NPIs and SDE

Von Fintel's view offers an explanation for NPI licensing in the antecedents of counterfactuals. 'any' is a negative polarity item:

(9) #I see any coins.
(10) I don't see any coins.

And 'any' is licensed in the antecedents of subjunctive conditionals:

(11) If I were to see any coins, I would take them. These observations call for explanation.

Bill Ladusaw (1980) famously argued that the key to NPI licensing is downward entailment. Specifically, NPIs are licensed exactly in syntactic positions where subset inferences are valid. Call the scope of such an expression an *environment*. In (10), the second argument position of 'see' is a downward entailing environment. Any substitution into that position with a term whose extension is a subset of *[*coin*]* will preserve truth. However, there are cases of NPI licensing environments that are not downward entailing (see von Fintel 1999 and Giannakidou 2006). For example, 'only' seems to license NPIs but is not downward entailing:

(12) Only Jim ate any beans.
(13) Only Jim ate any beans with mold on them.

The truth of (12) doesn't suffice for the truth of (13) since there may not be any moldy beans.[19] Similar points involve 'regret'. 'regret' licenses NPIs—such as 'ever' and 'any'—but fails to be downward entailing:

(14) Bernard regrets that he ever bought a guitar.
(15) Bernard regrets that he ever bought a polka-dotted guitar.

(14) doesn't entail (15) since Bernard bought a guitar and regretted it, but the guitar has no polka dots.[20]

[18] In any case, our observations have been independent of Karen Lewis' work and our discussion here differs from hers in that we have been considering what the impact of allowing resetting and non-expansion of the modal horizon could be on a von Fintel style semantics. In particular, we have considered how trying to ameliorate the issue on a view like von Fintel's seems to require invoking claims at a pragmatic level that have the unfortunate effect of making modal discourse utilizing counterfactuals too unlike discourse involving other modal vocabulary used to track possibilities.

[19] The claim that 'Only Jim ate some moldy beans' entails that Jim ate some beans is not without its critics. See, McCawley (1993: 312) for a similar example and discussion.

[20] See Giannakidou (2006: 577) for a similar example.

As von Fintel notices, the problem looks to be related to factivity presuppositions. Consider again the 'only' and the 'regret' cases: if the factivity presuppositions are satisfied, then the inferences look good. Accordingly, von Fintel proposed a modification of the Ladusaw approach: NPIs are licensed in environments that are downward entailing when all the con- textual and linguistic presuppositions of all the sentences at issue are satisfied. He (1999) pursues this idea to great effect by formulating and offering up the notion of Strawson entailment.

Why is the foregoing relevant? The antecedents of subjunctive conditionals license (some) NPIs. So his theory nicely predicts NPI licensing in the antecedents of conditionals: von Fintel's semantics renders the antecedents of subjunctive conditionals SDE environments.
Thus, while

(16) If John were taller, he could dunk. doesn't seem to entail:

(17) If John were taller but couldn't jump, he could dunk.

(16) does Strawson downward entail (17).

To see that the inference is Strawson valid we may reason as follows. If (16) is true and its presuppositions are satisfied, then there is either at least one world within the modal horizon at which John is taller and can't jump, or there is no such world. If there is such a world, then John can dunk at it. If there is no such world, then (17)'s presuppositions aren't satisfied so it can't constitute a counterexample to the SDE claim at issue. In either case, the inference is Strawson valid. Thus, von Fintel secures an answer for an otherwise puzzling question: why are NPIs licensed in the antecedents of subjunctive conditionals, given that the antecedents of conditionals aren't downward entailing?

By contrast, Lewis' semantics treats subjunctive conditionals as variably strict, and the antecedents of such conditionals don't validate subset inferences. As such, the Lewisian is left with no obvious explanation of why the antecedents of subjunctive conditionals license negative polarity items.[21]

3.2.1 Does SDE explain NPI Licensing?

In any case, all is not well for von Fintel's own explanation of NPI licensing. The antecedents of conditionals don't license all NPIs. For example, 'yet' appears to be an NPI.

[21] We hasten to add that strict conditionals are downward entailing and Lewis' account is one of subjunctive conditionals as variably *strict*. We think there is room for the Lewisian to offer an explanation of NPI licensing as a function of downward entailment should such a thing be desired. One might begin with the idea that NPI's are licensed in sentential contexts that are "no-variation downward monotone", where a sentential context is no- variation downward monotone just in case if the premise and conclusion of the inference involve quantification over the same set of worlds, then the sentential context is downward monotone. We won't explore this idea further here.

(18) I haven't seen her yet.
(19) #I have seen her yet.

