Linguist and Philos
DOI 10.1007/s10988-017-9209-8

Strict conditional accounts of counterfactuals

Cory Nichols[1]

Abstract von Fintel (Curr Stud Linguist Ser 36:123–152, 2001) and Gillies (Linguist Philos 30(3): 329–360, 2007) have proposed a dynamic strict conditional account of counterfactuals as an alternative to the standard variably strict account due to Stalnaker (Studies in logical theory, Blackwell, London, 1968) and Lewis (Counterfactuals, Blackwell, London, 1973). Von Fintel's view is motivated largely by so-called reverse Sobel sequences, about which the standard view seems to make the wrong predictions. (The other major motivation is data surrounding so-called negative polarity items, which I do not discuss here.) More recently Moss (Noûs 46 (3):561–586, 2012) has offered a pragmatic/epistemic explanation that purports to explain the data without requiring abandonment of the standard view. So far the small amount of subsequent literature has focused primarily on the original class of cases motivating the strict conditional view. What is needed in the debate is an examination of the predictions of the dynamic strict conditional account for a broader range of data. I undertake this task here, presenting a slew of cases that are problematic for the strict conditional view but not for Moss's view, and considering some possible responses. Ultimately I take my contribution to constitute a significant blow to the dynamic strict conditional view, though not a decisive verdict against it.

Special thanks to Boris Kment, Karen Lewis, Alan Hájek, Gideon Rosen, Paolo Santorio, Jack Woods, and Matt Moss for invaluable discussion and/or comments on earlier versions of this paper; to audiences at New York Philosophy of Language Workshop, Princeton Philosophical Society, the Eighth Barcelona Workshop on Issues in the Theory of Reference, Boğaziçi University, and the Second Belgrade Conference on Conditionals for extremely useful questions and feedback; and to anonymous reviewers in this journal for helpful suggestions.

✉ Cory Nichols
corymnichols@gmail.com

[1] Princeton University, Princeton, NJ, USA

Published online: 14 July 2017

Keywords Conditionals · Counterfactuals · Strict conditionals · Sobel sequences · Reverse Sobel sequences · von Fintel · Gillies · Moss

1 A peculiar asymmetry

Until fairly recently it seems to have gone unnoticed that counterfactuals—conditionals of the form *If A were the case, then C would be the case*—exhibit a surprising asymmetry: so-called *Sobel sequences* like those below are fine in one order but infelicitous in reverse. Suppose Jeff and Lars are delightful party guests alone, but together they always fight. Consider:

(PARTY): (a) If Jeff had come to the party, it would've been fun. (b) If Jeff and Lars had come to the party, it wouldn't have been fun.

(PARTY-R): (a) If Jeff and Lars had come to the party, it wouldn't have been fun. (b) #If Jeff had come to the party, it would've been fun.

The typical reactions are that PARTY sounds fine throughout but PARTY-R(b) is infelicitous. Here is another case: the Yankees have just won the World Series, and baseball legend Derek Jeter is sure to make an appearance at the victory parade. Consider:

(SOPHIE): (a) If Sophie went to the parade, she would see Jeter. (b) If Sophie went to the parade and got stuck in the back of the crowd, she wouldn't see Jeter.

(SOPHIE-R): (a) If Sophie went to the parade and got stuck in the back of the crowd, she wouldn't see Jeter. (b) #If Sophie went to the parade, she would see Jeter.

It isn't clear what accounts for this asymmetry. The standard semantics for counterfactuals is known as the variably strict account (VSA), due primarily to Lewis (1973, 1986) and Stalnaker (1968, 1981). On this view worlds are ranked according to their *closeness*, where the closeness of a world is a function of how *similar* it is in certain ways to the actual world. A counterfactual A $>$ C is true, then, iff (roughly) all the closest A-worlds are C-worlds.[1] Lewis puts the central thought nicely: "Roughly, a counterfactual is true if every world that makes the antecedent true *without gratuitous departure from actuality* is a world that also makes the consequent true" (1973, p. 41, emphasis added). But if facts about similarity comparisons between worlds underwrite the truth values of counterfactuals, why should their order of utterance make any difference? The similarity facts are not generally taken to be so deeply context-sensitive. Some further explanation is needed.

[1] Lewis and Stalnaker famously disagree about several further details of the variably strict account, and there are many controversial issues surrounding the notions of similarity and closeness. For a good overview see Bennett (2003). I can ignore these controversies for my purposes.

2 Strict conditionals

Philosophers had long dismissed the possibility of a strict conditional account of counterfactuals, according to which A $>$ C simply means *necessarily, if A then C*:

(SCA): A $>$ C is true iff $\Box$(A $\supset$ C)

The necessity operator seems too demanding: it might be true that if I went to the zoo (A) I would have fun (C); but surely there is *some* possible world where I go to the zoo and don't have fun (A&$\neg$C), e.g. if I get mauled by a lion; so A $>$ C is true but $\Box$(A $\supset$ C) is false. So the idea of treating counterfactuals as strict conditionals has historically not been taken seriously.

More recently, however, von Fintel (2001) and Gillies (2007) have given us reason to take the idea seriously.[2] In natural language we routinely quantify over restricted domains, and these restrictions can rapidly change throughout a conversation. Looking in the fridge, I say: "All the wine is gone; luckily, there's more in the cellar". Lest I be interpreted as contradicting myself, the first part of my utterance must be understood as quantifying over a different domain—something like *the immediately accessible wine*—from that of the second part of my utterance —something like *the wine at our disposal*. Von Fintel's view begins with the observation that strict conditionals systematically restricted in a similar way might preserve the standard truth conditions in ordinary cases while predicting non-standard truth conditions in idiosyncratic cases like reverse Sobel sequences. Lewis (1973), writing before these data were known, dismissed this style of analysis as ad hoc and defeatist. But von Fintel's account seems neither, and purports to explain the peculiar asymmetry mentioned above.

The view has two major parts. First, there is the strict conditional analysis: A $>$ C just means $\Box$(A $\supset$ C). Second, there is a *dynamic modal domain*: the set of worlds quantified over evolves throughout a conversation as speakers discuss new possibilities. More specifically, the operative modal domain D_c at any context point *c* is demarcated by the *modal horizon*, i.e. the "outer" limit that determines which worlds are included in the domain, and the modal horizon is expanded as necessary to accommodate new possibilities under discussion. Why? A quantified modal claim of the form *All φ-worlds are ψ-worlds* interpreted at a domain including no φ-worlds is vacuously true. So in order to give speakers a chance of saying something non-vacuous, conversational participants typically broaden the modal horizon until it reaches some A-worlds (A being the antecedent of the conditional), which demarcates a new domain D_{c*} at a new context c*. But when the domain already includes A-worlds no such change is needed, so $D_{c*} = D_c$.

A few further details are needed for this account to predict the asymmetry[3]:

[2] von Fintel and Gillies differ on some of the details. Most important for present purposes is that Gillies ultimately hedges on whether to appeal to a closeness-based ordering as von Fintel does. As a result his view is more difficult to evaluate, so I will focus on von Fintel's view. Even abandoning closeness, however, will only avoid the first of my four classes of problem cases discussed in the next section.

[3] These clauses and terminology are features of my preferred presentation of the view, not von Fintel's own.

(i) Expansion is *closeness-based*: worlds are ordered by closeness in the traditional way, and the modal horizon is understood as a location in the ordering. In other words, when the modal horizon is broadened, this amounts to admitting more distant worlds into D.
(ii) Expansion is *conservative*: when D expands to include some A-worlds, it expands only as much as is necessary to do so. Combined with (i), this means expansions will only include the *closest* A-worlds for the A inducing the expansion.
(iii) Expansion is *coarse-grained*: when D expands to include the closest A-worlds, it also includes all other worlds at least as close, whether or not they are A-worlds.[4]

Now we are in a position to see how the semantics works. Suppose there are no worlds in D_c where I go to the zoo. Then when I say at c that if I went to the zoo I would have fun, D_c expands until it includes some worlds where I go to the zoo. Because of (i) these will be the *closest* zoo-worlds; because of (ii) the expansion will stop there. The counterfactual is then evaluated as a strict conditional—$\Box$(*I go to the zoo* $\supset$ *I have fun*)—at the new context c* quantifying over the new domain D_{c*}, and thus is true iff every world in D_{c*} makes the antecedent false or the consequent true. The only worlds in D_{c*} where the antecedent is not false are (the closest) worlds where I go to the zoo; so if all those worlds make the consequent true, i.e. I have fun, the strict conditional is true. So A $>$ C is true so long as the closest A-worlds are C-worlds. So whenever expansion occurs the dynamic strict conditional account predicts the same truth conditions as the standard variably strict account.

