<thony@umich.edu>

Counterfactual Scorekeeping

Anthony S. Gillies
University of Michigan

November 30, 2005

1 Loophole

Counterfactuals like

(1) If Sophie had gone to the parade, she would have seen Pedro dance

are supposed—by the lights of orthodoxy at any rate—to mean something very different from what any strict conditional means. Glossing the details of that account a bit, (1) is true at i iff all the *nearest* worlds (to i) in which Sophie is a parade goer are also worlds in which she is a witness to Pedro's dancing. There is no modal operator such that a counterfactual like this amounts to that operator taking widescope over a material conditional with the same atecedent and consequent.

But there is a loophole, mentioned by Lewis just to be dismissed:

> It is still open to say that counterfactuals are vague strict conditionals . . . and that the vagueness is resolved—the strictness is fixed—by very local context: the antecedent itself. That is not altogether wrong, but it is defeatist. It consigns to the wastebasket of contextually resolved vagueness something much more amenable to systematic analysis than most of the rest of the mess in that wastebasket. (Lewis, 1973, p. 13)

Something in the neighborhood of this loophole is right. But it is no mere loophole, and exploiting it is not defeatist.

Very roughly: I will argue that counterfactuals like (1) are, indeed, strict conditionals after all. They amount to a necessity modal, scoped over a material conditional, just which such modal being a function of context. The denotation of the modal is a function of the set s of worlds over which it quantifies, s's value being a function of (among other things) material in the *if*-clause. A bit more precisely: counterfactuals carry an *entertainability presupposition* that their antecedents be possible with respect to the counterfactual domain. A *successful* assertion of a counterfactual in context can change the conversational score, selecting a domain that satisfies the presupposition. It is with respect to the score thus changed that the counterfactual is a strict conditional. If the utterance is unsuccessful—if there is disagreement between the conversational partners about whether the score should be so changed—then there is no accommodation of the entertainability presupposition and the story I want to tell will be idle. That is welcome: my interest is less in the *content* of counterfactuals than it is in getting straight about their *context change potential*—how successful utterances of stretches of counterfactual discourse change the information state of a hearer when she accepts the news conveyed by it.[1]

To see that this kind of story is not the stuff of defeatism we only have to see that the interaction between context and semantic value, mediated by a mechanism of modal accommodation, can be the stuff of formal and systematic analysis. And so we will. To see that this is not a mere loophole, we only have to see that facts about counterfactuals in context—the discourse dynamics surrounding them—are best got at by the kind of story I want to tell. And so we will.

2 Thinning

A hallmark of counterfactuals is that they do not allow for thinning of their antecedents—take a contingently true counterfactual: it is just not so that you can always retain its truth by conjoining an arbitrary bit to its antecedent. Lewis famously used just this fact about counterfactuals to argue that they cannot be strict conditionals. An example:

(2) a. If Sophie had gone to the parade, she would have seen Pedro dance; but of course,

[1] The problems I am interested here have to do with the relationship between counterfactual and context. So I will be ignoring how the semantics of counterfactuals interacts with factual discourse. Put another way: I am interested in problems about how conversation score is affected by counterfactuals and not how that score, once changed, might be *reset* by some bit of factual discourse.

b. if Sophie had gone to the parade and been stuck behind someone tall, she would not have seen Pedro dance.

If counterfactuals could be thinned—their antecedents strengthened—then conjunctions like (2) would be inconsistent.[2] But Sobel sequences like this are not inconsistent. And, of course, it is a fact about strict conditionals that they permit such thinning.

Take any strict conditional $\Box(p \rightarrow q)$, and suppose that the modal is given a very standard semantics relativized to a relevant domain s_c (assigned or inherited from context c):

(3) $[\![\Box\varphi]\!]^{c,i} = 1$ iff $s_c \subseteq [\![\varphi]\!]$

Analyzing (2) as a conjunction of such strict conditionals is a disaster.[3] Assuming (2a) is true, there seems to be no room for (2b) to be true as well. The set of worlds in which Sophie is a parade goer includes the worlds in which she is a parade goer *and* gets stuck behind someone tall—*a fortiori* the set of worlds in s_c in which Sophie is a parade goer includes the worlds in s_c in which she is a parade goer *and* gets stuck behind someone tall. Whence, if the former are included in the set of worlds where she is a witness to Pedro's dancing, so must be the latter. And that—barring our loophole—is pretty bad news for any strict conditional analysis.

But thinning cuts both ways. Although the conjunction in (2) is unremarkable, not so in reverse order:

(4) ??If Sophie had gone to the parade and been stuck behind someone tall, she would not have seen Pedro dance; but of course, if Sophie had gone to the parade, she would have seen Pedro dance.

---

[2] With just a little ingenuity we can produce sequences like this of indefinite length, the conditionals in them oscillating from a true counterfactual to a slightly thinned false one to a slightly more thinned true one to an even slightly more thinned false one . . . . Such sequences are often called "Sobel sequences" (Sobel, 1970).

[3] Three formal conventions. First: officially a context is a pair $c = \langle s_c, i \rangle$, and so the denotation function should be $[\![\cdot]\!]^{\langle s_c, i \rangle, i'}$ where i' is the index of evaluation. But we will ignore embeddings throughout; so let's be a bit sloppy, treating i as both index and part of the context. For the same reasons, we can omit reference to c outright. Second: nevermind that i appears only on the left-hand side of (3): s_c will, in general, be a function of i. Third: assume, for now, that if a strict conditional is true throughout a given relevant domain, then that domain has some antecedent worlds in it.

Far from unremarkable, here we have what sounds for all the world like a contradiction.[4]

We might try to explain away this asymmetry insisting that, despite appearances, (4) is equally unremarkable—we have merely elided the qualification *Sophie's view is unobstructed* in the antecedent of the second conditional. Making this qualification explicit:

(5) If Sophie had gone to the parade and had her view been unobstructed, she would have seen Pedro dance.

Swapping (5) for the second conjunct of (4) would give us a pretty unremarkable sequence of counterfactuals—about as unremarkable as (2).

This diagnosis is wide of the mark. Agreed that such a sequence is unremarkable. Nothing much follows from that. If the qualification is implicit in (4) it must also be implicit in the original Sobel sequence (2). Consider:

(6) a. If Sophie had gone to the parade and had her view been unobstructed, she would have seen Pedro dance; but of course,
b. if Sophie had gone to the parade and been stuck behind someone tall, she would not have seen Pedro dance.

But such a pair of counterfactuals could scarcely be evidence against thinning. The antecedent *Sophie goes to the parade and she is stuck behind someone tall* is simply not got from conjoining *Sophie is stuck behind somone tall* and *Sophie goes to the parade and her view is unobstructed*. So there is no elided qualification in (2), and this speaks—conclusively, I think—against positing it in (4).[5]

[4]This was, I think, first pointed out by Irene Heim and reported in von Fintel (1999). The classic Lewis example is:

(i) If the USA were to throw its nukes into the sea tomorrow, there would be war; but of course, if the USA and all the other superpowers were to throw their nukes into the sea tomorrow there would be peace.

Commuting the conjuncts is not good:

(ii) ??If the USA and all the other superpowers were to throw their nukes into the sea tomorrow there would be peace; but of course, if the USA were to throw its nukes into the sea tomorrow, there would be war.

We will look at von Fintel's analysis in Section 5.

[5]Similarly for the pair in footnote 4: there may be some temptation to argue that the second (unhappy) sequence has an elided *only* or *alone* in the antecedent of the second

The asymmetry between (2) and (4) is troublesome if, like the classic Stalnaker–Lewis semantics, counterfactuals are treated as variably strict conditionals. Since the nearest sphere permitting that Sophie goes to the parade is not guaranteed to include the nearest sphere in which both Sophie goes to the parade and is behind someone tall, (2) is rightly predicted to be consistent. Changing the order in which we find the nearest permitting spheres does nothing to change these facts, and so (4) is also predicted to be consistent.

The point of a Sobel sequence is that counterfactuals are resource-sensitive. The point of a Sobel sequence's ugly cousin—got by reversing the order of the counterfactuals—is that counterfactuals are resource-affecting. There is important interaction between counterfactual antecedents and the parameters of context relevant to their semantics. And that is something a story exploiting the loophole might shed some light on.

3 Gloss

Suppose counterfactuals are strict conditionals. Then they are some necessity modal, scoped over a plain conditional. Just *which* modal depends on context.

Here is an intuitive gloss of the interaction between context and counterfactual. One of the resources provided by context is a domain of worlds over which modals quantify. *If*-clauses presuppose that their complements are entertainable. In the case of a counterfactual *if*-clause, the presupposition is that the complement be possible relative to the domain. Call such a (modal) presupposition an *entertainability presupposition*. This presupposition projects to the entire conditional.[6] If the presupposition isn't met—*ceteris paribus* and within certain limits—it is accommodated, the domain undergoing a bit of change to meet the presupposition. It is with respect to this post-accommodation domain that the counterfactual is a strict conditional.

It is easy enough to see how a story along these lines might handle the delicate facts about thinning. Take a Sobel sequence:

(thinned) conditional. We had better resist such temptations.

[6]A familiar sort of example of a presupposition triggered in an antecedent projecting:

(iii) If Sophie realizes that there is no more ice cream, there will be trouble.

The presupposition that there is no more ice cream, triggered by the factive *realizes*, projects to the whole sentence. So if counterfactual antecedents trigger entertainability presuppositions we should expect them to project to the counterfactuals as a whole.