But 'yet' is not licensed in the antecedents of conditionals:

(20) # If I had seen her yet, I would have remembered it.

If SDE contexts license NPIs and conditional antecedents are SDE, it is not clear why 'yet' isn't licensed.

Second, and more importantly, von Fintel's treatment of the antecedents as SDE environments is only as compelling as the explanation of NPI licensing. But the claim that SDE environments are NPI licensing environments is highly controversial. The trouble starts with overgeneration. Recall that a sentence φ Strawson entails ψ just in case ψ is entailed by φ given that all the presuppositions of both φ and ψ are satisfied. Giannakidou (2006) shows that it's not hard to find counterexamples to the claim that SDE environments license NPIs. Consider the following cases:[22]

(21) John is unique in eating a vegetable.
(22) Uniquely, John ate some vegetables.

It is a presupposition of (23) and (24) that someone ate kale (as evidenced, e.g., by the negation test for presupposition).

(23) John is unique in eating kale.
(24) Uniquely, John ate some kale.

Recall that for Strawson entailment the contexts that matter are all and only those in which the presuppositions of *all* the involved sentences are satisfied. Hence, the environment: '*s* was unique in having eaten *x*' is Strawson downward monotone and (21)/(22) Strawson down- ward entail (23)/(24) (respectively). However these environments do not license NPIs such as 'any':

(25) #John is unique in eating any vegetable(s).
(26) #Uniquely, John ate any vegetables.

Similar points attend focus construction:

(27) # JOHN ate any vegetable.

Similar points also attend 'ever':

(28) #It was John who ever ate a vegetable.
(29) #It was John who ever ate a vegetable.
(30) #JOHN ever ate a vegetable.
(31) #Uniquely, John ever eating a vegetable.

Problematically, these sentences are similar to one's involving 'only' which clearly does license NPIs:

(32) Only John ate any vegetables.

The problem, then, is that the explanation of NPI licensing by environments that are Strawson-downward entailing is subject to apparent counterexamples (see Giannakidou 2006).

[22] These cases are ours but inspired by some of Giannakidou's cases involving 'only' and SDE (Giannakidou 2006: 579).

Giannakidou also provides counterexamples involving modals that license some NPIs while not generating SDE environments:

(33) John may order any meal.
(34) Any visitors must register at the front desk.
(35) The search committee can hire anyone they like.
(36) The search committee would interview anyone.

Of course, the environments after the NPIs above do support subset inferences:

(37) John may order any meal. ⇒ John may order any meal on the menu.

But, they aren't SDE:

(38) John may order a meal. ⇏ John may order an expensive meal.

Von Fintel's view is that NPIs are licensed just by environments that are SDE and this forms a key part of his argument that subjunctive conditionals are strict conditionals.

Interestingly, despite licensing occurrences of 'any', one can see that many other NPIs are *not* licensed by modals:

(39) ?John may/must ever go to Paris. (John may not ever go to Paris)
(40) ?John can be all that smart. (John can't be all that smart)
(41) ?The search committee would have much time to consider it. (The search committee wouldn't have much time to consider it.)

By contrast, the antecedents of subjunctive conditionals do license these NPIs:

(42) If John had ever gone to Paris. . .
(43) If John were/had been all that smart. . .

This seems right, but it doesn't remove the mystery: what is the connection between (some) NPI licensing and modality? The claim that the connection is grounded in SDE clearly overgenerates.

Von Fintel claimed that his semantics has the virtue of preserving a powerful explanation of NPI licensing. Namely, that NPI licensing occurs in SDE environments. But if modals are generally licensers of NPIs but not generally producers of SDE environments his semantics lacks the claimed virtue. Moreover, given the widely held view that subjunctive conditionals are modals, one can't defend von Fintel's semantics by suggesting that modals are a special case of NPI licensers that don't generate SDE environments.[23]

[23] It may be that presupposition satisfaction is part of the correct story of NPI licensing, but there is reason to be suspicious that the story involves a mere appeal to Strawson downward entailment. A promising line of explanation treats non-veridical contexts as the key to NPI licensing. A function f is veridical just in case $f(p)$ entails or presupposes p. Negation is thus non-veridical and, most relevant for our purposes, so are the antecedents of subjunctive conditionals. We refer the reader to Giannakidou (2006).

4. Need Pragmatics and True RSSs

4.1 Need Pragmatics

We have argued that the supposed advantages of a strict account of conditionals like von Fintel's are not clearly advantages at all and in fact are problematic. Here we argue that von Fintel's theory utilizes an overly limited range of tools to explain the predicted falsity of (2b). And to substantially improve the empirical coverage of theories like his, one would need pragmatic elements of the sort that Moss already appeals to directly in her treatment.

