Heim Sequences and Why Most Unqualified
'Would'–Counterfactuals Are Not True

I. Introduction

Sophie is considering going to the parade where baseball player Pedro Martinez will be featured on a float. She decides not to go. The following counterfactual seems true:

(1) a. If Sophie had gone to the parade she would have seen Pedro

But now consider what would have happened if Sophie had gone to the parade but had been stuck behind someone tall. It seems that (1b), below, is true as well:

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

Sequences like (1), commonly referred to as *Sobel sequences*[1], famously motivated David Lewis (1973) to reject the once popular strict conditional semantics of counterfactuals in favor of the now classic *variably* strict conditional semantics. The strict conditional analysis treats a counterfactual as a material conditional embedded under a necessity operator, where the domain of worlds in the operator's scope is determined by the context. In symbols, if '>' represents the counterfactual connective, '⊃' the material conditional connective and '□' the necessity operator, the strict conditional account says that for any two propositions ϕ and ψ, $\phi > \psi =_{def} \Box(\phi \supset \psi)$. Call the worlds in the necessity operator's domain (in a context) the *accessible worlds*.

What happens when we evaluate (1a) and (1b) using the strict conditional semantics? If (1a) is (nonvacuously) true, then at all of the accessible worlds at which

[1] Lewis (1973) thanks J. Howard Sobel for bringing sequences like (1) to his attention.

[2] Lewis considers the possibility that counterfactuals could be "vague strict conditionals based on similarity, and that vagueness is resolved – the strictness is fixed – by very local context: the antecedent itself." But he

Sophie goes to the parade, she sees Pedro. If that is so then Sophie sees Pedro at all of the accessible worlds at which she goes to the parade and is stuck behind someone tall. But in that case, (1b) is false (unless there are no accessible worlds where Sophie goes to the parade and is stuck behind someone tall – in which case (1b) is vacuously true). The problem for the strict conditional analysis is that it entails that (1a) and (1b) cannot both be nonvacuously true in a fixed context.[2]

Following Robert Stalnaker (1968), Lewis's (1973) solution is to order the accessible worlds according to their similarity to the world of assessment (*w*), based on a particular, contextually determined similarity metric. Pictorially we can imagine this, as Lewis does, by picturing a system of spheres centered around *w*. Each sphere represents a degree of similarity (or 'closeness') to *w*. The worlds represented in the innermost sphere are the worlds most similar to *w*. Moving outward, worlds represented in each successively larger sphere are worlds that are successively less similar to *w.* The worlds represented in the region outside the system of spheres are the inaccessible worlds. For Lewis (but using my notation) a counterfactual A>C is non-vacuously true just in case there is an A&C–world closer to *w* than any A& Not–C world. It is vacuously true just in case there are no accessible A-worlds, and it is false otherwise.[3]

According to Lewis's variably strict conditional semantics described above, if all of the Sophie-goes-to-the-parade-worlds most similar to the actual world are worlds where

[2] Lewis considers the possibility that counterfactuals could be "vague strict conditionals based on similarity, and that vagueness is resolved – the strictness is fixed – by very local context: the antecedent itself." But he rejects this, saying that it "…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." (1973: 13) As we will see, von Fintel (2001) and Gillies (2007) endorse a picture very similar to the one Lewis is rejecting here.
[3] Stalnaker's (1968) semantics is very similar. For Stalnaker 'A>C' is (nonvacuously) true just in case C is true at *the closest* A-world.

Sophie sees Pedro, then (1a) is true.[4] But the truth of (1a) does not preclude the truth of (1b). (1b) is also true if the nearest worlds at which Sophie goes to the parade *and* is stuck behind someone tall are worlds where Sophie does not see Pedro. As long as all of the worlds at which Sophie goes to the parade but is stuck behind someone tall are less similar than the most similar worlds where she goes to the parade and sees Pedro, both counterfactuals can be nonvacuously true.