But something different happens with reverse Sobel sequences. For convenience, let's say informally that φ is a *closer possibility than* ψ when the closest φ-worlds are closer than the closest ψ-worlds. In the forward versions of Sobel sequences, then, the first conditional's antecedent denotes a closer possibility than the one denoted by the second. In PARTY, for example, the possibility of Jeff being at the party without Lars must be closer than the possibility of both of them being there. After all, if it's true that it would've been fun if Jeff had come, and also true that it *wouldn't* have been fun if Jeff *and Lars* had come, then the closest worlds where Jeff

[4] For the interested reader, we could spell out the view more formally as follows. Let $\leq$ be an ordering relation such that $w \leq w'$ iff w is at least as close as w', let $\boldsymbol{\varphi}$ represent the set of worlds where φ is true, and let the abbreviation $\forall w_{\in \boldsymbol{\varphi}}$ represent the restricted quantification *all worlds* w *in* $\boldsymbol{\varphi}$. Then:

(SCA): A $>$ C uttered at a context c including domain D_c is interpreted as $\Box$(A $\supset$ C) at a new context c* quantifying over domain D_{c*}, such that:

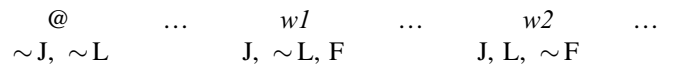
(i) if $D_c \cap \mathbf{A} \neq \varnothing$ then $D_{c*} = D_c$

(ii) if $D_c \cap \mathbf{A} = \varnothing$ then $D_{c*} = \{w\colon \forall w'_{\in \mathbf{A}}(\forall w''_{\in \mathbf{A}}(w' \leq w'') \supset w \leq w')\}$

In something closer to ordinary English: if the intersection of D_c and $\mathbf{A}$ is non-empty, i.e. there is an A-world in the initial domain, then the "new" domain D_{c*} is just D_c; but if the intersection is empty, i.e. the initial domain includes no A-worlds, then D_c is replaced by an expanded domain D_{c*} consisting of the set of worlds w such that: for any world w' in $\mathbf{A}$, if w' is at least as close as any world w'' in $\mathbf{A}$ (i.e. if w′ is a closest A-world), then w is at least as close as w'. (This last clunky bit says that D_{c*} is the set of ws at least as close as any closest A-world. This is necessary because expansion is coarse-grained, i.e. not every w included in an expansion need be an A-world.)

comes must be ones where Lars doesn't. Let's make this clearer with a diagram, where @ = the actual world, J = *Jeff comes to the party*, L = *Lars comes to the party*, F = *the party is fun*, and position from left to right represents distance from the actual world:

Diagram 1



At the actual world, @, Jeff and Lars didn't come. The closest worlds where Jeff does come, labeled *w1*, are closer than the closest worlds where both Jeff and Lars come, labeled *w2*. Supposing these worlds are not in the initial domain, SCA predicts that the domain for the forward version of the sequence will evolve in the normal way: PARTY(a) induces an expansion to include *w1*-worlds, where Jeff is at the party, but not *w2*-worlds, where Lars is there too; then (b) induces a second expansion to include the *w2*-worlds as well. But in the reverse case the larger (*w2*) expansion occurs first, so the smaller (*w1*) one is unnecessary. That is, an utterance of PARTY-R(a) (which is the same conditional as PARTY(b)) expands the domain to include *w2*-worlds, which of course also includes the closer *w1*-worlds; but then no domain change is needed to accommodate the subsequent utterance of PARTY-R (b), whose antecedent is *if Jeff had come to the party*, for the *w2*-worlds are themselves such worlds (as are the *w1*-worlds). Since no analogous method of domain contraction exists,[5] the counterfactual is evaluated at this domain, and is thus true iff every world in it makes the material conditional *Jeff comes to the party* ⊃ *it is fun* true. But *w2*-worlds are worlds where Jeff comes and it isn't fun, because Lars is there too. So the counterfactual is false in this context. (SOPHIE-R is exactly analogous.)

So a strict conditional semantics supplemented with the technology of a dynamic modal domain seems competent to get basic cases right, and appears to have an elegant explanation of the asymmetry exhibited by Sobel sequences.

Before moving on, it is worth noting that there is another major motivation for von Fintel's framework. Conditional antecedents are one of a handful of locutions that grammatically license the appearance of *negative polarity items* (NPIs), so-called because they characteristically occur in linguistically "negative" environments. Two paradigm NPIs are *any* and *ever*, which are permitted under the scope of negation, as in the sentences "We don't have any wine" or "I don't think she ever drinks", but not in their "positive" counterparts "We have any wine" and "I think she ever drinks". But NPIs turn out to occur in a handful of other, non-negative environments as well (the term is something of a misnomer), including the

[5] Von Fintel observes that in some cases we can eliminate the infelicity of a reverse Sobel sequence by explicitly signaling that the possibility just mentioned is to be ruled out, e.g.: "If Jeff and Lars had come to the party, it wouldn't have been fun. *But Lars wasn't at the party*. So if Jeff had come to the party, it would've been fun." In these cases, he says, the effect of the intermediate claim is to induce a domain contraction (to where is unclear). But in ordinary cases such contractions do not occur.

antecedents of conditionals, e.g. "If we'd drunk any wine, it would've been red" or "If she ever drinks, she drinks wine".

The further details of the relevant data and literature exceed the scope of this paper (and the expertise of its author). In short, it appears the unifying feature that explains the linguistic distribution of NPIs must be related to the entailment patterns validated by the environments they prefer,[6] so we would like a theory of conditionals whose logic agrees with those patterns. Von Fintel's dynamic strict conditional account does so, but Lewis's variably strict account does not, so this is a point in favor of von Fintel. All of this is orthogonal to the data I will discuss here, but the fact is worth mentioning, if only as a dialectical signpost, that there remains a second motivation for von Fintel's view that is completely unaddressed by this paper. So at best I will be able to claim by the end to have undermined one of the two.

3 Trouble for the strict conditional account

The discussion in the relevant literature focuses primarily on pairs of counterfactuals in which one antecedent is a logically strengthened version of the other, typically involving an additional conjunct. For example, *Jeff and Lars come to the party* is a strengthened version of *Jeff comes to the party*. But this constitutes a fairly narrow class of cases, and exploration of further data reveals several classes of cases about which the view, as it stands, makes bad predictions.

3.1 Intermediate worlds

Recall that domain expansion, according to SCA, is coarse-grained: when the domain expands to include the closest A-worlds for some counterfactual, it includes all other worlds at least as close. This turns out to be too permissive, as the following cases demonstrate.

First, given the earlier description of the party scenario, this seems true:

(NOTLARS): (a) If Jeff and Lars had come to the party, it wouldn't have been fun. (b) If Jeff but not Lars had come to the party, it would've been fun.

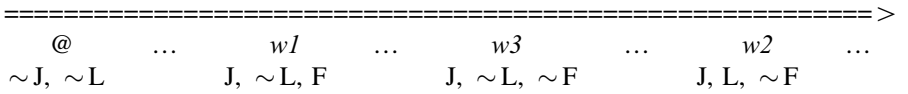
We know from a moment ago that the closest worlds where Jeff comes to the party are ones where Lars doesn't come, and it's fun (*w1*-worlds). And we know that the closest worlds where they both come, at which it's not fun (*w2*-worlds), are farther than these. (All this was represented by Diagram 1.) Now suppose further that the *w2*-worlds are *much* farther than the *w1*-worlds, i.e. involve a significantly greater degree of departure from actuality.[7] For example, suppose Jeff lives down the street

[6] Particularly patterns involving *downward entailment/monotonicity*. Von Fintel attributes this observation to Ladusaw (1979); see von Fintel (2001, pp. 132–133) for more on this, and von Fintel (1999) for more on NPIs in general.