Consider the following RSS-like sequence:

(44) a. If Sophie had gone to the parade and hadn't been stuck behind a tall-person, she would have seen Pedro.
b. #But, if Sophie had gone to the parade, she would have seen Pedro.

Clearly, (44b) is infelicitous. Also clear is that a strict application of the von Fintel semantics does nothing to predict the infelicity of (44b). After all, the minimal expansion of the modal horizon needed to include a world in which Sophie doesn't get stuck behind a person is no expansion at all. The infelicity remains but the explanation of the infelicity doesn't apply. As such, it looks like the von Fintel explanation (and ones that utilize similar mechanisms) of RSSs is no help in explaining similar sequences—sequences that seem importantly similar to RSSs.

On the other hand, Moss's view handles the case pretty easily: the antecedent raises to salience both the worlds at which Sophie doesn't get stuck behind a tall person as well as ones where she does. How does the antecedent make this happen? Unclear. Perhaps it is just a feature of our psychology that when we hear the possibility $\neg\phi$ mentioned and we can rule neither it or its negation out, both become contextually salient. In any case, whatever the correct story is about the mechanism at work, something like this is what happens.

What are the range of responses for the proponent of a von Fintel-type view? We will consider a few.

First, one might claim that von Fintel was trying to explain Sobel sequences and (44) doesn't comprise such a sequence. So why should von Fintel be on the hook to explain the infelicity of (44)? Our answer can probably be anticipated: A theory that explains likes alike should be preferred to one that does not. Sequences like (44) are clearly a lot like RSSs.

A second, more subtle, response: The purpose of the modal horizon is ultimately to allow participants to keep track of possibilities under discussion. An utterance of (44a) followed by (44b) typically indicates that the speaker wants to consider the contrast between worlds where the parade-going Sophie is not stuck behind a tall person with ones in which she is. Since the speaker indicates a desire to consider such worlds, cooperative partners will take their indirect salience to be rea-

son to include them in the modal horizon. As such, we will have worlds in the modal horizon in which Sophie goes to the parade and fails to see Pedro and as such, (44b) is predicted to be false on von Fintel's view after all.

Notice that the expansion of the modal horizon according to this explanation is not generated by the semantics. But it invokes the very resources that Moss calls upon to explain the infelicity—an appeal to pragmatics. Since these RSS-like sequences seem so similar to RSSs and SSs, this is highly suggestive that something like salience is at the heart of a good explanation of all Sobel-esque sequences.

Third, one might complain that our case is not like an RSS on account of conversational dynamics involving coherence. Plausibly, SSs and our RSSs depend on the discourse relation of contrast–typically marked by an occurrence of 'but' at the start of the second sentence.[24] Moreover, it is reasonable to complain that adding the word 'but' at the start of the second sentence in our RSS-like sequence generates infelicity by trying to mark a contrast that doesn't exist. (Both sentences mention a possibility in which Sophie sees Pedro.)

Fortunately, other cases seem to include the contrast characteristic of RSSs (and SSs) but still allow us to make the point we want to make:

(45) Context: Pedro participated in two parades today, one in the morning and one in the afternoon. Sophie, so far as we know, went to neither. However, we do know she couldn't have gotten out of work long enough to attend both parades but she could have gotten away to attend one of them.
   a. Heidi: If Sophie had gone to the morning parade and been stuck behind a tall person, she wouldn't have seen Pedro.
   b. Lana: # But if Sophie had gone to the afternoon parade, she would have seen Pedro.[25]

[24] This was pointed out to us by I-sen Chen.

[25] Moss offers a similar case in which the information that someone other than Sophie got stuck at the parade behind someone tall is asserted rather than mentioned in the antecedent of a conditional. Moss suggests:

> Consider the following sequence:
>
> (11a) Do you remember when Kate got stuck behind a tall person and missed seeing Pedro in her first baseball parade?
>
> (11b) # But if Sophie had gone to the New York Mets parade, she would have seen Pedro.
>
> (11a) is not a counterfactual. But it nevertheless raises the possibility that if Sophie had gone to the parade, she might have been stuck behind a tall person. My analysis predicts that (11b) is therefore infelicitous. Gillies and von Fintel do not predict this. Since (11a) is not a counterfactual, or even a modal sentence, it does not prompt any expansion of the domain over which counterfactuals quantify (Moss 2012: 578; her numbering).

As Moss points out, von Fintel and other strict conditional theorists may not be too worried about her sort of case, because such cases aren't sequences of counterfactuals. But with our sort of case, which does involve counterfactuals, this move is not available.