(7) If had been p, would have been q; but of course, if had been $(p \wedge r)$, would have been $\neg q$

The first conjunct presupposes (in the relevant sense) $\Diamond p$, and asserts that $\Box(p \rightarrow q)$. Suppose the intial domain s_c has no p-worlds. Assume we accommodate: s_c shifts to a slightly larger domain incuding some. The necessity modal then takes this shifted domain as input to its semantics. Suppose that all p-worlds in this posterior domain are q-worlds (so that the first counterfactual is true with respect to this domain). The second (thinned) conditional presupposes (in the relevant sense) $\Diamond(p \wedge r)$ and asserts $\Box(p \wedge r \rightarrow \neg q)$. But the domain s_c-shifted-by-the-first-conditional may well contain no $(p \wedge r)$-worlds—the nearest p-worlds, after all, need not include the nearest $(p \wedge r)$-worlds. And so the domain expands a bit further, the necessity modal for *this* conditional taking this even larger domain as input to its semantics. And it is quite possible that every $(p \wedge r)$-world in this new domain is a $\neg q$-world. So it is that the second counterfactual is a strict conditional over a different, larger domain than is the first. No wonder Sobel sequences can be consistent.

Things are different when we look at a Sobel sequence's ugly cousin. For if we *first* interpret the thinned counterfactual, our domain gets pretty big straightaway: the thinned conjunct presupposes $\Diamond(p \wedge r)$ and asserts $\Box(p \wedge r \rightarrow \neg q)$. Suppose that in the (comparatively larger) post-accommodation domain all the $(p \wedge r)$-worlds are $\neg q$-worlds. This domain is the input for interpretting the original (unthinned) counterfactual. But the presupposition $\Diamond p$ is already met here, and so there is no accommodating shift. But, by hypothesis, all of the $(p \wedge q)$-worlds in the big domain are $\neg q$-worlds. Whence not all of the p-worlds in this domain are q-worlds. The two strict conditionals end up quantifying over the same domain. No wonder a Sobel sequence's ugly cousin cannot be consistent.

What is left is to turn this intuitive gloss into a proper analysis. I will offer a first pass at an analysis, and then draw some initial comparisons. Finally, we will look at *might*-counterfactuals and refine the analysis.

4 Basics

The basic idea is simple. Suppose we divide the semantic labor of counterfactuals, factoring the meaning of a counterfactual into its entertainability presuppositions and its semantic value. Accommodating a missing presupposition can change the relevant contextual parameter for the assignment of semantic value. I have assumed that it is a domain—a set of worlds—that is assigned by or inherited from context. But it is a bit tidier to think of this parameter as an *nested set* of domains assigned by or inherited from context. Let's call them *hyperdomains*. So the semantics first computes the accommodation-induced changes to this parameter, and then passes the value of the parameter so changed to the semantic clause for the modals.

Hyperdomains are nested sets of domains—nested sets of worlds. But not all worlds are created equal. Some worlds—perhaps because they violate laws that we take to be non-negotiable when we entertain counterfactual antecedents or because our particular conversation presumes that such worlds are not relevant—are just not relevant for the truth of counterfactuals. Since such worlds are not relevant, they do not make it into a domain and *a fortiori* sets that include them do not make it into a hyperdomain.

An example: Jones invariably wears his hat on rainy days; on days with no rain, he wears his hat or leaves it home at random. Suppose that, as a matter of fact, it is rainy (and so Jones is hatted). Counterfactual antecedents like *If the weather had been fine* ask us to entertain various fine-weather possibilities, looking to sets of worlds consistent with the weather being fine. But no such domain will include worlds in which Jones has no hat at all. Nor will any of them include worlds in which Jones has different hat-wearing predilections, or worlds in which Jones forgets his hat on a rainy day. This is so even though: it is co-possible with fine weather that Jones has no hat, co-possible with fine weather that Jones has different hat-wearing predilections, and co-possible with fine weather that Jones forgets his hat on a rainy day. The invariance between Jones's hat-wearing and rainy weather is just not up for grabs, actually or counterfactually, and so such worlds do not make it into any of the domains over which the counterfactuals quantify.[7]

Given a set W of possible worlds, assume that for a given bit of counterfactual discourse we can settle upon some upper-bound $U \subseteq W$ encoding the information not up for grabs, actually or counterfactually.[8] A hyperdomain at i will have to respect this by not ordering any domain that is not included in U. Just as clearly, all of the nested domains are domains *around* i—the actual state of affairs is always relevant, though not decisive, to interpreting a

[7] See Pollock (1976) and Veltman (2005) for the special status of laws in the semantics of counterfactual constructions.

[8] I will be happy to assume that W is finite.

counterfactual. Begin by ordering the worlds in U relative to i, the ordering reflecting relative proximity between worlds. Admissible domains are sets of worlds marked off by the ordering: all the worlds not above a chosen w in the ordering form a domain. A hyperdomain is a nested collection of such admissible domains. Insisting that the ordering treat i as (uniquely) minimal guarantees both that i is a lower-bound on domains and that its information is present in all the weaker domains.

To say this properly I first need one bit of auxiliary notation. I have taken no stand on the nature of the indices involved in the story so far: counterfactuals might traffic in—have their semantic values at and their entertainability presuppositions induce accommodation of—*points* of evaluation (worlds, world–time pairs, or whatever) or they might instead traffic in—have their semantic values at and their entertainability presuppositions induce accommodation of—*sets* of such points. I would like to put off taking that stand.[9] But this generality means that we cannot—not straightaway, at any rate—say what we mean when we insist that the ordering over U is centered on i: if i is a point, then we mean that i is the (unique) minimal element in the ordering; if i is a set of such points, then we mean that every point in i is minimal in the ordering. This is more annoyance than problem, so we can define it away: let $\boldsymbol{i}$ be $\{i\}$ if i is a point, and let $\boldsymbol{i}$ be i itself if it is a set of such points. Then centering will always be centering on $\boldsymbol{i}$.[10]

Definition 1. Let $\preceq_{\boldsymbol{i}}$ be a (total) preorder of $U \subseteq W$ ($\preceq_{\boldsymbol{i}}$ is transitive and connected) centered on $\boldsymbol{i}$.

1. (ADMISSIBLE DOMAINS) $\mathbb{D}_i$ is the set of *admissible domains* (around i):

   $$s \in \mathbb{D}_i \text{ iff } \exists w \in U : s = \{v : v \preceq_{\boldsymbol{i}} w\}$$

2. (HYPERDOMAINS) A hyperdomain π (at i) is a $\subseteq$-nested subset of $\mathbb{D}_i$.[11]

A hyperdomain π is just a particular ordering of domains: $\langle s_n, s_m \rangle \in \pi$ only if $s_n \subseteq s_m$. Insisting that $\preceq_{\boldsymbol{i}}$ is centered on $\boldsymbol{i}$ guarantees both that $\boldsymbol{i}$ is the $\subseteq$-minimal admissible domain and that for every $s \in \mathbb{D}_i$ $\boldsymbol{i} \subseteq s$. This makes

[9] For now. Later, in the refined analysis, I will take i to be a set of worlds.

[10] In general, for any set X and $A \subseteq X$ whatever, we can conjure an ordering of X centered on A—the centering requirement just means that: (i) the ordering does not distinguish between any elements in A; (ii) and all elements in A are strictly below any element not in A.

[11] Just *which* subsets will become clear below.

precise the requirement that i be both a lower-bound on domains and that its information is present in all weaker domains.[12]

The intuitive picture is that the context change induced by a counterfactual amounts to a filter on the hyperdomain, only letting those domains pass through that satisfy the relevant entertainability presupposition. That trims the ordering.[13] The (default) initial hyperdomain at i, π_0, is $\mathbb{D}_i$ ordered by $\subseteq$; here no domain has been ruled out. $\pi_\perp = \emptyset$ represents the other limiting case; here we have ruled out too much, leaving no domain consistent with the non-negotiable U. Between the extremes are the hyperdomains reachable by accommodating. A relative modal figuring in counterfactual constructions—for now *must* ($\Box$) is the only relevant modal—then acts as a quantifier over the smallest (i.e., most informative) post-accommodation domain. Equivalently: the relative modal that figures in a *would*-counterfactual is a necessity modal indexed to the smallest post-accommodation domain and scoped over a plain conditional.

Assembling the pieces thus far gives us the following:

Definition 2.

1. COUNTERFACTUAL CCP

$$\pi|\textit{if had been } p\textit{, would have been } q| = \{\langle s_n, s_m\rangle \in \pi : s_n \cap [\![p]\!] \neq \emptyset \text{ and } s_m \cap [\![p]\!] \neq \emptyset\}$$

2. TRUTH-CONDITIONS (COUNTERFACTUALS)

$$[\![\textit{if had been } p\textit{, would have been } q]\!]^{\pi,i} = 1 \text{ iff } [\![\Box(p \to q)]\!]^{\pi',i} = 1$$

where $\pi' = \pi|\textit{if had been } p\textit{, would have been } q|$

[12] The appeal to such underlying orderings in constructing hyperdomains, I admit, leaves the impression that the kind of story I want to tell ultimately relies on the same formal apparatus that drives the (classic) variably strict semantics for counterfactuals. The impression is misleading. All that is needed is that we can order domains (sets of worlds) according to the relative ease with which we fall back to them to entertain possibilities. Nothing in that requires a similarity metric on worlds. We can tell the story that I want to tell—I'll gesture at just how toward the end of the paper—so that it is parametric on a choice of a fallback relation, omitting talk of similarity altogether.