[7] It's tempting to think and speak as though there is a metric on similarity/closeness, such that one could make good sense of a claim like "world *w* is three times as far from @ as world *w′* is". Lewis is skeptical

and nearly came to the party, but Lars was officiating a wedding in California, 3000 miles from the party in New York. Then *w1*-worlds wouldn't require much departure from actuality—say, Jeff changes his mind on a whim and decides to come after all—but *w2*-worlds would require a fair amount of departure from actuality—say, Lars is willing to disappoint his friends, skip their wedding, cancel his trip to the West Coast, and come to a party where Jeff will be. If the *w1*- and *w2*-worlds are this far apart then there are likely *intermediate worlds* in between them where Jeff but not Lars comes to the party, *but it still isn't fun*. For example, suppose Jeff is characteristically gregarious, but had his day gone a bit differently he would've been uncharacteristically ornery, which wouldn't have been fun. Or suppose the party almost ran out of beer, and an additional large group almost showed up who would've drunk up the last of it, which also wouldn't have been fun. If any worlds like these, where Jeff comes to the party without Lars but it isn't fun (*w3-worlds* in Diagram 2), require less departure from actuality than *w2*-worlds, then the ordering will be:

Diagram 2



And if this is the case NOTLARS(b) must be false. For (a) would expand the domain to include *w2*-worlds, thereby also including both *w1*- and *w3*-worlds; and since *w3*-worlds make the antecedent of (b) true but its consequent false—Jeff but not Lars comes to the party, but it isn't fun—(b) would be false. But intuitively, of course, it's true.

It is important to appreciate that it is the structure of the case that matters, not the particular details. This intermediate worlds problem can arise whenever more departure from actuality is required to make the antecedent of an earlier counterfactual true than to make a subsequent counterfactual's antecedent true but its consequent false. And of course there are no constraints on how far apart the various antecedent-possibilities in a counterfactual sequence may be. So once the recipe is clear, cases like the previous one are easy to cook up. Here is an even simpler one. Suppose Tina almost came to the party too, and it would've been fun. Then:

(TINA): (a) If Lars had come to the party, it would've been fun. (b) If Tina had come to the party, it would've been fun.

Footnote 7 continued
of this idea (1973, 50–52), but does not reject the possibility. But we needn't assume such a metric to make good sense of weaker claims like "world *w* is *very* far from @", just as I needn't assume that my preferences are determinate and precise (though they could be!) to believe that I *greatly* prefer chocolate to vanilla. In fact, it is sufficient for my cases here merely that we can make good sense of intuitions of the form: "world *w* involves at least as much departure from @ (of the relevant kind) as world *w*′ does". And indeed, these seem to be the very sorts of judgments that are presumed by proponents of similarity-based accounts to underlie our capacity to evaluate counterfactuals.

But since Tina almost came to the party, but Lars was never going to come, worlds where Tina comes are likely to be much closer than worlds where Lars comes. Then the same sorts of intermediate worlds, e.g. where Tina comes but we run out of beer, may be included in the domain expansion induced by (a), thus falsifying (b).

Analogous variants of SOPHIE are easy to come by too: suppose Sophie lives in Australia, and never seriously considered flying to New York for the parade, but had planned to watch it on TV. The following may well be true:

(ONTV): (a) If Sophie had gone to the parade, she would've seen Jeter. (b) If she had watched it on TV, she would've seen him too.

But if worlds where she watches on TV are much closer than ones where she goes in person, and it doesn't require much additional departure from actuality to find worlds where she watches on TV and somehow misses Jeter—e.g. if her TV reception cuts out—then these intermediate worlds will be included in the expansion induced by (a), thus falsifying (b). So intuitively such worlds ought not to be included in expansions; coarse-grained expansion is too permissive.

3.2 Falsifying antecedents

At this point the following line of thought is natural: the intermediate worlds problem was a result of coarse-grained expansion, which includes not just the worlds inducing the expansion (the closest A-worlds), but also any other world at least as close, some of which falsify subsequent conditionals in unexpected ways. So if we could screen off these problematic intermediate worlds in a principled way, perhaps we could avoid this problem.

Whether or not such worlds can in fact be successfully screened off will be discussed below when considering possible responses on behalf of SCA. But even if they can be, this solution will do nothing to avoid the remaining cases. In particular, this next class of cases involves sequences in which even the first conditionals' antecedent-worlds falsify the second conditionals, and these are the very worlds the relevant expansions are meant to include. Suppose Lars wanted to go to the beach instead of the wedding. Then this might well be true:

(BEACH): (a) If Lars had come to the party, it would've been fun. (b) If he hadn't been at a wedding on the West Coast that day, he would've gone to the beach.

Since Lars is great fun at parties, (a) is presumably true. And we can easily imagine that (b) is true. But the closest worlds where the antecedent of (a) is true, i.e. where Lars is at the party in New York, are worlds that make (b) false, i.e. where it's true that he isn't at the wedding on the West Coast, but it's false that he goes to the beach (assuming he can only do one of the three). So even an expansion including *only* the closest antecedent-worlds for (a) would make the antecedent of (b) true but its consequent false.

An analogous version of the SOPHIE/ONTV case is easy to imagine:

(COUCH): (a) If Sophie had gone to the parade, she would've seen Jeter. (b) If she hadn't fallen asleep on her couch just before it began, she would've watched it on TV.

Given that Sophie lives in Australia, worlds where she goes to the parade in New York are worlds where it's true that she doesn't fall asleep on her couch just before it begins, but where it's false that she watches it on TV.

In the previous section we observed that coarse-grained expansion would lead to problems by admitting falsifying intermediate worlds into the domain. But even the most fine-grained expansion possible, adding to the domain only the antecedent-worlds for the conditional inducing the expansion, would lead to problems as well.

3.3 Complex evolution

At this point another line of thought is natural: perhaps the original version of SCA, according to which domain expansions by default endure, is too simple. Perhaps what these cases show is that domains are often reset or contracted. So perhaps if we could explain when expansions endure and when they do not, we could avoid this problem.

Whether or not there is an adequate solution to the previous cases in this vicinity will also be discussed in the Sect. 4 below. But even if a mechanism for domain contraction or resetting is built into the formal framework, with a corresponding reliable method for predicting when expansions endure, the semantics will not yet be sophisticated enough to deal with the following cases. Suppose the problematic cases above, such as BEACH and COUCH, really did induce some sort of domain contraction or resetting. Then we would expect subsequent domain expansions to continue to occur in the usual way: when there are no A-worlds in the domain, the modal horizon is broadened just enough to include the closest ones, and the counterfactual in this context should have the ordinary truth conditions. But this cannot be right. This would mean that after a putative contraction we would generally *not* find the same sorts of infelicities observed in the original reverse Sobel sequences. For what causes these infelicities, according to SCA, is when an enduring earlier expansion causes a later conditional to be evaluated at the previously expanded domain rather than the domain at which it would ordinarily be evaluated. So in cases where such an expansion is reversed, we should generally not expect it to be able to generate the same sort of infelicity in later conditionals.

But in fact we find the opposite. Consider a variant of BEACH:

(BEACH2): (a) If Jeff and Lars had come to the party, it wouldn't have been fun. (b) If Lars hadn't been at a wedding on the West Coast that day, he would've gone to the beach. (c) #If Lars had come to the party, it would've been fun.

The tension between (a) and (c) of this sequence is exactly analogous to the one between (a) and (b) in the original PARTY-R,[8] so the explanation ought to be

[8] The meticulous reader will have noticed that BEACH2(c) is about Lars coming to the party, rather than Jeff. The example is simpler this way. We can suppose that if Lars had come Jeff would still not have

essentially the same. This means the expansion induced by BEACH2(a), after being reversed for the evaluation of (b) (as is the current proposal), would then have to be *reinstated* for the evaluation of (c), rather than (c) inducing its own expansion in the usual way. For if (c) induced its own ordinary expansion, the domain at that point would include the closest worlds where Lars comes to the party, at which Jeff does not come and it is fun, but not the worlds where they both come and fight. So we would evaluate (c) in the usual way, at the domain including only the closest A-worlds, and should judge it to be true. But it is in fact infelicitous. So something additional must be added to the semantics to reinstate the earlier domain in this case, even though it is not necessary for the accommodation of the new antecedent.