The only way we can hear (45b) as felicitous is if we understand its utterer to be communicating, in part, that if Sophie had gone to the second parade she wouldn't have been stuck behind a tall person. The important point for us is that on von Fintel's account the infelicity of the second sentence is not predicted. This is because on his account the antecedent of the first conditional broadens the modal horizon so as to let in worlds at which Sophie goes to the morning parade and is stuck behind a tall person. However, his story does not tell us that the modal horizon is expanded in such a way to include worlds where Sophie gets stuck behind a tall person at the evening parade. One might think that what the antecedent of the first conditional makes salient is some generic possibility in which Sophie gets stuck behind a tall person at some parade, so the modal horizon gets expanded in this way. But that is not predicted by von Fintel's semantics.

We can foresee a response on von Fintel's behalf: One might worry that the semantics von Fintel offers requires expanding the modal horizon to include all the worlds at least as close to ours as the closest antecedent worlds. Some of those worlds may not satisfy the antecedent.[26] So, for example, in the case of (45), perhaps the right thing to say is that the afternoon parade worlds where Sophie is stuck behind someone tall are as close as the morning parade worlds, and so an utterance of (45a) enriches the modal horizon with those worlds as well. In that case, we could explain the infelicity of (45b) by reference to worlds already in the modal horizon that are inconsistent with the truth of (45b). See Arregui (2009). This reply is on to something but it would require us to treat similarity very heavy handedly. Against such heavy handedness, notice that we can clearly setup the context in ways that make Sophie's going to the afternoon parade much more dissimilar to the world of utterance than the morning parade. Moreover, we can change the case to include cases that seem rather dissimilar yet the relevant judgments still obtain:

(46) Context: Sophie and Pedro love parades, but both fear subways and go out of their ways to avoid riding them.
   a. Heidi: If Sophie had gone to the morning parade and been stuck behind a tall person, she wouldn't have seen Pedro.
   b. Lana: # But if Sophie and Pedro had stood near each other on the subway, she would have seen Pedro.

The mere mention of having one's view obstructed by a tall person seems to destroy the felicity of a whole range of subjunctive conditionals somewhat independently of the nearness of the possibilities implicated in the antecedents of those conditionals.

4.1.1 Karen Lewis on the Moss-Lewis View

Recently, Karen Lewis (2017) has offered some criticism of the Moss-Lewis view that we consider below. K. Lewis (2017) offers a hybrid view

[26] Thanks to an anonymous reviewer for emphasizing this to us.

of sorts, according to which RSSs are inconsistent, but this has nothing to do with any modal horizon because her semantics for counterfactuals involves an ordering on possible worlds as a function of two distinct proper- ties of worlds: similarity and relevance. As with D. Lewis' view, the semantics she offers is one according to which $\phi > \psi$ is true iff every ϕ world nearest to the actual world is a ψ world. But when a world is relevant enough to conversational purposes this may have the consequence that the world is nearer to the actual world than some more similar but less relevant worlds. For discussion of K. Lewis' notion of relevance and how it interacts with salience and similarity, we refer the reader to K. Lewis' 2016 and 2017.

We think K. Lewis' account is fine as far as it goes, though we worry about how relevance may interact with other semantic phenomena, such as modal subordination, where salience, not relevance seems to hold the key to which worlds are tracked. (A possibility may be made salient yet simply be so irrelevant to conversational purposes that no world that realizes the possibility is ranked among the nearest worlds.) In any case, given that salience doesn't suffice for relevance, the entailment doing the explanatory work for von Fintel is absent on K. Lewis' view. We take this to be a distinguishing feature between von Fintel–style semantic views and that class of views we group with the Moss-Lewis approach because they share a reliance on features of pragmatics to explain the infelicities at issue.

We feel obliged to offer some response to some of Karen Lewis' recent (Lewis 2017) criticisms of the Moss-Lewis view. K. Lewis' criticisms are fourfold:

1 Moss cannot give a theory-neutral account of what gets raised to salience (Lewis 2017: 23).
2 Moss cannot account for infelicitous Heim sequences involving possibilities not among the most similar worlds satisfying the antecedent (Lewis 2017: 12, 23).
3 In predicting what and how many possibilities are raised to salience in various cases, Moss' view may become too complicated (Lewis 2017: 15, 23).
4 Unlike K. Lewis' view and the Fintelian view, Moss' view has difficulty handling retraction sequences (Lewis 2017: 15, 23).

We find problems 1., 2., and 4 to be the most pressing and will consider them in turn.