[13] This is similar to the picture of accommodation in Beaver (1999).

3. TRUTH-CONDITIONS (*must*)

$$[\![\Box\varphi]\!]^{\pi,i} = 1 \text{ iff } s_\pi \subseteq [\![\varphi]\!]$$

where φ is non-modal and s_π is the $\subseteq$-minimal domain in π

Interpretation of a counterfactual takes a prior hyperdomain as argument, manipulates it so that the entertainability presuppositions are met in every domain it, and outputs the posterior hyperdomain. This is the context that enters into the truth-conditions for the counterfactual: the smallest—i.e., most factually informative—surviving domain is the set of worlds relative to which the counterfactual is a strict conditional. The posterior hyperdomain is also the context that serves as input to the interpretation of the next bit of counterfactual discourse, should there be any.

5 ***Would*s**

So far I have been happy enough to only consider *would*-counterfactuals. The proposal is that such conditionals are in fact strict: they are some necessity modal, scoped over a plain conditional. Just *which* necessity modal depends on context.

For limiting-case sequences of counterfactuals—counterfactual discourses that stretch only one conditional long—there is no predictive difference between the truth-values assigned by the variably strict semantics and a special case of the story I have told. Assume that i is a world, not a set of such worlds. Then

(8) If had been p, would have been q

is true at i by the lights of the (classic) variably strict semantics iff it is true at $\langle\pi_0, i\rangle$ by the lights of our strict semantics. The variably strict semantics says that (8) is true at i iff the set of ($\preceq_i$-)nearest p-worlds are q-worlds; if i is a p-world, then the counterfactual is true at i iff q is true at i. The strict semantics agrees: we begin with the default hyperdomain π_0 around i, and figure the CCP of the counterfactual, accommodating any entertainability presuppositions that are not met. If i is a p-world, then $\{i\}$—the minimal element in π_0—admits the presupposition, and accommodation idles. Thus, since $s_{\pi_0} = \{i\}$, if p is true at i, then (8) is true at $\langle\pi_0, i\rangle$ iff $\{i\} \cap [\![p]\!] \subseteq [\![q]\!]$—that is, iff q is true at i as well. If p is not true at i, then the CCP is non-trivial; the post-accommodation hyperdomain π' orders all domains from π_0 that allow p. The minimal such domain ($s_{\pi'}$) is just the set of ($\preceq_i$-)nearest worlds. But then $\Box(p \rightarrow q)$ is true at $\langle \pi', i \rangle$ iff $s_{\pi'} \cap [\![p]\!] \subseteq [\![q]\!]$—that is, iff the ($\preceq_i$-)nearest p-worlds are q-worlds.

It is on non-limit-case sequences that differences emerge. But not in Sobel sequences. Here, though, the explanations for the phenomenon are different, and the explanation on offer from a strict semantics like the one we have been considering makes way for predicting the asymmetry between a Sobel sequence and its ugly cousin.

Take a Sobel sequence like that in (2):

(2) a. If Sophie had gone to the parade, she would have seen Pedro dance; but of course,
b. if Sophie had gone to the parade and been stuck behind someone tall, she would not have seen Pedro dance.

Suppose that (2a) is true at $\langle \pi_0, i \rangle$. If we assume that Sophie was not a parade goer at i ($\neg p$), the context change induced by the antecedent is non-trivial and π_0 gets changed to π_1, the ordering of those domains from π_0 compatible with Sophie's parade going (p). In the minimal such domain s_{π_1} all the worlds where Sophie is a parade goer are worlds where she is a witness to Pedro's dancing; $\Box(p \rightarrow q)$ is true at $\langle \pi_1, i \rangle$.

Continuing with the thinned (2b), we first figure its effect on context. Given the assumptions of the story about Sophie and the parade—the same assumptions needed to make this Sobel sequence, by the lights of the variably strict semantics, a case against thinning—accommodating the entertainability presuppositions will be non-trivial. There will be lots of domains ordered in π_1 that do *not* allow that Sophie goes to the parade and is stuck behind someone inordinately tall ($p \wedge r$), the minimal domain s_{π_1} in π_1 among them. So we throw out such domains as unduly restrictive. The resulting domains are ordered in π_2, and we can easily imagine (given assumptions about Sophie's height and the relative difficulty she would have in seeing over a really tall parade goer) that the smallest such domain is one whose only ($p \wedge r$)-worlds are $\neg q$-worlds; in that case $\Box(p \wedge r \rightarrow \neg q)$ is true at $\langle \pi_2, i \rangle$. So the truth of (2a) at $\langle \pi_0, i \rangle$ does nothing to preclude the truth of (2b) at $\langle \pi_1, i \rangle$. Our contextual parameter has accommodated the extra information introduced by the thinned antecedent. This effect on context shows itself in the contents: the two counterfactuals are each strict conditionals, each with a different modal governing its strength. The first is weaker than the second.

Of course the variably strict semantics predicts the same consistency of (2a)–(2b). The ($\preceq_i$-)nearest $(p \wedge r)$-worlds need not be among the ($\preceq_i$-)nearest p-worlds. Whence it follows that the latter's being included in $[\![q]\!]$ does not preclude the former from being included in $[\![\neg q]\!]$. But this explanation of the phenomenon is decidedly different—it straightaway predicts no difference between a counterfactual discourse unfolding as in (2a)–(2b) and the defective reverse order in (4).

A strict semantics like the one we have been considering predicts this asymmetry. Interpreting the thinned (2b) in π_0 and following this with (2a) produces something very different from interpreting a counterfactual discourse that unfolds from (2a) to (2b). By hypothesis i is not a $(p \wedge r)$-world, and so $\pi_0|$*if had been* $p \wedge r$*, would have been* $\neg q|$ will differ from π_0 by removing all domains that do not have a world at which Sophie goes to the parade and is stuck behind someone tall $(p \wedge r)$, and will order those that are left. This is just the hyperdomain π_2. Given the plausible assumptions about her relative height and the relative dissimilarity between i and worlds in which (say) Sophie is on stilts, all of the Sophie-goes-and-is-stuck-behind-someone-tall worlds in the smallest domain in π_2 will be worlds in which she does not see Pedro dance. Formally: $s_{\pi_2} \cap [\![p \wedge r]\!] \subseteq [\![\neg q]\!]$, and so (2b) is true at $\langle \pi_0, i \rangle$. But the score has changed—π_2 is the input context for whatever counterfactual comes next. And in this case it is (2a) that comes next. But since by hypothesis s_{π_2} is compatible with *Sophie goes to the parade and is stuck behind someone tall* $(p \wedge r)$, it is also compatible with *Sophie goes to the parade* (p). Since hyperdomains are $\subseteq$-nested, every domain ordered by π_2 has such witnessing worlds. And so accommodation on the antecedent of (2a) idles—its only entertainability presupposition (that p is possible) is met. That means that (2a) is true at $\langle \pi_2, i \rangle$ iff every p-world in s_{π_2} is a q-world. But by assumption *some* of these p-worlds in s_{π_2} are also r-worlds—and all of *those* are $\neg q$ worlds. So there is just no way for (2a) to be true here, given the truth of (2b). And that means that conjunctions representing such a discourse—conjunctions like (4)—are seriously defective. Successful interpretation of the first conjunct creates a context in which interpreting the second is doomed to failure.

Roughly put: the classic variably strict semantics predicts that whatever context effects are induced by a counterfactual antecedent are forgotten once we have interpreted the consequent. We see each counterfactual in a stretch of counterfactual discourse as though it were the first. Not so for a strict semantics like the one we have been considering. Accommodating counterfactual antecedents changes the score for good. And this can have consequences downstream.

This kind of explanation of the delicate facts about thinning exploits something very much like the the loophole Lewis mentions. My way of exploiting it—for *would*-counterfactuals—agrees, plus or minus just a bit, with the account of counterfactuals von Fintel (1999) proposes.[14] The idea is simply that conditionals are (two-place) quantifiers, and a counterfactual at i, in particular, is a quantifier over a "modal horizon"—a set of relevant worlds at i. Thus a counterfactual like (8) has a logical form along the lines of

(9) $would(p)(q)$

The semantics says that such a quantifier at i takes a contextually inherited modal horizon (domain) D_i, and that it is a universal quantifier over this domain restricted by p. But universal quantifiers carry an existence presupposition on their first argument. Here that amounts to a presupposition that there are p-worlds in D_i. If the presupposition is not met, then we accommodate. Assuming a well-behaved ordering around each point of evaluation i that drives the accommodation, and assuming that the default $D_i^0 = \{i\}$, we have a picture no different from the strict semantics we have been considering:

Definition 3.

1. COUNTERFACTUAL CCP

$$D_i|\textit{if had been p, would have been q}| = D_i \cup \{w : \forall v \in \min(p, \preceq_i) \Rightarrow w \preceq_i v\}$$

2. TRUTH-CONDITIONS

$$[\![would(p)(q)]\!]^{D_i,i} = 1 \text{ iff } D_i|\textit{if had been p, would have been q}| \cap [\![p]\!] \subseteq [\![q]\!]$$

Assume that i is a point of evaluation and (as we have thus far) that the structure of hyperdomains around i are got by appeal to $\preceq_{\boldsymbol{i}}$. Then the story

[14]The accounts, as I say, agree on *would*-counterfactual sequences. But his motivations are a good deal loftier than mine: he is trying to make the semantics of counterfactuals fit better with their NPI licensing behavior. The version of von Fintel's proposal in the text is a bit of a reconstruction: it is equivalent to the form he gives, but it suits my purposes better.

as I have told it and von Fintel's telling of it come to the same thing. The default modal horizon at i, D_i^0, is just s_{π_0}. Accommodating the possibility that p into a modal horizon amounts to adding all worlds between i and the nearest p-world and treating the counterfactual as strict over this set. That is just the same as eliminating domains that do not permit p, and treating the counterfactual as a strict conditional over the smallest domain left.