An exactly analogous variant of the Sophie case is just as easy to construct:

(COUCH2): (a) If Sophie had gone to the parade and gotten stuck in the back, she wouldn't have seen Jeter. (b) If she hadn't fallen asleep on her couch just before it began, she would've watched it on TV. (c) #If she had gone to the parade, she would've seen Jeter.

Since worlds where Sophie is at the parade in New York are *a fortiori* worlds where she does not fall asleep on her couch in Australia just before it begins, but where she also doesn't watch it on TV, these worlds would falsify (b). So the domain expansion that includes them, induced by (a), must be reversed for the evaluation of (b). But then if (c) induced another expansion in the normal way the domain would expand just enough to include the closest worlds where Sophie goes to the parade, at which she sees Jeter, so (c) should be true. But it is in fact infelicitous. So, again, the earlier domain must be reinstated for some reason, even though this is not necessary for the accommodation of the new antecedent.

These cases show that even if some mechanism for domain contraction is built into the account that can distinguish between the cases that would require such contractions and the cases that would require, rather, a persistence of the previous expansion, SCA will still fail to predict these later-occurring infelicities. Still more complexity would need to be added to the view.

3.4 Independent antecedents

Finally, even if all the previous cases could be accounted for via additional complexity in the semantics/dynamics—and that is a big *if*!—there are cases that apparently generate the same infelicity, but which seem to elude any analysis of the kind SCA has to offer:

(SASHA): (a) If Sophie's twin sister Sasha, who is just like her in almost every way, went to the parade and got stuck in the back, she wouldn't see Jeter. (b) #If Sophie went to the parade she would see Jeter.

Footnote 8 continued
come, so that the closest worlds where Lars comes are ones where the party is fun. This is compatible with the claim that *if* they both had come, it wouldn't have been fun.

(GROUP): (a) If Jeff went to the parade and got stuck in the back, he wouldn't see Jeter. (b) If Lars went to the parade and got stuck in the back, he wouldn't see Jeter. (c) If Tina went to the parade and got stuck in the back, she wouldn't see Jeter. (d) If you or I went to the parade and got stuck in the back, we wouldn't see Jeter. (e) #If Sophie went to the parade she would see Jeter.

In these cases, unlike the other cases generating the infelicity, there is no direct logical relationship between the infelicitous sequence-final conditionals and the earlier conditionals in the sequences. In SASHA, the proposition that *Sasha* does not see Jeter is perfectly compatible with the proposition that *Sophie* sees Jeter, so the inclusion in the domain of worlds where the former is true has no truth-conditional bearing on conditionals about the latter. (Similar remarks apply to the second example, of course.)

Furthermore, note that it is immaterial which worlds are closer, the Sasha-worlds or the Sophie-worlds—we don't even need to know anything about who was more likely to go to the parade in order to recognize the infelicity of SASHA(b). And suppose again that Sophie lives in Australia and never seriously considered going to the parade, and suppose as well that Sasha lives a few blocks from the parade and almost went. Then the possibility of Sasha going is presumably much closer than the possibility of Sophie going. So an utterance of (a) should only expand the domain to include the closest Sasha-worlds, which would not include any Sophie-worlds. So a subsequent utterance of (b) should expand the domain again to include the closest Sophie-worlds, which are by hypothesis all ones where she sees Jeter. Whatever the explanation of the infelicity of SASHA(b) is, it seems like it *can't* be the same as SCA's explanation of the infelicity of SOPHIE(b). But intuitively their explanation should be essentially the same. This suggests the original explanation was wrong.

Readers familiar with the relevant literature may have noticed the similarity of this case to one mentioned by Moss (2012, p. 21). She observes that the same infelicity found in reverse Sobel sequences can be generated by certain non-modal claims, like the following (adapted from her case):

(OFTEN): (a) Often when people go to parades they get stuck in the back of the crowd and can't see the thing they came to see. (b) #If Sophie went to the parade, she would see Jeter.

Since a construction like (a) is "not a counterfactual, or even a modal sentence, it does not prompt any expansion of the domain over which counterfactuals quantify" (ibid), and thus does not fall under the purview of the SCA explanation of infelicities of this kind. Moss claims this datum is a point in favor of her account, since she appeals to a more general pragmatic/epistemic principle that plausibly applies to cases like OFTEN as well as sequences of conditionals. (See Sect. 5 below for a more detailed discussion of her view.)

But a reasonable response on behalf of von Fintel might go something like this: "It is a nice *bonus feature* of your account that it also explains analogous infelicities generated by non-counterfactual constructions. But this virtue of your account is not a *defect* in mine, for all I have offered is an account of counterfactuals. For all I have

said, analogous mechanisms exist elsewhere that have similar effects on modal domains to generate analogous infelicities. But these data are not my responsibility." However, SASHA and GROUP *are* sequences of counterfactuals, and whatever putative effects on the modal domains are responsible for the infelicity of the sequence-final conditionals, they are effects generated by the earlier counterfactuals. So apparently these data *do* fall under the purview of von Fintel's account, and an explanation of them *is* his responsibility, not just a nice bonus. In principle it is possible that there is an additional factor, outside the scope of an analysis of counterfactuals, interacting with the ordinary mechanisms at play in counterfactual discourse that replicates the same or a similar effect. But there is more pressure on the strict conditional account to be able to explain data of the kind it purports to be *an account of* than there is to explain data of another kind that exhibits similar behavior. And given the structure of the framework, it isn't clear at all how such a story could go.

3.4.1 Recap

Let's summarize the findings of the current section before moving on to possible solutions. The *intermediate worlds cases* showed that coarse-grained expansion would in many cases allow worlds into the domain that would falsify subsequent conditionals that were intuitively felicitous. The *falsifying antecedents cases* showed that even the most fine-grained method of expansion, including only the antecedent-worlds of the conditional invoking the expansion, will deliver similar incorrect predictions about other sequences that are intuitively felicitous. The *complex evolution cases* showed that even if domain expansions were reversed for some reason in the previous cases, the account would fail to predict later infelicities generated by conditionals occurring after these putative reversals. And the *independent antecedents cases* showed that even if all the previous cases could be accounted for, there are other counterfactual sequences apparently generating the same infelicity that seem to elude explanation of the kind offered by the strict conditional theory.

4 Responses

In this section I consider possible responses to my problem cases on behalf of the strict conditional account. I find none of them to be fully satisfying.

4.1 Antecedent-world-only expansion

One option that can be quickly ruled out is the idea of trading in coarse-grained expansion for *maximally fine-grained* expansion, i.e. expansion that includes *only* the closest antecedent-worlds for the conditional inducing the expansion. In the case of PARTY-R, for example, an A-world-only expansion induced by (a) would include only the closest worlds where Jeff and Lars both come to the party. The new

domain would thus exclude the problematic intermediate worlds that threatened to falsify (b), where Jeff but not Lars comes to the party, but, say, Jeff is unusually ornery so it isn't fun after all. This wouldn't address the other problem cases, but it would avoid the intermediate worlds cases, which would be a good first step.

This option is a non-starter, however, for it validates a disastrous logic of counterfactuals, at least in contexts with sufficiently small initial domains. Here are three examples:

Antecedent-to-consequent movement (ACM):
$D \cap \mathbf{A} = \emptyset$
$(A\&B) > C$
$\therefore A > B$

Antecedent conjunct elimination (ACE):
$D \cap \mathbf{A} = \emptyset$
$(A\&B) > C$
$\therefore A > C$

Antecedent entailment substitution (AES):
$D \cap \mathbf{B} = \emptyset$
$A > C$
$A \vDash B$
$\therefore B > C$

ACM and ACE go hand-in-hand. Whenever the domain D contains no A-worlds (i.e. $D \cap \mathbf{A} = \emptyset$), an utterance of $(A\&B) > C$ will expand D to include the closest A&B-worlds. These will now be the *only* A-worlds in D, since there were previously none. And of course they are all B-worlds, since they are all A&B-worlds. They must also all be C-worlds, since $(A\&B) > C$ is true. So all A-worlds in D will be both B-worlds and C-worlds. So $A > B$ will be true (ACM), and $A > C$ will be true (ACE).