Regarding the first problem, recall that Moss' view is that for an RSS having the form

(n) a. $(P \wedge R) > \neg Q$
b. $P > Q$

b. is infelicitous if a. raises to salience a possibility that cannot be ruled out and that is incompatible with b. The problem is that which possibilities are inconsistent with a. is sensitive to which semantic theory of subjunctive conditionals is correct. For example, positing that $P >$

R is raised to salience works for D. Lewis but not Stalnaker, and vice versa for $P > R$. And positing that both are raised to salience works for Stalnaker but not for D. Lewis (Lewis 2017: 13).

Accordingly, K. Lewis suggests that the best option for the proponent of Moss' view is to hold that what gets raised to salience is that there is a $P \wedge R$ world among the nearest P-worlds (which works whether Lewis' or Stalnaker's semantics is correct). But this move, she argues, saddles Moss with the second problem.

Regarding the second problem, K. Lewis offers the following case:

(47) Context (quoting K. Lewis 2017: 14): Suppose Logan and Claudia are in the same group of friends, and so often end up at the same social events. But Logan and Claudia notoriously do not get along, and when they are together, it is absolutely no fun to be around them (though each on his or her own is a very fun person). The conversational participants are discussing a recent party which neither Logan nor Claudia attended because they both had work to do. While Logan is very attentive—he is unlikely to miss a deadline if he can help it—Claudia is easily distracted from her work and almost went to the party, only to be swayed at the last second not to go by her guilty conscience.
   a. If Logan and Claudia had come to the party, it would have been no fun at all.
   b. # But of course, if Logan had come to the party, it would have been fun.
   c. # But of course, if Claudia had come to the party, it would have been fun.

K. Lewis' analysis is that b. and c. are infelicitous but that the Moss view cannot explain the infelicity of c. This is because we know it was always unlikely that Logan would have attended. So we know that if Claudia had attended, Logan would not have. Hence, EI cannot explain the infelicity here because we can rule out that Logan and Claudia would have both been at the party.

We believe the case does not demonstrate K. Lewis' conclusion. The problem is that the context gives us by stipulation that it is unlikely that Logan would attend. But unlikeliness (in any of the many ways of precisifying the notion) does not make for greater distance from the actual world. Suppose a die will be rolled and that you have taken a bet that it will come up six. Observe (the fairly well-known point) that just because it is unlikely you will win, this does not mean that worlds where you win are farther away than worlds where you don't. Rather, the correct judgment seems to be that exactly one-sixth of the most similar worlds are worlds where you win. The unlikeliness of an event does not on its own make worlds where that event occurs more distant.

Notice further that if we change the context to read as follows, then 47c sounds fine:

(48) Context (compare K. Lewis 2017: 14): Suppose Logan and Claudia are in the same group of friends, and so often end up at the same social events. But Logan and Claudia notoriously do not get along, and when they are together, it is absolutely no fun to be around them (though each on his or her own is a very fun person). The conversational participants are discussing a recent party which neither Logan nor Claudia attended because they both had work to do. Logan is very attentive to work deadlines and the possibility that he would attend is ruled out. On the other hand, Claudia is easily distracted from her work and almost went to the party, only to be swayed at the last second not to go by her guilty conscience.

a. If Logan and Claudia had come to the party, it would have been no fun at all.

b. # But of course, if Logan had come to the party, it would have been fun.

c. But of course, if Claudia had come to the party, it would have been fun.

Keep in mind that if K. Lewis' claim is correct that we and the conversational participants could rule out Logan's attendance all along, then the language in our modified context should make no difference to the relevant felicity judgments. But it does.

Turning to the fourth problem now, K. Lewis offers the following example of a *retraction sequence*:

(49) a. A: If Sophie had gone to the parade, she would have seen Pedro dance.

b. B: But of course, if Sophie had gone to the parade and been stuck behind someone tall, she wouldn't have seen Pedro dance.

c. A: Alright, I guess then, if Sophie had gone to the parade, she might not have seen Pedro dance.

The sequence is felicitous, and as K. Lewis puts it, "At first, this seems like just the sort of sequence that (EI), or at least something in its spirit, is poised to account for. According to the account, [the above sequence] raises to salience the counterfactual *If Sophie had gone to the parade, she might have been stuck behind someone tall*. Since this makes [(49a)] unassertable in this context, given that it conflicts with a possibility raised to salience, it would make sense that a speaker might go back and amend or retract her initial assertion, even though it was perfectly good in the original context" (Lewis 2017: 16 [brackets added]). The problem that K. Lewis brings up is that on D. Lewis' analysis 49c has the logical form $P > \neg Q$, and on his analysis this is logically equivalent to $\neg(P > Q)$. But then 49a is logically inconsistent with 49c.

Since 49a and 49c are both true by hypothesis, K. Lewis' example seems to have the consequence that either the D. Lewis analysis of

might counterfactuals is false or the Moss-Lewis view has the unpalatable consequence that there are very many felicitous but false might counterfactuals like 49c.