We might well wonder if anything as exotic as an analysis that exploits the loophole is really needed to explain the delicate facts about thinning. If we allow shifts in context to enter into the explanation, then amending the variably strict semantics accordingly can predict the kind of asymmetry between a Sobel sequence and its ugly cousin.[15] The central apparatus in the variably strict semantics is an ordering over possible worlds, centered on the point of evaluation (or, what comes to the same thing, a well-behaved partition- or premise-function from worlds to sets of sets of worlds). A counterfactual is then true at a point iff its consequent is true at all the closest antecedent worlds in the ordering. But suppose we allow the ordering to evolve, changing as counterfactual assumptions are made: the effect of a successful utterance of a counterfactual *if had been p, would have been q* is to promote the closest p-worlds in the ordering so that they are among the closest world *simpliciter*. This amounts to changing the ordering by lumping together antecedent-facts, and that means that the ordering becomes coarser as more counterfactual antecedents get interpreted.

Formally put:

Definition 4. Given an ordering $\preceq_{\boldsymbol{i}}$ (with field F) and $P \subseteq W$, let $f_{\preceq_{\boldsymbol{i}}}(P)$ be the set of $\preceq_{\boldsymbol{i}}$-minimal worlds in P: $f_{\preceq_{\boldsymbol{i}}}(P) = \{w \in P : \text{ if } v \in P \text{ then } w \preceq_{\boldsymbol{i}} v\}$.

1. (ORDERING CCP)

$$\preceq_{\boldsymbol{i}} |\textit{if had been p, would have been q}| = \preceq_{\boldsymbol{i}} \cup \{\langle w, v\rangle : w \in f_{\preceq_{\boldsymbol{i}}}([\![p]\!]) \text{ and } v \in F\}$$

2. (VARIABLY STRICT TRUTH-CONDITIONS)

$$[\![\textit{if had been p, would have been q}]\!]^{\preceq_i, i} = 1 \text{ iff } f_{\preceq'_i}([\![p]\!]) \subseteq [\![q]\!]$$

where $\preceq'_{\boldsymbol{i}} = \preceq_{\boldsymbol{i}} |\textit{if had been p, would have been q}|$

[15] Something like this alternative was suggested by Angelika Kratzer (p.c.) to Kai von Fintel who in turn posed it to me.

So long as the changed ordering gets passed downstream, such an amendment would explain the asymmetry. Assume an initial ordering $\preceq_i^0$ centered on a world i. Interpreting the first conjunct of (2a) results in an ordering $\preceq_i^1$ in which the ($\preceq_i^0$-)nearest p-worlds are promoted to the closest worlds *simpliciter*. It is then true iff all the ($\preceq_i^1$-)nearest p-worlds are q-worlds—iff all the closest worlds *simpliciter* make the material conditional $p \to q$ true. Interpreting (2b) then coarsens the ordering further, promoting the ($\preceq_i^1$-)closest $(p \wedge r)$-worlds in $\preceq_i^2$. It is true iff, in the posterior ordering, all the closest worlds *simpliciter* make the material conditional $(p \wedge r) \to \neg q$ true. And this is surely possible. Reversing the conjuncts makes a difference because the ($\preceq_i^2$-)nearest worlds *simpliciter* contains p-worlds that are not q-worlds, and these witness the falsity of the first conjunct. And that is enough to predict the asymmetry between a Sobel sequence and its ugly cousin.

Amending the variably strict semantics in this way is not altogether wrong—for sequences of *would*-counterfactuals, it agrees with the strict conditional stories—but it is defeatist. The semantics is now variably strict in name only. Making an ordering coarser idles exactly when accommodation idles, and the smallest post-accommodation domain coincides exactly with the set of nearest worlds *simpliciter*. Variability—that the set of nearest p-worlds neither includes nor is included in the set of nearest $(p \wedge r)$-worlds—plays no role in explaining why counterfactuals do not allow for thinning. That is a role now played by pointing to different orderings of similarity, one for the first conjunct in a Sobel sequence and a coarser one for the second. Nor does the variability recorded in the ordering bear any serious weight in assigning truth-conditions. Once the ordering is coarsened by a counterfactual antecedent *if had been p*, a counterfactual is simply a strict conditional over the set of nearest worlds *simpliciter*. The amended variably strict semantics is an inelegant notational variant of the strict conditional semantics.

I conclude that the amended version of the variably strict semantics is not right. We do better with some version or other of the strict conditional story. But *neither* my way nor von Fintel's way of cashing that out is quite right. They get the data about Sobel sequences right, but they mislocate the phenomenon. This turns up when we look to the context effects triggered by *might*-counterfactuals. I can see how to amend what I think is distinctive about the story I have been telling, but can see no way to amend von Fintel's story.

6 *Might*s

The sort of asymmetry between a Sobel sequence and its ugly cousin is not confined to counterfactual antecedents. What is distinctive about Sobel sequences is that the thinned counterfactual represents a way of weakening the unthinned conditional connection between antecedent and consequent. In the examples we have considered, it is weakened to the point of reversal: the connection between p and q is flipped, in the presence of r, to a connection between p and $\neg q$. Since accommodation is comparably easier to do than undo, no wonder sequences of counterfactuals do not happily commute.

But we might well weaken a claimed counterfactual connection between p and q by calling attention to a substantially weaker connection between p and something incompatible with q. That is one thing *might*-counterfactuals are good for, and the relevant discourses exploiting them exhibit the same resistance to commutation that Sobel sequences do.

Some examples:

(10) a. If Sophie had gone to the parade, she would have seen Pedro dance; but, of course,
b. if Sophie had gone to the parade, she might have been stuck behind someone tall and then wouldn't have seen Pedro dance

(11) a. If Hans had come to the party he would have had fun; but, of course,
b. if Hans had come to the party, he might have run into Anna and they would have had a huge fight, and that would not have been any fun at all

These are consistent, even if complicated, stretches of counterfactual discourse. Call such stretches *Hegel sequences*.[16] As with a Sobel sequence, a Hegel sequence's ugly cousin is dramatically worse:

(12) a. ??If Sophie had gone to the parade, she might have been stuck behind someone tall and then wouldn't have seen seen Pedro dance;

[16] Kratzer (1981b) considers "Hegel's counterfactual": if we have a view of the world that links all facts together, then changing one fact may change all facts; and so a partition function that respects this will treat a counterfactual *if had been p, would have been q* as true iff p entails q. The kinds of sequences involving *might*-counterfactuals that I am interested in seem to push us, even if just a bit, toward Hegel's counterfactual.

but, of course, if Sophie had gone to the parade, she would have seen Pedro dance

b. ??If Hans had come to the party, he might have run into Anna and they would have had a huge fight, and that would not have been any fun at all; but, of course, if Hans had come to the party he would have had fun

These are pretty bad—about as bad as their Sobel counterparts. Sobel sequences (and their ugly cousins) show how entertainability presuppositions triggered in counterfactual *antecedents* can contribute to a certain kind of shifty behavior of whole conditionals. Hegel sequences (and their ugly cousins) show how entertainability presuppositions triggered in counterfactual *consequents* can contribute to the same kind of shifty behavior. This is good reason to hold out for a unified account of both.

The variably strict semantics, of course, has no trouble predicting that a Hegel sequence's ugly cousin is somehow defective. For assume that *would*- and *might*-counterfactuals are duals. Then the first conjunct of (12a) is true at i just when some of the ($\preceq_i$-)nearest worlds where Sophie is a parade goer are worlds where she gets stuck behind someone tall and misses out on seeing Pedro dance. But those same worlds are exactly those that make the counterexample to the second conjunct true. The trick is to predict that the Hegel sequences are consistent, that their ugly cousins are not, and to do both while still treating *would*s and *might*s—if not the conditional constructions involving them, then at least the unary modals that figure prominently in them—as duals.

That is a trick that a story exploiting the loophole seems just right for. That is good news since the phenomena surrounding Hegel sequences looks an awful lot like the phenomena surrounding Sobel sequences. But not every story exploiting the loophole is up to pulling this off.

Part of the strength and appeal of von Fintel's proposal lies in assimilating entertainability presuppositions to a more familiar and less exotic fact about quantifier domains. A counterfactual *if had been p, would have been q* is just a two-place quantifier $would(p)(q)$, and so the entertainability presupposition triggered by the antecedent is just an instance of an existence presupposition on the quantifier's first argument. That is very tidy. I can claim no such strength since I have only gestured that a counterfactual antecedent presupposes in some sense or other $\Diamond p$ without saying how or why or in just what sense.

But this strength in von Fintel's proposal cuts both ways. Suppose we extend von Fintel's story in the obvious way to *might*-counterfactuals—the relevant logical form is got by swapping the universal *would* for the existential *might*:

(13) $might(p)(\neg q)$

Assimilating the triggering of entertainability presuppositions to existence presuppositions on the quantifier in (13) does not seem right. The phenomenon is not where we would expect it, given that kind of assimilation. The action in (10b) and (11b) is in the consequent, and that means that that is where the shiftiness is triggered. But, given the quantifier presupposition diagnosis, we would not at all expect this: quantifiers carry existence presuppositions on the *first* argument, but the phenomenon seems to be in its second. I am happy enough to agree that

(14) Every student smokes Reds

presupposes that there are some (relevant) students. But this does not presuppose that there are (relevant) Reds.