Apply this to the case of PARTY-R: suppose there are no worlds in D where Jeff comes to the party. An utterance of (a) will expand D to add the closest worlds where Jeff and Lars come to the party. Then these will be the only worlds in D where Jeff comes to the party, so it will be true that if Jeff came to the party, then Lars would come to the party. And since all these worlds are worlds where the party isn't fun, it will be true that if Jeff came to the party it wouldn't be fun. But if the closest worlds where Jeff comes to the party are ones where Lars doesn't come, and it's fun, these should both be false.

An even more extreme case will underscore the implausibility of ACM and ACE. Suppose there are no worlds in D where I go to the zoo. Then an utterance of "If I went to the zoo and got mauled by a lion, I would be traumatized" will add to D the closest worlds where I go to the zoo and get mauled by a lion. Then these will be the only worlds in D where I go to the zoo, so an utterance of "If I went to the zoo, I would get mauled by a lion" will be true. And if all these worlds are worlds where I am traumatized, an utterance of "If I went to the zoo, I would be traumatized" will be true as well. But these (one hopes) are both false.

AES is similar. Suppose A entails B. An utterance of $A > C$ will expand D to include the closest A-worlds, which will also be B-worlds due to the entailment. If

there were previously no B-worlds in *D*, these will now be the only B-worlds in *D*. If A > C is true, these will all be C-worlds, so B > C will be true.

Apply this to a variant of the previous case: supposing *I go to the zoo* entails *I leave the house*, if there are no worlds in *D* where I leave the house, then an utterance of "If I went to the zoo, I would see an elephant" will expand *D* to include the closest worlds where I go to the zoo, which will be the only worlds in *D* where I leave the house; if I see an elephant at these worlds, then I see an elephant at every world in *D* where I leave the house; so an utterance of "If I left the house, I would see an elephant" will be true. But this is obviously false.

These logical consequences are devastating. In addition, this strategy would only avoid the intermediate worlds cases. So this option can be ruled out conclusively.

4.2 Medium-grained expansion

Coarse-grained expansion made SCA vulnerable to the intermediate worlds problem, and maximally fine-grained expansion validated unacceptable logical principles. But perhaps something in between would get things right: perhaps if domains expanded by adding more than just the closest A-worlds, but less than every other world at least as close, maybe they would add enough worlds to avoid validating the unacceptable logical principles, but not enough to include the problematic intermediate worlds.

There are too many possible versions of this approach to rule it out with the same confidence as the previous one, but there is reason for pessimism. Let's work backwards from the problem to the potential solution. In the intermediate worlds cases, the problematic worlds were ones that falsified the second conditional in a given sequence by departing from actuality enough to make its antecedent true *and then some*, making the consequent false via some additional departure. Recall that in NOTLARS the problematic intermediate worlds falsifying (b) (*If Jeff but not Lars had come to the party, it would've been fun*) were ones where Jeff comes to the party and Lars doesn't, but something else makes the party unfun, e.g. Jeff is uncharacteristically ornery or another group shows up and drinks all the beer. It is natural to feel that these worlds ought to be irrelevant, since the extra features of them that falsify the conditional—Jeff being ornery, the extra group showing up—were not crucial to bringing about the truth of the antecedent (*Jeff but not Lars comes to the party*). They seem to depart in irrelevant ways, reminding one of what Lewis insisted the closeness ordering ought to avoid: *gratuitous departure from actuality*. A sensible thought, then, is: What if domain expansions systematically excluded worlds that differ in ways not relevant to the antecedent?[9]

[9] A nearly equivalent variant of this view, perhaps closer in spirit to Lewis's account, would be to count worlds that differ in ways not relevant to A as farther than the closest A-worlds. On this view, at least formally, coarse-grained expansion is preserved, but the problematic intermediate worlds are counted as too far in the relevant ordering to fall within the modal horizon. Ignoring some technical differences, this would deliver more or less the same results as the proposal currently under consideration, since the same worlds would be included in or excluded from the domain, albeit in virtue of being too distant, rather than being simply irrelevant. As a result my objections to the current proposal would apply, *mutatis mutandis*, to this variant as well.

The trouble is, the problematic intermediate worlds are irrelevant to the evaluation of (b), but the domain expansion that admits them is induced by (a), the previous conditional, *with a different antecedent*. So for this solution to work, they would have to be excluded from the expansion on the basis of irrelevance to (a). But of course there is no guarantee that the same things will be relevant to the antecedents of both (a) and (b). Consider the intermediate worlds where Jeff is unusually ornery so the party isn't fun. On the current proposal these worlds would be excluded from the expansion induced by (a) on grounds of irrelevance. But suppose the closest worlds where the antecedent of (a)—*Jeff and Lars come to the party*—is true *are* ones where Jeff is unusually ornery. Perhaps he never would've showed up to the same party as Lars in the first place unless he was looking for a fight. Then the departure from actuality involved in making Jeff ornery *is* required to make the antecedent of (a) true, so it *is* relevant after all. In this case the intermediate worlds falsifying (b) would not depart in ways irrelevant to (a), so they would not be excluded from an expansion induced by (a). And in that case they would be in D to falsify (b) after all.

Here is a second problem. Suppose as we just did that the closest worlds where Jeff and Lars both come are ones where Jeff is ornery. And suppose that the closest worlds where Jeff alone comes are ones where he is characteristically gregarious. And suppose finally that in the actual world Jeff was in a lukewarm mood, which is why he skipped the party. In this case, relative to the evaluation of (a), worlds where Jeff is gregarious seem to depart from actuality gratuitously. After all, Jeff's actual mood was lukewarm, and in the closest antecedent-worlds for (a) he is ornery, so any worlds where he is gregarious depart in a way that is irrelevant to (a). So on these grounds an expansion induced by (a) should exclude any worlds where Jeff is gregarious, among which are the closest worlds where Jeff (but not Lars) comes to the party. After such an expansion the only worlds in D where Jeff and not Lars come will be ones where Jeff is not his usual, gregarious self. Supposing this would not be fun, NOTLARS(b) would then be false, and it would be true instead that if Jeff but not Lars came, then Jeff would not be his usual, gregarious self, and it would not be fun.

Finally, note that the closest A-worlds for a conditional like (b) may be the falsifying intermediate worlds for some other true conditional (b′). So, for the reasons just mentioned, we *would* want them to be included by a preceding expansion for the purposes of evaluating (b); but, to avoid the intermediate worlds problem, we *would not* want them to be included in such an expansion for the purposes of evaluating (b′). For example, suppose Jeff was invited to two parties, ours and Sasha's, and if he had been in a better mood he would've gone to a party, but it would've been Sasha's and not ours. Then the following sequence seems felicitous:

(MOOD): (a) If Jeff and Lars had come to the party, it wouldn't have been fun.
(b) If Jeff had been in a better mood that day, he would've gone to Sasha's party.

Note that MOOD(a) is the same conditional as NOTLARS(a), so (assuming a relevantly similar context) they should induce the same domain expansion. And for

NOTLARS(b) to be true, this expansion ought to include the closest worlds where Jeff (but not Lars) is at our party, at which he is in a gregarious mood and the party is fun.[10] But these very worlds are ones in which he is in a better mood (than his actual lukewarm mood), but he is at our party instead of Sasha's. So for MOOD (b) to be true, the expansion ought *not* to include these worlds. But of course the expansion must either include or exclude the worlds where a gregarious Jeff comes to our party; it cannot do both.

It may be premature to rule out the possibility that some other version of the current proposal will be successful, but given these initial difficulties I cannot imagine what it would be.

4.3 Sophisticated domains

My first two classes of problem cases—the intermediate worlds cases and the falsifying antecedents cases—involve sequences of counterfactuals in which (according to SCA) the first conditional would induce an expansion that (according to me) would inadvertently include worlds that would falsify the second conditional when evaluated at this expanded domain. Alternate methods of expansion that would exclude some of these worlds led to undesirable results. So perhaps the proper solution is that expansions turn out to be more fragile, i.e. easily reversed in some way, than the simple version of SCA allows, and so the second conditionals in these cases are not in fact evaluated at the expanded domains.

A few questions are immediately raised by this proposal, two pertaining to the technical details, and a third concerning a related explanatory burden. The first technical question is: When an expansion is reversed, what is the resulting state of the modal domain? It might return to its most recent previous state, or some other earlier state, or a null state, or the singleton set containing the actual world, or something else. Different answers to this question will make different predictions, but I will set this issue aside entirely here.