Whether or not the theoretical cost of positing lots of felicitous and false sentences is too high, we see little reason why a proponent of the Moss-Lewis view cannot simply accept that utterances of might conditionals in natural language are very often syntactically ambiguous between D. Lewis' favored analysis and the Stalnaker analysis according to which *If P then might not Q* is equivalent to saying *It is epistemically possible that if P then not Q*. And we think that the context surrounding K. Lewis' retraction sequence heavily suggests an epistemic reading of the occurrence of 'might'.

In sum, we quite like K. Lewis' account of counterfactuals and what it tells us about the effect of relevance on their truth values. However, we think that while the criticisms she offers of the Moss-Lews view are interesting and worth thinking through carefully, we do not think they ultimately show the Moss-Lewis view to be in trouble.

4.2 True RSSs

The previous section aimed to show that a proper explanation of RSS-like phenomena in general involve a pragmatic dimension. One might worry that this leaves open whether von Fintel has the correct story about RSSs in particular. In this section, we consider RSSs directly and argue, pace von Fintel's semantics, that their coordinates are not semantically inconsistent.

We start from the following case involving an RSS.

(50) Context: Ours is a world at which Sophie did in fact go to the parade this morning, and she saw Pedro there. However, neither Heidi or Lana know this. Part of their conversation proceeds as follows.
   a. Heidi: If Sophie had gone to the parade this morning and had her view (of Pedro) blocked by a tall person, she wouldn't have seen Pedro.
   b. Lana: #But if Sophie had gone to the parade this morning, she would have seen Pedro.

Lana's utterance is infelicitous in context. However, even a counterfactual skeptic like Al Hájek will accept the truth of both of these subjunctive conditionals (since (50a) has an antecedent that entails its consequent and (50b) has a true antecedent and consequent—i.e., it is a true-true subjunctive conditional. Many philosophers accept Strong Centering, and that principle guarantees the truth of true-true subjunctive conditionals.[27]

But Strong Centering is ultimately inessential to the issue here—all that would be needed to show that the Fintelian semantics is incorrect

[27] By *Strong Centering* we mean the inference principle: *(SC)* $\phi \wedge \psi \Rightarrow \phi > \psi$.

would be a demonstration of the existence of some true-true subjunctive conditionals like (50b) that are true and for reasons not defeated by utterances of conditionals like (50a). But on von Fintel's view, an utterance of (50b) is inconsistent, so *cannot* be true as uttered. But we think it is pretty clear that if Heidi and Lana were to go on to disagree about the truth of Heidi's utterance, we could easily settle the issue in favor of Heidi's side by pointing out that we were at the parade with Sophie and we know she saw Pedro.

Now, it still it remains open to the Fintelian to simply hold on to the claim that no RSS can have simultaneously true coordinates. But here are two more, we think very unpalatable consequences of this insistence:

First, if von Fintel's semantics is correct, we get a surprising disanalogy between subjunctive conditionals and other modals. Uttering a subjunctive conditional when it is known that the antecedent is true typically generates infelicity. [28]

People very rarely utter subjunctive conditionals when they know that both antecedent and consequent are true and such utterances typically generate infelicity. But notice how similar this looks to what happens to utterances of other modal expressions in analogous circumstances:

(51) Context: We are at a party and have just both observed that Jordan is in attendance. Then I say one of the following:
  a. Jordan might attend the party tonight.
  b. Jordan should at least make an appearance at the party tonight.

It is certainly odd for a speaker to say that someone *might* or *should* attend a party when participants in the conversation know the person has attended, and the oddness here is similar to the oddness of uttering a conditional in the subjunctive mood when conversational participants know that the antecedent is true. However, despite the oddness of (51a) and (51b), the natural judgment, we think, is that it is perfectly true that Jordan might attend the party tonight given that he is in fact in attendance tonight, and it may well be true that Jordan should attend regardless of whether I say anything about that. The point is that the Fintelian must claim that unlike in cases involving other modals, analogous utterances of subjunctive conditionals cannot be true when uttered in an RSS.

Further, if we avail ourselves of rigidifying operator we can construct sequences of conditionals that come out true on von Fintel's semantics, but which appear to generate infelicity in exactly the way that RSSs do: Let '*Actually*' be an operator that shifts the circumstance of evaluation to the actual world of utterance (in a manner something like that introduced in Kaplan (1989)). Now consider the following variant of (50):

[28] But see Iatridou (2000) for discussion.