So far so good. Unlike the strict conditional semantics, the variably strict conditional semantics handles Sobel sequences very well. But there is a problem. The problem of Heim sequences[5], attributed to Irene Heim and first discussed by Kai von Fintel (2001) and Anthony Gillies (2007), is that if the order of (1a) and (1b) is reversed, it no longer seems that both counterfactuals are true:

(2) a. If Sophie had gone to the parade and been stuck behind someone tall, she would not have seen Pedro.
b. #But if Sophie had gone to the parade she would have seen Pedro.[6]

Given that both (1a) and (1b) seem true, the infelicity of (2b) is unexpected if the variably strict conditional analysis is correct. If the closest worlds where Sophie goes to the parade are worlds where she sees Pedro (as (1a) says), and if the closest worlds where she goes to the parade and is stuck behind someone tall are worlds where she does not see Pedro (as (1b) says), then, according to the variably strict conditional semantics, the counterfactuals in sequence (2) should both be true, regardless of which is uttered first.

[4] For ease of exposition I will generally speak as though the 'Limit Assumption', the assumption that there is always a set of most similar antecedent-worlds, holds. Lewis (1973) denies the Limit Assumption but nothing hangs on that here.
[5] Some previous authors have referred to sequences like (2) as *reverse Sobel sequences.* I follow Karen Lewis (forthcoming) in referring to them as *Heim sequences.*
[6] The "#" sign is used to indicate the seeming infelicity of the utterance.

This challenge for the classic model has been taken very seriously in the literature on counterfactuals. It has led some theorists (von Fintel (2001), Gillies (2007)) to reject the variably strict conditional semantics entirely, and instead endorse a variation of the original strict conditional account. Others (Ichikawa (2011), Karen Lewis (forthcoming)) have argued that the problem motivates a contextualist rendering of counterfactuals similar to contextualist accounts of knowledge or taste. And an entirely different kind of response comes from Sarah Moss (2008), who argues that, in fact, the classic semantics can handle Heim sequences like (2) just fine. On her view the infelicity of (2b) has a pragmatic explanation and should not be attributed to the counterfactual being *false*.

It is my contention that none of these reactions to the problem of Heim sequences is the right reaction. After showing why I think each of the proposals is inadequate, I will defend a novel way to make sense of the troublesome sequences. The solution I endorse avoids the problems faced by the alternative analyses. In addition, there is good independent reason to think that it is right. There is, however, a difficulty for my view: its truth entails that many ordinarily accepted counterfactuals are *not* true. I will argue that this (apparent) cost is an acceptable one. Before defending my own solution to the problem, however, we should take a brief look at the three extant proposals.

II. Von Fintel and Gillies: the Dynamic Semantic Solution

In response to the Heim sequence problem, von Fintel (2001) and Gillies (2007) have each (independently) argued that Lewis and Stalnaker were wrong to reject a strict conditional analysis in favor of the variably strict conditional analysis in the first place. Instead, they contend, the strict conditional semantics is basically correct – it just needs some tweaking. Von Fintel and Gillies defend a variation of the strict conditional semantics according to which, as part of its *meaning*, a counterfactual utterance 'A>C' updates the domain of worlds that it, and subsequent counterfactuals, quantify over: in particular, 'A>C' demands that there are accessible A-worlds in the domain.[7] These *dynamic semantic* analyses account for the infelicity of (2b) as follows. As part of its meaning, (2a) demands that there are at least some accessible worlds where Sophie goes to the parade and gets stuck behind someone tall. But if that is so then (2b) is false: it is not the case that Sophie sees Pedro at all accessible worlds at which she goes to the parade. However, although (2b) is false, (1a) need not be. Because (1a) is asserted prior to (1b), at the time of (1a)'s utterance there has been no demand that there be accessible worlds where Sophie goes to the parade and gets stuck behind someone tall. And (1b) need not be false, either. (1a) demands only that there are some accessible worlds where Sophie goes to the parade. Accommodating the (weaker) demands of (1b) requires that we bring into the domain worlds where Sophie gets stuck behind someone tall, and at these worlds, Sophie does not see Pedro. Thus, on this picture, (2a) and (2b) are semantically inconsistent, though (1a) and (1b) are not. This accounts for the felicity of sequence (1) and the infelicity of sequence (2). As we will see in the next section, however, there are good reasons to reject the dynamic semantic account.