Thinking of the shiftiness of counterfactuals in terms of quantifier domain presupposition also commits us to thinking of entertainability presuppositions as *real* presuppositions.[17] This would be a good thing if the distribution of facts lines up in a neat way. But I do not think it does. A typical *might*-counterfactual *if had been p, might have been* $\neg q$ triggers some shiftiness to the effect that $\Diamond(p \wedge \neg q)$, but this material fails all standard presupposition tests: it is not backgrounded, the negated *might*-counterfactual doesn't trigger it, and—what I take to be decisive—it fails the "Hey, wait a minute" test.[18] If you utter (11b), I cannot reply with

[17] I have called these "entertainability presuppositions" largely because I can think of no better name. There is *some* presupposition-y feel to them, but I hesitate to think of them as a straightforward modal variant of of what goes on with other classes of well-known presupposition triggers.

[18] If φ presupposes ψ a hearer can legitimately complain when her conversational partner utters φ by stopping her in her tracks—"Hey, wait a minute. I had no idea that ψ." Compare:

(iv) a. *A*: If Sophie realizes that there is no more ice cream, there will be trouble
b. *B*: Hey, wait a minute. I had no idea there was no more ice cream.
c. *C*: ?? Hey, wait a minute. I had no idea Sophie would get mad.

See von Fintel (2003) for a semantic account of these pragmatic facts about presupposition (at least presuppositions about certain European monarchs).

(15) ??Hey, wait a minute. I had no idea that Hans might have run into Anna at the party.

Informing (or, perhaps, reminding) me that Hans might have run into Anna at the party is why you said what you did. So that I had no idea that might have happened cuts little ice.

Although the failure here contrasts rather sharply with the behavior of other presuppositions, it does pattern well with the sort of entertainability triggered by counterfactual *if*-clauses. You say *if Hans had come to the party, we would have run out of punch.* I cannot register a complaint

(16) ??Hey, wait a minute. I had no idea that Hans might have come to the party.

That Sobel sequences cannot be happily commuted points to shifty behavior of counterfactual antecedents. That Hegel sequences cannot be happily commuted points to shifty behavior of counterfactual consequents. It is better to have one explanation for both phenomena, and that means that this shiftiness is not the shiftiness of accommodating existence presuppositions on quantifier domains.

7 Schmontent

I want to revisit and refine the strict conditional account, where the work is done by a relative modal and not by a quantifier. Of course, modals behave in *some* respects like quantifiers over sets of worlds. But this does not mean that they must inherit *all* the semantic properties of quantifiers. Even given the quantificational view of modality, there is room for thinking that the relationship between a counterfactual and its entertainability presuppositions is not quite the relationship between a quantifier and its existence presuppositions.

Both von Fintel's proposal and the strict conditional account divide semantic labor. We are pretending that the only relevant changes to context are changes by accommodation—expanding domains to make the entertainability presuppositions of modal constructions met. Contents are figured by reference to post-accommodation domains. Each story thereby dived semantic labor: there is the CCP-assigning bit of semantic machinery, and then there is the (truth-conditional) content-assigning bit of semantic machinery.

I do not think this is the way to go. Insisting on dividing semantic labor in the way we have when it comes to *might*-counterfactuals raises problems I cannot solve. Very roughly: the problem is that if we follow the example of *would*-counterfactuals for how the labor is divided, all (or, anyway, too many) *might*-counterfactuals are everywhere true; if we do not follow the *woulds*, then we introduce truth-value gaps where there were none before. Since it is best to hold out for a unified story of *might*- and *would*-counterfactuals, we should not divide semantic labor in this way.

Less roughly: consider a *might*-counterfactual.

(17) a. If had been p, might have been q
b. $might(p)(q)$

The shiftiness of such constructions lives on the shiftiness of *might*, a two-place relative modal operator. This shiftiness is a bit peculiar—such existential relative modals do seem to be governed by an accommodation mechanism, but accommodating here *makes* the modal claim true. We are trying to see if Ruud is at to the party, and have been going by the bikes and scooters parked outside. Ruud's is nowhere to be seen (though I have spotted Alex's), and so we conclude that he is not at the party. But you bring up a possibility we had been ignoring:

(18) Maybe he came with Alex on her Vespa.

Accommodating the possibility makes what you said true. An existential modal claim needs just one witnessing world to come out true, so the entertainability of its complement entails the truth of the modal claim. Hence this sort of modal accommodation will pretty much guarantee that the relative modal is true. That means that our accommodation mechanism for *might*-counterfactuals will—if no one objects and accommodation does its job—invariably land us in a context in which the counterfactual is true.[19]

This behavior is peculiar—accommodating *other* sorts of presupposition certainly does not work this way. But peculiarity does not a problem make. That accommodating an existential relative modal—taking on board its entertainability presupposition—makes the modal true is just a fact about the

[19]See Lewis (1979, pp. 246–247). This kind of behavior is, presumably, why Kratzer describes accommodation for relative modalities by saying "This is black magic, but it works in many cases" (Kratzer, 1981a, p. 311).

semantics of these constructions. There is a real problem in the vicinity, given a division of labor in the semantics between the context-changing bits and the context-sensitive bits. The semantics for *might* would seem to be completely straightforward. Consider an utterance of *if had been p, might have been q* in a context where the domain over which the *might* quantifies is s_c. Relativize interpretation to some-domain-or-other s^* and have it call the context-affecting bits of the semantics:

(19) $[\![might(p)(q)]\!]^{s^*,i} = 1$ iff $s_c|might(p)(q)| \cap [\![p \wedge q]\!] \neq \emptyset$

And now we ask: When a *might*-counterfactual is uttered in a context with domain s_c, what set of worlds should $s_c|might(p)(q)|$ return? And: When so uttered, what value should s^* take?

Take these in turn. Constrain the class of candidate-CCPs by constraining $s_c|might(p)(q)|$. Any plausible candidate-CCP for $might(p)(q)$ must return a superset of s_c containing some $(p \wedge q)$-worlds; if s_c is such a superset, accommodation idles. Accommodation mechanisms flouting these constraints will not predict the shiftiness of Hegel sequences.

But now the second question is more than just a little awkward. The obvious thought is that *might*-counterfactuals follow the lead of *would*s: the semantic value of a *might*-counterfactual, uttered in a context with domain s_c, should be relative to the prior (pre-accommodation) domain s_c: identify s^* with s_c. That will not do. It entails that if p and q are logically compatible (modulo U), then *if had been p, might have been q* is everywhere true. This is obviously the case if $s_c \cap [\![p \wedge q]\!] \neq \emptyset$, in which case accommodation idles. So suppose that s_c has no $(p \wedge q)$-worlds—this is just the kind of context that allows for Hegel sequences, and in which accommodation has work to do. But $s_c|might(p)(q)|$ will in any case contain $(p \wedge q)$-worlds, making the counterfactual true with respect to s_c. Whence it is everywhere true.

So suppose *might*-counterfactuals do not follow the lead of *would*s. Even though *if had been p, might have been q* is *uttered* in one context s_c, it is always *evaluated* with respect to some different context—if, that is, it happens to be uttered in a context in which accommodation has any work to do. And that means that, for interesting choices of s_c—including the kinds the consistency of Hegel sequences trade on—the interpretation function becomes undefined: there just is no fact of the matter about the content of $might(p)(q)$ at s_c. I have no in-principle beef with gappiness, but this gappiness is unwelcome.

So assuming that we divide semantic labor, *might*-counterfactuals pose a dilemma. That is more than a little embarrassing. They are not everywhere true, and they do not induce gaps. But the truth of the matter is nearby: successfully uttering a *might*-counterfactual always lands you in a context in which it is true. That is something hard to say if we divide semantic labor. But we can say it pretty easily if we take the semantics of counterfactuals to be CCPs all the way down, dispensing with the division of semantic labor by dispensing with the assignment of contents of the normal sort. I can see how to amend accordingly my way of telling a strict conditional story about counterfactuals, but cannot see how to do that with von Fintel's way.

8 Refinement

Counterfactuals are shifty through and through. Sobel sequences (and their ugly cousins) point to shiftiness triggered by counterfactual antecedents; Hegel sequences (and their ugly cousins) point to shiftiness triggered by the consequents of *might*-counterfactuals. But these two phenomena are really one, and this can be explained by a strict conditional account that locates the source of shiftiness in an accommodation mechanism governing the relative modal *might*.

Here is how. Counterfactual antecedents test incoming contexts to see if the antecedent is possible with respect to that context. If not, and assuming the conversation does not derail, accommodation makes it so and the conditional is interpreted in the changed context. A *would*-counterfactual is a strict conditional where the universal modal quantifies over the post-accommodation domain. It tests the context again: are all antecedent worlds consequent worlds? If so, fine; if not, too bad. A *might*-counterfactual is dual to the corresponding *would*: thus it is the conjunction of antecedent and consequent, under the scope of *might*. This invites another test of the context, but (assuming antecedent and consequent are compatible modulo the laws) it is a test that will either succeed or be accommodated. Either way the effect on context is that after successfully updating the information at hand with a *might*-counterfactual, the contextually relevant domain will contain some antecedent-plus-consequent worlds. The refined account will make this precise. The resulting proposal will be a semantics for stretches of counterfactual discourse that identifies their meaning with how they affect conversational score.