Second, how are the complex evolution cases to be accounted for within the semantic framework? A reversed expansion cannot simply be "erased", because the same infelicity can be generated after a conditional that purportedly induces such a reversal. Recall:

(COUCH2) (a) If Sophie had gone to the parade and gotten stuck in the back, she wouldn't have seen Jeter. (b) If she hadn't fallen asleep on her couch just before it began, she would've watched it on TV. (c) #If she had gone to the parade, she would've seen Jeter.

According to SCA, first (a) induces an expansion to include the closest worlds where Sophie goes to the parade and gets stuck in the back. Call this domain D_1. Next, on the current proposal, (b) induces a domain contraction of some kind to exclude these worlds and any intermediate worlds that falsify (b). Either this

[10] We would also want such worlds, where Jeff but not Lars comes to our party, to be included in order to avoid the entailment from "If Jeff and Lars had come to the party, it wouldn't have been fun" to "#If Jeff had come to our party, Lars would've come as well" (an instance of the undesirable ACM inference from the previous section).

contraction first reduces the domain to some minimal state (i.e. just the actual world, or no worlds at all), after which another expansion occurs to include the closest worlds where Sophie doesn't fall asleep on her couch before the parade, or the domain simply contracts directly to this point. Call this domain D_2. Since the closest worlds where Sophie doesn't fall asleep on the couch are ones where she watches the parade on TV, and these are closer than any worlds where she goes to the parade in person, D_2 will include no worlds where she goes in person. So (c) would then induce another expansion, this time to include the closest worlds where Sophie goes in person, at which she *doesn't* get stuck in the back, and does see Jeter. Call this domain D_3. So (c) should be true, since D_3 includes none of the worlds where she goes to the parade and misses Jeter. But it is in fact infelicitous.

So the mechanics of domain expansion must be more complex. The explanation of the infelicity of (c) ought to be essentially the same as that of the original SOPHIE-R(b), since COUCH2 minus its (b) conditional is effectively identical to SOPHIE-R.[11] So COUCH2(c) must expand the domain not to D_3, as predicted by the simple mechanics, but rather to D_1, the domain including worlds where Sophie goes to the parade and gets stuck in the back, which was previously associated with the (a) conditional. But why should it do that? The antecedent of (c) is *Sophie goes to the parade*, not *Sophie goes to the parade and gets stuck in the back*. Recall that, according to SCA, expansions occur simply to accommodate new possibilities being introduced, so that speakers have a chance of saying something non-trivial. This was a purely mechanistic explanation in terms of an independently plausible conversational process. But in the present case the normal method of expansion must be circumvented and replaced by something more complex. This cries out for explanation.

So the second task for the current proposal is to provide a more sophisticated mechanics for domain expansion to account for cases like COUCH2, which require reinstating domains from previous points in the discourse. I will not fuss over the formal details here, but I suspect the best option is to build into the account some sort of running record of previous domains and the conditionals associated with them, and a relation R between conditionals (or antecedents, perhaps) that holds whenever one conditional is to be evaluated at the domain associated with another earlier conditional. (The R-relation, of course, is a black box in the theory that will have to be replaced by a more detailed account of what prompts the reinstatement of an earlier domain. But it will help in the meantime to have a sort of placeholder for the missing part of the theory.) There is nothing terribly wrong with this additional complexity per se, though the account will have suffered some loss of theoretical simplicity and elegance vis-à-vis the original, simpler SCA.

More important than the technical apparatus, however—and this brings us to the third and most pressing question for this proposal—will be the choice of what plays the role of the R-relation in the theory. Of course it is not enough simply to say that

[11] The only difference between them is that COUCH2 is a past tense counterfactual while SOPHIE-R is present tense—"if Sophie *had gone* to the parade, she *would've* seen Jeter", as opposed to "if Sophie *went* to the parade, she *would* see Jeter". This was done to maintain tense agreement between COUCH2 and COUCH, which made more sense as a past tense counterfactual with an assumption that Sophie in actuality did fall asleep on the couch.

some or other relation exists that unifies the cases generating the infelicity. The account of *R* will have to not only make the right predictions about which particular cases call for reinstatements, but also provide some non-ad hoc explanation of why such reinstatement does or does not occur. This predictive and explanatory power will be the measure of success for this proposal.

It is important to appreciate the significance of this explanatory challenge. Von Fintel's original view provides an account of how modal domains behave throughout a discourse, which makes certain desired predictions about a particular class of data. The mechanics of domain evolution alone provided an explanation of such behavior. These predictions and the corresponding explanation provided the primary motivation for the theory. But my examples show that the theory overgenerates and makes undesirable predictions about a variety of other, similar data. So the scope of data actually *accounted for* by the account, as it stands, turns out to be narrower than expected. To be complete the account will have to be amended to cover the new data as well; to do this it will have to build domain contraction into the mechanics. But now with both expansion and contraction in the mechanics, the picture is more complicated: we had a story about when expansions occur, but that story turned out to be insufficient; we now need a story about when contractions occur and why, in order to restore the theory's scope of explanation to full generality.

What might such a story look like? It will help to take inventory of the data in need of explanation. Let's revisit a sample of the earlier cases for comparison, reprinted here for convenience:

(SOPHIE-R): (a) If Sophie went to the parade and got stuck in the back of the crowd, she wouldn't see Jeter. (b) #If Sophie went to the parade, she would see Jeter.

(ONTV): (a) If Sophie had gone to the parade, she would've seen Jeter. (b) If she had watched it on TV, she would've seen him too.

(COUCH): (a) If Sophie had gone to the parade, she would've seen Jeter. (b) If she hadn't fallen asleep on her couch just before it began, she would've watched it on TV.

(COUCH2): (a) If Sophie had gone to the parade and gotten stuck in the back, she wouldn't have seen Jeter. (b) If she hadn't fallen asleep on her couch just before it began, she would've watched it on TV. (c) #If she had gone to the parade, she would've seen Jeter.

(SASHA): (a) If Sophie's twin sister Sasha, who is just like her in almost every way, went to the parade and got stuck in the back, she wouldn't see Jeter. (b) #If Sophie went to the parade she would see Jeter.

In the classic reverse Sobel sequences like SOPHIE-R, the first conditional expands *D* to include the closest A-worlds, and these worlds falsify the second conditional when it is evaluated at *D*. So the aforementioned *R*-relation between conditionals or antecedents of course holds in these cases. In the intermediate worlds and falsifying antecedents cases like ONTV and COUCH, the same sort of expansion occurs, so according to the simple version of SCA the second conditional should be falsified, either by the problematic intermediate worlds or again by the

first conditional's A-worlds, respectively. But in these cases the second conditional is intuitively true, so the putative expansion must be reversed, so *R* must *not* hold. In the complex evolution cases like COUCH2, the third conditional is infelicitous, and it is the first conditional that is responsible for its infelicity, so *R* must hold between these. But the second conditional is perfectly felicitous, so here as well the putative expansion induced by the first conditional must be reversed, so *R* must *not* hold between the second conditional and the first. Since the first conditional must be able to produce the infelicity of the third even after the reversal occurs, the *R*-relation must hold across domain contractions, so to speak, and in these cases an earlier state of *D* must be reinstated. And in the independent antecedents cases like SASHA, the closest A-worlds for the first conditional do not falsify the second conditional, nor necessarily do any of the other worlds included in the expansion, yet the second conditional is infelicitous in a way that is analogous to the other cases. And in this case there is no previous state of *D*, e.g. one including worlds where Sophie goes to the parade and does not see Jeter, that could be reinstated to falsify the second conditional. So here the infelicity is completely mysterious.

So what special relationship obtains between the relevant conditionals or antecedents in SOPHIE-R, COUCH2, and SASHA, but not in ONTV or COUCH? Isolate the first two sequences reprinted above (SOPHIE-R and ONTV), and one might think the relation was entailment-based: *Sophie goes to the parade and gets stuck in the back of the crowd* entails *Sophie goes to the parade*, but *Sophie goes to the parade* does not entail *Sophie watches the parade on TV*. But the remaining sequences defeat this hypothesis: *Sophie goes to the parade* and *Sophie goes to the parade and gets stuck in the back* do entail *Sophie doesn't fall asleep on her couch just before the parade begins*—not logically, of course, but in some significant sense of entailment—but these pairs are not infelicitous. And *Sophie's twin sister Sasha… goes to the parade* does not in any sense entail *Sophie goes to the parade*, but this sequence is infelicitous.