(52) Context: Ours is a world at which Sophie did in fact go to the parade this morning, and she saw Pedro there. However, neither Heidi or Lana know this. Part of their conversation proceeds as follows.
   a. Heidi: If *Actually* (Sophie had gone to the parade this morning) and Sophie had been stuck behind a tall person, She wouldn't have seen Pedro.
   b. Lana: #But if *Actually* (Sophie had gone to the parade this morning), *Actually* (she would have seen Pedro).

More colloquially:

   a. Heidi: If Sophie had actually gone to the parade this morning and she had gotten stuck behind a tall person, she wouldn't have seen Pedro.
   b. Lana: #But if Sophie had actually gone to the parade this morning, she actually would have seen Pedro.

Notice that the first conditional is non-vacuously true because its first conjunct is necessarily true and every world in the modal horizon that satisfies the second conjunct also satisfies the consequent. And the second coordinate is true because both antecedent and consequent are necessarily true given that Sophie actually went to the parade and saw Pedro. However, despite the fact that the Fintelian semantics tells us both sentences of 52b are true, the sequence looks like it generates infelicity *in exactly the way* that "real" RSSs do.

Of course it is open to the Fintelian to deny that such an operator could be added to English (not to mention impossible that such an operator already is already present in our language). But that should look implausible to anyone who isn't already committed to the Fintelian picture. Again, our point is not that the Fintelian is unable to hold on to their view in the face of the above considerations, but rather that for anyone with an open mind, the above considerations show some serious drawbacks of the view and should count heavily against it.

We also wish to point out that our (50) is somewhat different than the following case presented by Moss:

(53) Context (quoting Moss 2012: 574): Suppose John and Mary are our mutual friends. John was going to ask Mary to marry him, but chickened out at the last minute. I know Mary much better than you do, and you ask me whether Mary might have said yes if John had proposed. I tell you that I swore to Mary that I would never actually tell anyone that information, which means that strictly speaking, I cannot answer your question. But I say that I will go so far as to tell you two facts:
   a. If John had proposed to Mary and she had said yes, he would have been really happy.
   b. But if John had proposed, he would have been really unhappy.

We think that a proponent of von Fintel's view has a reasonable explanation of what is going on in (53). Let's look more closely, then, at von Fintel's claim that the second sentence of an RSS will be false and why he thinks that the truth conditions of such sentences are sensitive to context in this way.

Von Fintel writes:

> What I mean by "no longer true" is not that the objective facts have changed. It is the parameters of the discourse that have changed so that the proposition expressed by the first counterfactual in the initial context can no longer be ex- pressed by the same linguistic expression in the new context. Compare the fact that the claim that France is hexagonal may be true in a context where it is preceded by Italy has the shape of a boot, but may cease to be true in a later context where the standards of precision have been sharpened (von Fintel 2001: 146, note 8).

The Fintelian can say that the utterance of (53b) after (53a) *is* inconsistent and, hence, surprising.[29] This effect of surprise is capitalized upon in order trigger the inference on the part of the listener. Assuming the speaker is taken to be reliable and in this instance committed to the activity of cooperative communication, the hearer must perform a pragmatic repair on the modal horizon so as to make the second utterance come out true, and the only way to do this consistent with (53a) is to kick all the worlds out of the horizon where John asks and Mary accepts. This is a pragmatic explanation to be sure, but it is precisely because the utterance of (53b) was initially inconsistent, as predicted by von Fintel, that the oddity of the utterance was striking enough to spur the pragmatic repair and the intended inference. This looks okay for von Fintel's view.

Our case involving Lana and Heidi is different. In our case, the second coordinate remains infelicitous by the lights of Heidi and any other listener unaware that Sophie did in fact go see Pedro. So the proponent of von Fintel's theory can't say that it is some accommodation effect making the second RSS coordinate ultimately come out true. Rather, Sophie went to the parade and saw Pedro so, contra von Fintel, both (50a) and (50b) are true together despite the generated infelicity.

This still leaves the questions of why Lana's utterance is infelicitous in context. However, the Moss-Lewis view supplies a story that we think should be counted among the best candidate explanations: Lana's utterance is infelicitous because it runs afoul of the EI principle: The (nonactual) possibility of being stuck behind a tall person should Sophie have attended the parade was brought to salience by Heidi's utterance of (50a), and neither Heidi or Lana were in an epistemically appropriate position to rule that possibility out. And the Moss-Lewis style explanation of the infelicity generated in 52b is the same, which is exactly what a theorist should desire.

[29] Moss describes such an explanation (Moss 2012: 578–579).

5. Conclusion

In this paper we considered the case for von Fintel-style semantic theories of subjunctive conditionals. A great deal hangs on this as von Fintel's semantics (and ones like it) recommend a fairly non-conservative approach to the semantics of subjunctive conditionals. Given the generally agreed upon links between subjunctive conditionals and modals, the departure is most likely not limited to subjunctive conditionals (as discussed in Section 3.1).