It is useful, though not required, to think of the proposal as offering a semantics of counterfactual constructions that is mediated twice over. Counterfactuals first get represented in terms of *must* and *might*. The accommodation-inducing behavior of the relative modals is figured and those are projected into the formal representation in terms of $\Box$ and $\Diamond$. That $might(p)$ triggers accommodation so that it comes out true can be modeled as $might(p)$ presupposing $\Diamond p$ and asserting $\Diamond p$. Since $must(p)$ has no such presuppositions it simply asserts $\Box p$. The semantics assigns values—CCPs—over the fragment with boxes, diamonds, and presupposition operators.[20]

The fiction is useful, so I adopt it. For *would*s things go pretty much as before, using $\partial\varphi$ to represent that φ is, in the relevant sense, presupposed:

(20) a. If had been p, would have been q
b. $might(p); must(p \to q)$
c. $(\partial\Diamond p; \Diamond p); \Box(p \to q)$

Since the entertainability presupposition is importantly different from other presuppositions—it does not pass the "Hey, wait a minute test"—we do just as well to treat it as something that the conditional flat-out claims. That is really what we have done here.

The case with *might*-counterfactuals:

(21) a. If had been p, might have been q
b. $might(p); might(p \wedge q)$
c. $(\partial\Diamond p; \Diamond p); (\partial\Diamond(p \wedge q); \Diamond(p \wedge q))$

But (21c) is clearly overkill: since $\Diamond\varphi$ is an existential modal, the second sequence will (on any sensible semantics) entail the first. So we can replace (21c) with the equivalent, and simpler:

(21) c′. $\partial\Diamond(p \wedge q); \Diamond(p \wedge q)$

[20]The picture can be formalized. Define the translation function Tr—from the (quasi) English expression of counterfactuals to the relevant sentences of the fragment with boxes, diamonds, and presupposition operators—as follows:

1. $\mathsf{Tr}(\text{if had been } p\text{, would have been } q) = \mathsf{Tr}(might\ p; must(p \to q))$
2. $\mathsf{Tr}(\text{if had been } p\text{, might have been } q) = \mathsf{Tr}(might\ p; might(p \wedge q))$
3. $\mathsf{Tr}(\phi; \psi) = \mathsf{Tr}(\phi); \mathsf{Tr}(\psi)$
4. $\mathsf{Tr}(might\ \phi) = \partial\Diamond\phi; \Diamond\phi$
5. $\mathsf{Tr}(must\ \phi) = \Box\phi$

Assuming—as I am happy to—that *might* and *must* are duals straightaway yields that $might(p \wedge q)$ abbreviates $\neg must(p \to \neg q)$, whence that *might*- and *would*-counterfactuals are duals.

The semantic value of a counterfactual is the semantic value of its representation in the fragment with boxes, diamonds, and presuppositional operators.

I confessed early on that my interest is less in the content of counterfactuals than it is in getting straight about their context change potential—how successful utterances of stretches of counterfactual discourse change the information state of a hearer when she accepts the news conveyed by it. So, as before, assume a set $U \subseteq W$ of worlds. Given a set i of worlds—characterizing what is settled in the conversation or the hearer's factual information—I will say how $\langle \pi, i \rangle$ is changed by a successful utterance of a counterfactual by saying how it impacts the hyperdomain π.[21] Given i we form the set of admissible domains $\mathbb{D}_i$ as before; hyperdomains around i are as before. Since π is a function of i, we generally suppress i and treat the semantics as taking hyperdomains to hyperdomains.

Changes to hyperdomains are changes brought by looking to the would-be changes of the domains that make them up. The would-be changes on individual domains record both the satisfaction of the entertainability presuppositions associated with the modals—if they are not met the would-be update is undefined—and the quantificational force of those modals. These effects then percolate to the hyperdomain.

It is easier to digest this if we break it into two separate definitions. First, domains:

Definition 5.

1. DOMAIN CCP

   a) $s[\Diamond\varphi] = \{w \in s : s \cap [\![\varphi]\!] \neq \emptyset\}$

   b) $s[\Box\varphi] = \{w \in s : s \subseteq [\![\varphi]\!]\}$

   c) $s[\partial\varphi] = s$ if $s[\varphi] \neq \emptyset$

2. TRUTH AT DOMAINS

   $s \vDash \varphi$ iff $s[\varphi] = s$

Relative modals test the domain: *might* testing that domain has some φ-worlds, *must* that it has only φ-worlds. (These clauses are not defined if φ

[21] Formally, this gloss of the machinery is not obligatory: i could just as well be a point of evaluation.

carries another modal, but that situation will not arise for us.) The effect of presupposed material—what we might better call *non-proffered* material—is to test for the satisfaction of that material. A relative modal is true at a domain just in case the information it carries is already present in that domain.

We now lift the whole process to hyperdomains:

Definition 6.

1. HYPERDOMAIN CCP

   a) $\pi + \Diamond\varphi = \{\langle s_n, s_m\rangle \in \pi : s_\pi \vDash \Diamond\varphi\}$

   b) $\pi + \Box\varphi = \{\langle s_n, s_m\rangle \in \pi : s_\pi \vDash \Box\varphi\}$

   c) $\pi + \partial\varphi = \{\langle s_n, s_m\rangle \in \pi : \exists\langle s'_n, s'_m\rangle \in \pi$ such that $s'_n[\varphi] = s_n$ and $s'_m[\varphi] = s_m\}$

2. TRUTH AT HYPERDOMAINS

   $$\pi \vDash \varphi \text{ iff } \pi + \varphi = \pi$$

It follows straightaway that $\Box\varphi$ is true at a hyperdomain π iff the minimal domain in π has only φ-worlds in it. For $\Box\varphi$ is true at π iff π is a fixed point of $+\Box\varphi$ iff $s_\pi \vDash \Box\varphi$. And this iff s_π is a fixed point of $[\Box\varphi]$ iff all the worlds in s_π are φ-worlds. Similarly for $\Diamond\varphi$: it is true at π iff the minimal domain in π has some φ-worlds in it.

An entertainability presupposition like $\partial\Diamond\varphi$ affects a context $\langle\pi, i\rangle$ by seeing just what domains in π allow φ. Suppose s is a domain ordered by π. If s has no φ-worlds, then $s[\Diamond\varphi] = \emptyset$ and hence $s[\partial\Diamond\varphi]$ will be undefined. Whence no pair in π in which s occurs will survive the update of π with $\partial\Diamond\varphi$. Entertainability presuppositions do just what we wanted to hyperdomains: they filter the ordering, eliminating domains that do not meet the presupposition.

Counterfactuals induce just the changes that their representations induce, and are true just when their induced changes idle. Officially:

Definition 7.

1. *would*-COUNTERFACTUAL CCP

   $$\pi + \textit{if had been } p\textit{, would have been } q = \pi + \partial\Diamond p; \Box(p \to q)$$

2. *might*-COUNTERFACTUAL CCP

   $$\pi + \textit{if had been } p\textit{, might have been } q = \pi + \partial\Diamond(p \wedge q); \Diamond(p \wedge q)$$

3. TRUTH

$$\langle \pi, i \rangle \vDash \textit{if had been p, would/might have been q} \text{ iff}$$
$$\pi + \textit{if had been p, would/might have been q} = \pi$$

This is definitely an analysis in the spirit of the loophole: *would*-counterfactuals are strict conditionals, a universal modal scoped over a material conditional—just which modal a function of context; *might*-counterfactuals are dual to *would*s; and the semantics of each construction is driven by the semantics given to the unary relative modals that figure prominently in them. But it is not defeatist.

Defeatist or not, however, this would not be worth the bother if it buys no better explanation of the shiftiness of counterfactual antecedents and consequents. Saying just when stretches of counterfactual discourse are consistent gives us the makings of that explanation. There are two semantic concepts nearby. A sequence of counterfactuals is *consistent* just in case it can be interpreted without collpase into absurdity. That sequence is *cohesive* iff there is a non-trivial hyperdomain that is a fixed point of the update with the sequence.

Definition 8.

1. (CONSISTENCY) $\varphi_1; \ldots; \varphi_n$ is *consistent* iff there is a $\langle \pi, i \rangle$ such that $(\pi + \varphi_1) + \cdots + \varphi_n \neq \pi_\perp$

2. (COHESIVENESS) $\varphi_1; \ldots; \varphi_n$ is *cohesive* iff there is a $\langle \pi, i \rangle$ such that $(\pi + \varphi_1) + \cdots + \varphi_n = \pi \neq \pi_\perp$

Cohesiveness implies consistency, but not conversely. Consistency requires that the possibility of successful update; cohesiveness requires the possibility of a non-absurd fixed point of an update. Whence, while consistency requires only that the information lead somewhere without collapse, cohesiveness requires that all the information in the sequence hang together in a single state. That means that cohesiveness, but not consistency, is sensitive to accommodating shifts—such shifts are compatible with consistency but not cohesiveness. The explanation is now straightforward: Sobel sequences and Hegel sequences are consistent, but not cohesive. Their ugly cousins are neither.