Likewise, at a glance it might appear to be some syntactic/structural relationship between the conditionals in the sequences that is relevant. In the infelicitous SOPHIE-R and (the infelicitous portion of) COUCH2, *Sophie goes to the parade and gets stuck in the back of the crowd* contains as a syntactic constituent *Sophie goes to the parade*, but the same cannot be said of the conditionals in the felicitous ONTV, COUCH, or (the felicitous portion of) COUCH2. But neither can the same be said of the infelicitous SASHA. Moreover, SOPHIE-R could easily be rephrased to eliminate this structural relationship between the antecedents:

(SOPHIE-R′): (a) If Sophie went to the parade and got stuck in the back of the crowd, she wouldn't see Jeter. (b) #If the group of parade attendees included Sophie, she would see Jeter.

(SOPHIE-R″): (a) If Sophie went to the parade and got stuck in the back of the crowd, she wouldn't see Jeter. (b) #If the parade were attended by Sophie, she would see Jeter.

The infelicity survives. Clearly it is not explained by syntactic structure.

Neither a logical nor syntactic relationship is the *R* we are looking for. What more subtle feature might explain the relevant patterns? Admittedly there are too

many possibilities in logical space to take exhaustive inventory of the options. But one more possibility worth considering—partly because it has arisen during informal discussion of my data—is the notion of *topic*. Suppose the above cases where I have claimed SCA incorrectly predicts infelicity in fact involved mid-sequence topic changes, and suppose this were the sort of thing that could somehow change the operative modal domain. Then even when the first conditional in a sequence expanded the domain to include some problematic worlds, if the second conditional induced a topic change then it would not be evaluated at this earlier domain, and the infelicity would not be predicted. Domains could not be simply "reset", of course, as we learned from the complex evolution cases. But perhaps if each topic were associated with its own dynamic domain evolving in the manner initially proposed by von Fintel, then *unification under a single topic* could play the role of the *R*-relation.

But it is doubtful any plausible notion of topic will be up to the task. First we can rule out the *commonsense* notion of topic. Most of my problem cases intuitively involve no change of topic in the ordinary sense, though verification of this claim is best left to the reader. It is corroborated, however, by the bizarreness (signaled by **?** below), if not quite infelicity, of inserting an explicit signal of topic change mid-sequence (as one often does to smooth over an abrupt change of topic):

(ONTV?): (a) If Sophie had gone to the parade, she would've seen Jeter. (b) **?**On a different topic, if she had watched it on TV, she would've seen him too.

(COUCH?): (a) If Sophie had gone to the parade, she would've seen Jeter. (b) **?**On a different topic, if she hadn't fallen asleep on her couch just before it began, she would've watched it on TV.

Additionally, there are a few theoretical notions of topic established within the relevant linguistics literature,[12] but they are unlikely to do the job either. *Grammatical* notions of topic are largely grounded in sentence structure (at least in English) and anaphora resolution (see, e.g. Roberts 2011; Cornish 2006). But both conditionals in ONTV, for instance, have the same general structure with the same grammatical subject (viz. Sophie, modulo the substitution of "she" for "Sophie" in (b)). And it was demonstrated a moment ago that the grammatical structure of the infelicitous cases could be rearranged without change to the infelicity of the sequence. Moreover, all the anaphora in ONTV(b) are bound by constituents of (a). Together these facts strongly suggest that the transition from (a) to (b) does not even constitute a change of *sentence* topic, let alone a change of *discourse* topic (which would be the more appropriate candidate for the *R*-relation).

The other notable linguistic conception of topic identifies or associates topics with *Questions Under Discussion* (QUDs)—implicit background questions that guide discourse along various paths of inquiry, thus organizing conversational contributions correspondingly (see, e.g. Roberts 2012). For example, consider the sequence:

[12] Thanks to an anonymous referee for this journal for pointing this out.

(DRINKS): Jack had the beer; Una had the wine; I had the whiskey.

According to QUD theory, these statements form a coherent narrative because they answer a cluster of related questions, such as: *What did Jack have?*; *What did Una have?*; and *What did [the speaker] have?*, or *Who had the beer?*; *Who had the wine?*; and *Who had the whiskey?*; as well as the more general question: *Who had what?* But it is unlikely the concept of QUDs tracks the patterns of in/felicity observed in our data. For example, the most natural candidates for QUDs unifying the conditionals in the original PARTY under a single topic would be: *What if Jeff came to the party?* or *If Jeff came to the party, what would it be like?* But it's hard to imagine why these QUDs should not also unite the conditionals in NOTLARS in precisely the same way.

Finally, and more generally, it is unlikely that any notion of topic *whatsoever* will be adequate to completely avoid intermediate worlds cases, given their structure. Recall that if the first conditional's A-worlds are significantly farther from actuality than the second conditional's A-worlds, there will likely be some worlds in between that falsify the second conditional. All that is needed for this situation to arise is that some scenario making the second conditional's antecedent true but its consequent false should require less departure from actuality than the antecedent of the first conditional. But this relationship is a matter of comparative similarity, which has nothing inherently to do with topic, and it is extremely unlikely that these two disparate notions will turn out to be systematically coordinated in a way that could provide the needed theoretical resources. More specifically, it is unlikely that any notion of topic worthy of the name will guarantee, for any two possibilities falling under the same topic, that the closest worlds realizing each will always fall within a narrow enough threshold of closeness to each other so that no problematic intermediate worlds exist between them.

So whatever special relation holds between the pairs of conditionals or antecedents that generate the infelicity, it does not seem to be logical, structural, or topic-based. I believe that no other candidate exists that will prove to be suitable, though I cannot claim to have decisively established this here, only that these few initially plausible ones are not; further work would be needed to establish the more general negative claim. We are therefore in the suboptimal position of having gone as far as identifying the work that would need to be done to save the theory—identifying a satisfactory *R*-relation—without being able to go so far as to confidently conclude that the work cannot be done. Should we throw up our hands and call it a draw? I think not. Pending a satisfactory answer to this challenge for the strict conditional view, an abductive case could be made against it via a compelling case for a competing account of what unites the infelicitous sequences. So it will be helpful to end my discussion with a brief exploration of the major alternative solution on offer, due primarily to Sarah Moss. I will stop short of actually endorsing Moss's view, but I think it is more promising than the strict conditional view, and thus worth mentioning given the state of the dialectic.

5 Moss's alternative solution

There is an alternative style of explanation of the asymmetry of Sobel sequences, due originally to Moss (2012), which is worth discussing for comparison. According to Moss, the infelicitous reverse sequences can be explained without abandoning the traditional variably strict semantics, by appeal to an independently plausible pragmatic/epistemic principle, viz. (roughly):

(PEP): For possibilities P and Q and speaker S, it is infelicitous for S to assert Q if:
(i) P is salient;
(ii) P and Q are incompatible; and
(iii) S is unable to rule out P.

PEP seems independently plausible due to its applicability to non-conditional cases. Suppose you and I are waiting for a New Jersey Transit train, which is due momentarily. Compare the following sequences:

(TRAIN): Me: Our train will be here any minute.
You: New Jersey Transit trains often run very late.
(TRAIN-R): You: New Jersey Transit trains often run very late.
Me: #Our train will be here any minute.

In the first case, when the possibility of the train being very late has not yet been mentioned, it may be perfectly felicitous for me to outright assert that our train will arrive any minute. But in the second case, once the possibility of lateness is salient, the very same assertion is infelicitous. Moss's principle, together with some reasonable assumptions about salience, seems to explain this: when our attention is focused on our train, mentioning the fact that NJ Transit trains are frequently very late naturally raises to salience the possibility that *our* NJ Transit train is very late. (Let this be P.) Our train being here any minute (let this be Q) is incompatible with its being very late. Unless I have some way of ruling out P—say, if I can see our train down the track—it is infelicitous to assert Q. But when P is not salient, Q may be perfectly assertible.[13]

What does this have to do with counterfactuals? According to Moss, analogous explanations apply to the infelicitous reverse Sobel sequences. In SOPHIE-R, for example, (a) makes salient the possibility that if Sophie went to the parade she might get stuck in the back and not see Jeter, which is incompatible with—or at least clashes with[14]—the claim that if she went to the parade she *would* see him.[15]

[13] Note that this makes no commitments about *truth* or *falsity*. If we have not considered the possibility that our train is late, but it is, it may be perfectly *assertible* that the train will arrive any minute, even if it is false.