The motivation for von Fintel-style theories comes in large part from intuitions regarding the truth values of conditionals in a sequence. There are long standing questions in the philosophy of language regarding our methodology: when and how far can we trust our ability to separate truth value judgments from judgments of felicity? After all, trained philosophers know there is a difference between falsehoods and infelicitous truths. Given that we are good at this, why can't we settle the dispute more directly by intuiting whether or not the second sentence of an RSS is true? If it's false then it is hard to see how a view like von Fintel's doesn't win the day. If it's infelicitous but true, it looks like views like von Fintel's are basically non-starters. It's downright odd that we have to try to settle these disputes indirectly. But that's where we are. And to that extent, we think the case for von-Fintel semantics is pretty weak and the case against it pretty strong.

References

Arregui, A. 2009. "On similarity in counterfactuals." *Linguistics and Philosophy* 32 (3): 245–278.

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

von Fintel, K. 1999. "NPI licensing, Strawson entailment, and context dependency." *Journal of Semantics* 16: 97–148.

von Fintel, K. 2001. Counterfactuals in a dynamic context." In M. Kenstowicz (ed.). *Kan Hale: A Life in Language*. Cambridge: MIT Press, 123–152.

von Fintel, K. 2012. "Subjunctive conditionals." In D. Graff Fara and G. Russel (eds.). *The Routledge Companion to the Philosophy of Language*. New York: Routledge, 466–477.

von Fintel, K. and T. Gillies. 2012. "Should von Fintel and Gillies be mothballed?" Invited talk at conference "What if? On the Meaning Epistemology, and Scientific Relevance of Counterfactual Claims and Thought Experiments", University of Konstanz. Konstanz Germany. October 27, 2012. Slides retrieved from http://web.mit.edu/fintel/fintel-gillies-2012-mothball-konstanz.pdf on 4/22/2018.

Giannakidou, A. 2006. "Only, emotive factive verbs, and the dual nature of polarity dependency." *Language* 82 (3): 575–603.

Giannakidou, A. 2019. "3. Negative and positive polarity items". In P. Portner, C. Maienborn and K. von Heusinger (ed.). *Semantics - Sentence and Information Structure*. Berlin, Boston: De Gruyter Mouton, 69–134.

Gillies, T. 2007. "Counterfactual scorekeeping." *Linguistics and Philosophy* 30: 329–360.

Hiddleston, E. 2005. "A causal theory of counterfactuals." *Noûs* 39 (4): 632–657.

Iatridou, S. 2000. "The grammatical ingredients of counterfactuality." *Linguistic Inquiry* 31 (2): 231–270.

Kaplan, D. 1989. "Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals." In *Themes from Kaplan*. Oxford: Oxford University Press.

Kaufmann, S. 2017. "The limit assumption." *Semantics and Pragmatics* 10 (18): 1–29.

Ladusaw, W. 1980. *Polarity Sensitivity as Inherent Scope Relations*. New York: Garland.

Lewis, D. 1973a. "Causation." *Journal of Philosophy* 70: 556–67.

Lewis, D. 1973b. *Counterfactuals*. London: Blackwell Publishing.

Lewis, K. 2016. "Elusive counterfactuals." *Noûs* 50 (2): 286–313.

Lewis, K. 2018. "Counterfactual discourse in context." *Noûs* 52 (3): 481–507.

McCawley, J. 1993. *Everything That Linguists Have Always Wanted to Know about Logic. . . But Were Ashamed to Ask* (2nd ed.). Chicago: University of Chicago.

Moss, S. 2012. "On the pragmatics of counterfactuals." *Noûs* 46 (3): 561–586.

Myers, R. 2003. *The Basics of Chemistry*. Santa Barbara: Greenwood Publishing Group.

Nichols, C. 2017. "Strict conditional accounts of counterfactuals." *Linguistics and Philosophy* 40 (6): 621–645.

Nute, D. 1980. *Topics in Conditional Logic, Volume 20*. Dordrecht: D. Reidel Publishing.

Pollock, J. 1976. *Subjunctive Reasoning*. Dordrecht: Reidel.

Stalnaker, R. C. 1968. A theory of conditionals. In R. Stalnaker (ed.). *Studies in Logical Theory*. London: Blackwell Publishing, 98–112.

Starr, W. 2019. "Counterfactuals." In E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition). https://plato.stanford.edu/archives/spr2019/entries/counterfactuals

Strawson, P. F. 1950. "On referring." *Mind* 59 (235): 320–344.