Sobel sequences are explained as before. You utter a simple counterfactual *if had been p, would have been q*. I update $\langle \pi_0, i \rangle$ accordingly. The change to π_0 induced by $\partial \Diamond p; \Box(p \to q)$ is, first, to make room for $\Diamond p$. Assume that

i contains no p-worlds. Then $\pi_0 + \partial\Diamond p$ will allow $\langle s_n, s_m \rangle$ from π_0 to pass through to π_1 just in case each of the s's have a p-world. It is easy to imagine that $\Box(p \rightarrow q)$ is true at π_1 in virtue of its being true at s_{π_1}—that is, in virtue of the ($\preceq_i$-)nearest p-worlds being q-worlds. But now you utter the thinned *if had been* $(p \wedge r)$*, then would have been* $\neg q$. Updating begins where it left off: I update π_1 with $\partial\Diamond(p \wedge r)$ and update the result with $\Box(p \wedge r \rightarrow \neg q)$. So π_1 is refined further to π_2: all ordered domains in π_1 not allowing $(p \wedge r)$ are removed. It is easy to imagine the facts about relative proxmity allowing that in the smallest domain left all of the $(p \wedge r)$-worlds are $\neg q$-worlds. And so, since $s_{\pi_2} \cap [\![p \wedge r]\!] \subseteq [\![\neg q]\!]$, $\Box(p \wedge r \rightarrow \neg q)$ is true at π_2. In the normal case, such a π_2 will not be absurd and so the sequence is consistent.

But not so in reverse order. Jumping straightaway to π_2 from π_0 is very different. Although it is still fine that $\Box(p \wedge r \rightarrow \neg q)$ is true at π_2, attempting to interpret the unthinned, simple counterfactual would now be a disaster. Updating π_2 with $\partial\Diamond p$ idles—every domain ordered in π_2 allows $(p \wedge q)$, and *a fortiori* allows p. But on the assumption that $\Box(p \wedge r \rightarrow \neg q)$ is true at π_2, it follows that $s_{\pi_2} \cap [\![p]\!] \not\subseteq [\![q]\!]$. Hence $s_{\pi_2}[\Box(p \rightarrow q)] = \emptyset$. And so the hyperdomain collapses upon updating with the simple counterfactual: $\pi_2 + \Box(p \rightarrow q) = \pi_\perp$. So a Sobel sequence's ugly cousin is inconsistent.

The explanation for Hegel sequences is exactly the same—that is an advantage worth claiming. You utter *if Hans had come to the party, he would have had fun*. Update $\langle \pi'_0, i \rangle$ accordingly, first by updating π'_0 with the entertainability presupposition $\partial\Diamond p$. Only those domains ordered by π'_0 that have $p-$worlds survive to π'_1. And, as before, we test $\Box(p \rightarrow q)$ here. Assume that, indeed, $s_{\pi'_1} \cap [\![p]\!] \subseteq [\![q]\!]$. Now you continue: *But, of course, if Hans had come to the party, he might have run into Anna and they would have had a huge fight, and that would not have been any fun*. For simplicity, let's gloss this as *if had been p, might have been r* where $[\![r]\!] \subseteq [\![\neg q]\!]$. Update accordingly, first resolving the effects of accommodation: $\pi'_1 + \partial\Diamond(p \wedge r)$. We thus get rid of any domains ordered by π'_1 that have no $(p \wedge r)$-worlds—$s_{\pi'_1}$ will be one such casualty. In the resulting hyperdomain π'_2 every domain will have some $(p \wedge r)$-worlds, and hence some $(p \wedge \neg q)$-worlds. Those worlds—in particular those in $s_{\pi'_2}$—are sufficient to guarantee the truth of $\Diamond(p \wedge \neg q)$ at π'_2. Such a π'_2 will not be absurd and so the sequence is consistent.

But not so in reverse order. Jumping straightaway to π'_2 from π'_0 is very different. It is still true at π'_2 that $\Diamond(p \wedge \neg q)$. But attempting to update π'_2 with $\partial\Diamond p; \Box(p \rightarrow q)$ would be a disaster. The entertainability presupposition is met, and so accommodation idles. But no non-absurd hyperdomain—and so not π_2'—that makes $\Diamond(p \wedge \neg q)$ true can be updated with $\Box(p \to q)$ without collapse. Since $s_{\pi_2'} \cap [\![p \wedge \neg q]\!] \neq \emptyset$, it follows that $s_{\pi_2'} \cap [\![p]\!] \not\subseteq [\![q]\!]$—and so $\pi_2' + \Box(p \to q)$ will be absurd. So a Hegel sequence's ugly cousin is inconsistent.

Sobel and Hegel sequences are consistent. But the information they carry cannot all hang together at once. Both require an accommodating shift midway through—that is the mark of incohesiveness. In each case some non-trivial shift is needed to make that bit of counterfactual discourse make sense. But once shifted, there is no guarantee that the preceding stretch could be truthfully uttered in the context thus changed. That is why they cannot be consistently reversed.

9 Reformulations

I want to offer two reformulations of the strict conditional analysis. Two features loom large in the analysis: there is a nested set of domains that accommodation operates as a filter on, and that set is generated in a more or less straightforward way from an underlying ordering recording relative proximity between worlds. The first feature is inessential in the sense that we might instead—with equivalent results—treat accommodation as a means of making (unstructured) domains ever-larger. The second feature is inessential in the sense that nothing I have said at all depends on making any assumptions about an underlying ordering of proximity between worlds. The reformulations are meant to make this clear.

The first reformulation takes the semantics of counterfactual discourse to affect and be affected by domains, not hyperdomains. As before, admissible domains around i are got by forming sets around i in a way that is faithful to the underlying ordering $\preceq_{\boldsymbol{i}}$. Each domain around i is simply a sphere centered on i. Before, we separated the accommodation-induced effects of *might* from its quantificational-effects. That is why *might*(p) gave rise to the representations $\partial\Diamond p$ and $\Diamond p$. I still think that is useful and we can stick to that if we like. But there are other options. We can model the impact that *might* has on a context without this separation, making accommodation part of the update directly. The picture is that accommodation triggers a domain shift by making the prior domain bigger. I assume that a domain shifts to allow the possibility that φ only if $[\![\varphi]\!] \neq \emptyset$.

Definition 9.

1. DOMAIN SHIFT
   $s \diamond \varphi = s'$ iff s' is the smallest set in $\mathbb{D}_i$ such that $s \subseteq s'$ and $s' \cap [\![\varphi]\!] \neq \emptyset$.

2. DOMAIN CCP

   a) $s[\Diamond\varphi] = \{w \in s \diamond \varphi : s \diamond \varphi \cap [\![\varphi]\!] \neq \emptyset\}$

   b) $s[\Box\varphi] = \{w \in s : s \subseteq [\![\varphi]\!]\}$

3. COUNTERFACTUALS

   a) s[*if had been p, would have been q*] $= s[\Diamond p][\Box(p \to q)]$

   b) s[*if had been p, might have been q*] $= s[\Diamond p][\Diamond(p \wedge q)]$

Consistency and cohesiveness are redefined to appeal to domains instead of hyperdomains:

Definition 10.

1. (CONSISTENCY, DOMAIN-WISE) $\varphi_1; \ldots; \varphi_n$ is *consistent*$_d$ iff there is an $\langle s, i\rangle$ such that $s[\varphi_1]\ldots[\varphi_n] \neq \emptyset$

2. (COHESIVENESS, DOMAIN-WISE) $\varphi_1; \ldots; \varphi_n$ is *cohesive*$_d$ iff there is an $\langle s, i\rangle$ such that $s[\varphi_1]\ldots[\varphi_n] = s \neq \emptyset$

Assume that the default, initial domain is $s_0 = i$. Clearly $s_{\pi_0} = s_0$. And, just as clearly, if π results from updating π_0 with a stretch of counterfactual discourse $\varphi_1; \ldots; \varphi_n$ and if s results from updating s_0 with that same stretch, then $s_\pi = s$. That stretch is consistent (cohesive) iff it is consistent$_d$ (cohesive$_d$). The two stories are notational variants.

But there is a (small) point to the variation. The reformulation invokes a change operation on sets of worlds. This change operation corresponds to a specific contraction operator in belief dynamics—*severe withdrawal*.[22] Contraction functions model a limit case of belief change where beliefs are removed but not added. Since belief states are in a state of logical equilibrium, constructing such operations from states to states is non-trivial. Each such construction represents some trade-off between conservatism—keep believing as much as you can—and egalitarianism—treat like beliefs alike, where likeness is measured by resistance to change or importance or whatever. Severe withdrawal favors

[22] See Rott and Pagnucco (1999).

the egalitarian over the conservative. Accommodation as we have put it coincides exactly with severe withdrawal: contracting a belief set by removing φ corresponds exactly to the problem of accommodating $\Diamond\neg\varphi$ in a set of worlds characterizing that belief set. Different stories about accommodation might have opted for different constraints, and corresponded to more conservative change operations. We might have missed this connection between accommodation and the dynamics of belief if we did not bother to reformulate.

The second reformulation serves a different purpose. Accommodation looms large in any strict conditional story exploiting the loophole. As I have told the story, the dynamics of counterfactual domains are driven in large part by an underlying ordering of worlds. That leaves the impression that the strict conditional story ultimately relies on the same apparatus that drives the classic variably strict semantics for counterfactuals. But, even setting aside issues of contextual dynamics, there are problems with *any* similarity-based semantics for counterfactuals—facts and predictions do not happily align.[23] Better to tell the story in a way that is independent of this issue.