[14] It is controversial whether so-called *might* counterfactuals—"If A were the case, C *might* be the case"—are *incompatible* with their consequent-negated ordinary (*would*) counterparts—"If A were the case, C *would not* be the case"—or whether they merely clash in some other way. (See especially Stalnaker 1981.) In either case, we would expect infelicity to result from co-assertion.

[15] The positive accounts of counterfactuals defended by Ichikawa (2011), Lewis (2016, 2017), and Nichols (ms) offer solutions to the problem of reverse Sobel sequences that bear similarities to the Moss-style explanation. An extensive comparison of the solutions on the table is a topic for future research, but

Moss's account needs to be developed further, but it is worth pointing out that the cases discussed in this paper do not appear to be problematic for her style of explanation. First, recall the intermediate worlds cases: in these cases the expansions induced by the first conditionals unexpectedly included worlds that falsified the second conditionals. But Moss's explanation appeals to no such expansions, so these intermediate worlds would not be quantified over by the second conditionals. And the corresponding possibilities are not mentioned, either, so they would not be salient. So this problem never arises at all for a Moss-style explanation.

Neither do the falsifying antecedents cases: here the closest antecedent-worlds for the first conditionals themselves falsify the second conditionals, so including these worlds in the domain incorrectly predicts infelicity. In BEACH, e.g. worlds where Jeff comes to the party are *a fortiori* worlds where he isn't at a wedding thousands of miles away, but he also isn't at the beach, so (b) should be false. But, again, Moss does not appeal to domain expansions, so these antecedent-worlds needn't be relevant to subsequent conditionals. And even if BEACH(a) makes salient some possibilities about what would or might have happened if Jeff had come to the party, this has no immediate bearing on our judgments about what he would've done if he weren't at the wedding.

So Moss's view outperforms von Fintel's view on the first two classes of cases simply by not predicting infelicity where there is none. More impressive, however, is that her view correctly predicts infelicity in the complex evolution cases. Salience can endure throughout an extended stretch of conversation, even if the salient item is not constantly attended to. Consider the following variation of TRAIN-R:

AMTRAK: You: New Jersey Transit trains often run very late. Amtrak trains, however, are extremely punctual, and very clean. Metro-North trains are quite clean too.
Me: #Our (NJ Transit) train will be here any minute.

Clearly the salience of the possibility of lateness is able to survive a brief interlude, since my assertion is still infelicitous. The same observation applies to complex evolution cases like COUCH2. If (a) makes salient the possibility that if Sophie went to the parade she might get stuck in the back, there is no reason to expect this salience to vanish as soon as a slightly different possibility is mentioned. So (c) should be expected to clash with (a) even after the brief digression created by (b).

Moss's view even seems likely to predict the infelicity of independent antecedents cases like SASHA and GROUP. In the train case, making salient the general possibility of NJ Transit trains being late thereby made salient the particular possibility of our NJ Transit train being late. Similarly, making one possibility P salient often thereby makes some similar possibility P′ salient as well:

Footnote 15 continued
it is worth mentioning that there are several proposed variations of the style of explanation discussed in this section.

(HANDS): You: For all I know I'm a brain in a vat! I may not be a real person! I may not even have a body!
Me: #Thank god *I* have a body.

Since your speech makes salient the possibility that you are a bodiless brain in a vat, it automatically makes salient the possibility that I am one as well, even though you have said nothing about me. Likewise, in SASHA, since Sasha and Sophie are so similar, the salience of the possibility that Sasha might get stuck in the back if she went to the parade plausibly brings about the salience of the corresponding possibility involving Sophie. And in GROUP, the salience of several different people possibly getting stuck in the back of the parade plausibly makes salient the general possibility of one getting stuck in the back of the parade, which naturally brings about the salience of the same possibility involving Sophie. Since this is a possibility in which she wouldn't see Jeter, it is infelicitous in these cases to assert that if she went to the parade she *would* see him. Moss's view might have to be developed more to accurately predict these sorts of salience relations, but there is clear independent motivation for their existence, and, with any luck, an account of them could be largely inherited from an adequate account of salience.

Finally, it is important to fully appreciate the difference between the two styles of explanation on offer. According to Moss, counterfactual assertions (and other types) can raise to salience certain related possibilities that interfere with the assertibility of subsequent counterfactual claims that would otherwise be assertible. These relationships between assertions, possibilities, and salience are largely independent of any ordering on worlds. But von Fintel's explanation is quite different: counterfactual assertions can expand the domain of modal quantification outward (in the direction of greater world-distance); and this may include worlds that interfere with the interpretation of subsequent counterfactuals about nearer possibilities, since these may be evaluated at the expanded domains including not just their *closest* A-worlds (i.e. the only ones relevant on the standard truth conditions), but others as well. The infelicity under discussion, by design, thus occurs only at the moments when we turn our attention from farther away possibilities towards closer ones. And, as a corollary, the standard truth conditions for counterfactuals are replicated only when sequences of counterfactual discourse are arranged in an outwardly progressing order. But why should they be so arranged? In ordinary discourse we are not generally concerned with ordering the series of possibilities under discussion according to their distance from actuality. The in/felicity patterns displayed by sequences of counterfactuals seem to be tracking something else.

So Moss's view seems fairly well equipped to handle my problem cases, insofar as they arise for her view at all. This ought not to be taken as an all-things-considered verdict in favor of her account over von Fintel's: my considerations could conceivably be outweighed by other problems for her view and/or virtues of his that I have not discussed. And there are other views in the vicinity of Moss's that may fare better in the long run. But with respect to the cases discussed within this paper, Moss's explanation is at a clear advantage.

6 Conclusion

One of the two primary motivations for the dynamic strict conditional account of counterfactuals is the asymmetry exhibited by reverse Sobel sequences. I have presented four classes of sequences about which this account, as it stands, makes incorrect predictions. I have also considered a few possible extensions of the account that one might have expected to offer some improvement, but which look unpromising upon closer inspection. Though moments of this discussion were inconclusive, we now have at the very least a clearer view of the work that would lie ahead for the strict conditional view. Finally, I have argued that the major competing style of explanation on the table seems to handle these cases fairly easily or avoids them altogether. Further research must be done to more decisively adjudicate between these options. And it bears repeating that the second major motivation for the strict conditional view, concerning so-called negative polarity items, remains completely unaddressed by this paper. But I take myself to have raised a series of objections that at least will need to be addressed by proponents of the strict conditional view, and at most have undermined one of the two arguments for it.

References

Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.

Cornish, F. (2006). Discourse anaphora. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., pp. 631–638). Oxford: Elsevier.

Gillies, A. S. (2007). Counterfactual scorekeeping. *Linguistics and Philosophy, 30*(3), 329–360.

Ichikawa, J. (2011). Quantifiers, knowledge, and counterfactuals. *Philosophy and Phenomenological Research, 82*(2), 287–313.

Ladusaw, W. A. (1979). *Polarity sensitivity as inherent scope relations*. Ph.D. Dissertation, University of Texas, Austin.

Lewis, D. (1973). *Counterfactuals*. London: Blackwell.

Lewis, D. (1986). Counterfactual dependence and time's arrow. Reprinted in *Philosophical papers* (Vol. 2, pp. 32–52). Oxford: Oxford University Press.

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

Lewis, K. (2017). Counterfactual discourse in context. *Noûs*. http://onlinelibrary.wiley.com/doi/10.1111/nous.12194/abstract.

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

Nichols, C. (ms). Rethinking similarity (under review).

Roberts, C. (2011). Topics. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (Vol. 33.2, pp. 1908–1934). Mouton de Gruyter.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics, 6*, 1–69.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). London: Blackwell.

Stalnaker, R. C. (1981). A defense of conditional excluded middle. In W. Harper, R. C. Stalnaker, & G. Pearce (Eds.), *Ifs* (pp. 87–104). Dordrecht: Reidel.

von Fintel, K. (1999). NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics, 16*(2), 97–148.

von Fintel, K. (2001). Counterfactuals in a dynamic context. *Current Studies in Linguistics Series, 36,* 123–152.