The strict conditional analysis exploits hyperdomains—ordered sets of domains around a given i. These hyperdomains inherit their structure from an underlying ordering: accommodation trims a hyperdomain, and this amounts to taking a system of spheres and successively eliminating the innermost spheres. But such systems need not be generated by orderings of overall comparative similarity. All that is needed is a *fallback relation* and a proposition—these generate a *hyperproposition*; hyperdomains inherit both their name and their structure accordingly.[24]

Given a proposition x, suppose we have associated with it a fallback relation F—a relation recording the series of propositions we fall back to if retreating from x. A hyperproposition about x is just the closure of x under the fallback relation: $H_x = x^* = \{y : xFy\}$. The structure of a hyperproposition is determined by the properties of the fallback relation. Assume at least this structure: F is reflexive, transitive, inclusive, and quasi-connected.[25] Quasi-connectedness plus inclusion entail that F induces a linear order. Where W is finite, this implies the Limit Assumption for F.[26] Orderings of similarity

[23]See Kratzer (1989) and Veltman (2005).

[24]See Fuhrmann (1999) for more on the structure of hyperpropositions. Grove (1988) was the first to use (a version of) them in belief dynamics.

[25]Inclusion: xFy implies $x \subseteq y$; Quasi-Connectedness: xFy and xFz implies yFz or zFy.

[26]Limit Assumption: if H_x is a hyperproposition about x and y is a proposition, there is

induce one choice of F—that is why a system of spheres is equivalent to a set-valued selection function—but others are possible.[27] Although officially a hyperproposition H_x is a set of propositions, it is sometimes convenient to think of it simultaneously as the ordering of H_x by F. Let's not fuss over the difference.

The set $\mathbb{D}_i$ of admissible counterfactual domains (at i, with repsect to a choice of fallback relation F) is simply the hyperproposition $H_{\boldsymbol{i}}$. A hyperdomain around i is an ordered set of admissible domains, the structure of the ordering inherited from the structure of the hyperproposition $H_{\boldsymbol{i}}$. From here, the analysis goes as before. What I have said about counterfactual score can be said so that it is parametric on a choice of F.

10 Conjecture

I have tried to make a case for exploiting a loophole. The loophole takes *would*-counterfactuals to be strict conditionals, a universal modal scoped over a material conditional. But the force of that modal is highly context dependent. Most of the work lies in getting straight about the interaction between context and semantic value. It is not a *mere* loophole because the facts about counterfactuals in context push us toward it.

If counterfactuals are just strict conditionals, what of indicatives? I think they are strict, too, amounting to a universal (epistemic) modal scoped over a material.[28] The interpretation of the epistemic modal is highly context dependent. Most of the work lies in getting straight about the interaction between context and semantic value. The semantics of the conditional constructions are largely the same. They differ in what domains they test and so on what parameter accommodation operates.

Let i be the set of worlds compatible with the facts I have. The change induced by an indicative *if p, q* amounts to testing i with the strict $\Box(p \rightarrow q)$. So far—if there are p-worlds in i—this looks just like the counterfactual. What if there are no antecedent worlds? Well, then we accommodate. But here a difference emerges. Accommodation triggered by counterfactuals—either in the *would*- or *might*- flavors—has the result of shifting the counterfactual domain. Accommodation triggered by indicatives, on the other hand, shifts the set of

a smallest $z \in H_x$ such that $z \setminus x \neq \emptyset$.

[27] Veltman's (2005) counterfactual retraction operator induces a particularly nice fallback relation since the fallbacks thus induced respect dependencies between facts in a context.

[28] See Gillies (2004).

worlds contending for the actual world. That is a different thing and predicts that accommodation in the two constructions will be sensitive to different pressures. Since accommodation triggered by an indicative introduces new worlds previously ignored or ruled out as contenders, we should expect accommodation here to interact in interestingly different ways with factual discourse. Or so I conjecture.

Bibliography

Asher, N. (2005). "*Were*, *would*, *might* and a compositional account of counterfactuals," ms., The University of Texas at Austin.

Beaver, D. (1992). "The kinematics of presupposition," in: Dekker, P. and M. Stokhof (eds.), *Proceedings of the Eigth Amsterdam Colloquium* (Amsterdam: ILLC).

Beaver, D. (1999). "Presupposition accommodation: A plea for common sense," in: Moss, L., J. Ginzburg, and M. de Rijk (eds.), *Logic, Language, and Computation, Vol. 2* (Stanford: CSLI).

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. New York: Oxford University Press.

von Fintel, K. (1998). "The presuppositon of subjunctive conditionals," in: Sauerland, U. and O. Percus (eds.), *The Interpretive Tract*, MITWPL **25**: 29–44.

von Fintel, K. (1999). "Counterfactuals in a dynamic context," in: Kenstowicz, M. (ed.), *Ken Hale: A Life in Language* (Cambridge: MIT Press).

von Fintel, K. (2003). "Would you believe it? The King of France is back!" in: Reimer, M. and A. Bezuidenhout (eds.), *Descriptions and Beyond* (New York: Oxford University Press).

Frank, A. (1997). *Context Dependence in Modal Constructions*. Ph.D. dissertation, Fakultät Philosophie: Universität Stuttgart.

Fuhrmann, A. (1999). "When hyperpropositions meet . . . ," *Journal of Philosophical Logic* **28**: 559–574.

Gillies, A. S. (2004). "Epistemic conditionals and conditional epistemics," *Noûs* **38**: 585–616.

Groenendijk, J. and M. Stokhof (1991). "Dynamic predicate logic," *Linguistics and Philosophy* **14**: 39–100.

Groenendijk, J. M. Stokhof, and F. Veltman (1996). "Coreference and modality," in: S. Lappin (ed.), *Handbook of Contemporary Semantic Theory* (Oxford: Blackwell).

Grove, A. “Two modellings for theory change,” *Journal of Philosophical Logic* **17**: 157–170.

Kratzer, A. (1977). “What ‘must’ and ‘can’ must and can mean,” *Linguistics and Philosophy* **1**: 337–355.

Kratzer, A. (1981a). “The notional category of modality,” in: Eijkmeyer, H.-J., and H. Rieser (eds.), *Words, Worlds, and Contexts* (Berlin: de Gruyter). Reprinted in: Portner, P and B. H. Partee (eds.) (2002), *Formal Semantics: The Essential Readings* (Oxford: Blackwell).

Kratzer, A. (1981b). “Partition and revision: The semantics of counterfactuals,” *Journal of Philosophical Logic*, **10**: 201–216.

Kratzer, A. (1989). “An investigation of the lumps of thought,” *Linguistics and Philosophy* **12**: 607–653.

Lewis, D. (1973). *Counterfactuals*. Cambridge: Harvard University Press.

Lewis, D. (1975). “Adverbs of quantification,” in: Keenan, E. L. (ed.), *Formal Semantics of Natural Language* (Cambridge: Cambridge University Press). Reprinted in: D. Lewis (1998), *Papers in Philosophical Logic* (Cambridge: Cambridge University Press).

Lewis, D. (1979). “Scorekeeping in a language game,” *Journal of Philosophical Logic* **8**: 339–359. Reprinted in: Lewis, D. (1983) *Philosophical Papers, Vol. I* (New York: Oxford University Press).

McCawley, J. (1996). “Conversational scorekeeping and the interpretation of conditional sentences,” in: Shibatani, M. and S. A. Thompson (eds.), *Grammatical Constructions* (Oxford: Clarendon Press).

Pollock, J. L. (1976). *Subjunctive Reasoning*. Dordrecht: D. Reidel.

Ramsey, F. P. (1929). “General propositions and causality,” reprinted in: Ramsey, F. P. (1978), *Foundations: Essays in Philosophy, Logic, Mathematics, and Economics* (London: Routledge and Kegan Paul).

Roberts, C. (1987). *Modal Subordination, Anaphora, and Distributivity*. Ph.D. dissertation, Department of Linguisitics: University of Massachusetts, Amherst.

Roberts, C. (1996). "Anaphora in intensional contexts," in: S. Lappin (ed.), *Handbook of Contemporary Semantic Theory* (Oxford: Blackwell).

Rott, H. and M. Pagnucco (1999). "Severe withdrawal (and recovery)," *Journal of Philosophical Logic* **28**: 501–547.

Sobel, J. H. (1970). "Utilitarianisms: simple and general," *Inquiry* **13**: 394–449.

Stalnaker, R. (1968). "A theory of conditionals," in N. Rescher (ed.), *Studies in Logical Theory* (Oxford: Basil Blackwell), pp.98–112.

Stalnaker, R. (1975). "Indicative conditionals," *Philosophia* **5**(3): 269–286.

Stalnaker, R. (1981). "A defense of conditional excluded middle," in: Harper, W., R. Stalnaker, and G. Pearce (eds.), *Ifs*. Dordrecht: D. Reidel.

Tichy, P. (1976). "A counterexample to the Stalnaker–Lewis analysis of counterfactuals," *Philosophical Studies* **29**: 271–273.

Veltman, F. (1976). "Prejudices, presuppositions and the theory of counterfactuals," in: Groenendijk, J. and M. Stokhof (eds.), *Amsterdam Papers in Formal Semantics, vol. I* (Amsterdam: Centrale Interfaculteit, Universiteit van Amsterdam).

Veltman, F. (1985). *Logics for Conditionals*. Ph.D. dissertation, Department of Philosophy: University of Amsterdam.

Veltman, F. (1996). "Defaults in update semantics," *Journal of Philosophical Logic* **25**: 221–261.

Veltman, F. (2005). "Making counterfactual assumptions," *Journal of Semantics* **22**: 159–180.