III. Moss: the Pragmatic Solution

Moss (2012) has defended an entirely pragmatic way to account for the infelicity of Heim sequences. And, as she has shown, her solution is superior to the solution of Von

[7] The details regarding exactly how this occurs (and which distinguish von Fintel's account from Gillies's) are not important for my purposes here.

Fintel and Gillies in at least two important ways. On Moss's view, (2b) is *not* infelicitous because it is *false*, as on the dynamic semantic analysis. Rather, (2b) is infelicitous for the same basic reason that in general, when an assertion is made in a context in which there is a salient possibility that is such that (i) it cannot be ruled out by the speaker and (ii) it is incompatible with the assertion, the assertion is infelicitous. Since the speaker cannot rule out that if Sophie had gone to the parade she would have been stuck behind someone tall and so not seen Pedro, the possibility, when raised to salience by (2a), makes (2b) pragmatically infelicitous. (According to Moss (1a) does not similarly make (1b) infelicitous because that Sophie would have seen Pedro (had she gone to the parade) is intuitively no longer part of the common ground once (1b) has been uttered.[8])

Pragmatic infelicity of the kind Moss describes is a widely occurring phenomenon instantiated by a wide variety of different kinds of utterances, not just counterfactual ones. Here is one of Moss's examples. Suppose you and I are looking at the zebras at the zoo, and we have the following exchange:

(7) a. That animal was born with stripes.
b. But cleverly disguised mules are not born with stripes.

As Moss writes, "[t]his reply may be a non sequitur, perhaps even a little annoying. But otherwise there is nothing wrong with [the] reply." (2012: 567) Things change when we reverse the order of (a) and (b), however:[9]

[8] Moss claims that this asymmetry (between (1) and (2)) in whether the first proposition remains common ground once the second sentence of the sequence has been uttered is a ubiquitous kind of asymmetry. See Moss (2012: 570-571)

[9] There are ways to read (8) such that it sounds perfectly fine. The person uttering (8b) has the option to resist taking seriously the possibility made salient in (8a), i.e., that the animal could be a disguised mule with stripes (maybe the possibility seems too outrageous). As we will see shortly, Moss can account for this as well.

(8) a. Cleverly disguised mules are not born with stripes
b. #But that animal was born with stripes.

Moss concludes,

> So why is [(8)] bad, while [(7)] is okay? Here is one intuitive answer: in the above scenario, [(8b)] is infelicitous because [(8a)] raises the possibility that the caged animal is a cleverly disguised mule, and the speaker of [(8b)] cannot rule out this possibility. So [(8b)] is infelicitous because in the above scenario, it is an epistemically irresponsible thing to say. (2012: 568)

Moss gives two important considerations that rule in favor of her pragmatic explanation of the infelicity of Heim sequences over the semantic explanation given by von Fintel and Gillies. The first is that, as we have seen, her explanation is far more general. It accounts for a much wider range of data.

The second consideration is that Moss's solution can provide a natural explanation for why in some cases the second counterfactual of a Heim sequence does *not* sound infelicitous. And there are a lot of good examples of sequences that are like this. Moss gives the following helpful example (2012: 574):

> Suppose John and Mary are our mutual friends. John was going to ask Mary to marry him, but chickened out at the last minute. I know Mary much better than you do, and you ask me whether Mary might have said yes if John had proposed. I tell you that I swore to Mary that I would never actually tell anyone that information, which means that strictly speaking, I cannot answer your question. But I say that I will go so far as to tell you two facts:

(9) a. If John had proposed to Mary and she had said yes, he would have been really happy.
b. But if John had proposed, he would have been really unhappy.

Here the second counterfactual in the Heim sequence is felicitous, and Moss can explain why. The speaker can rule out that Mary would have said yes, had John asked (of course, this is exactly what the speaker is intending to communicate by asserting the two counterfactuals). And if the speaker can rule out that Mary would have said yes, then the case does not meet Moss's first criterion, stated above, for instantiating the kind of pragmatic infelicity instantiated by sequences like (2) and (8).

Moss's pragmatic explanation of the infelicity is very appealing. There are, however, some problems with it. Karen Lewis (forthcoming) has recently advanced a series of forceful objections to Moss's account. In section V of this paper I will offer a novel argument which aims to show that, at the very least, Moss's account cannot tell the full story. The most important difference between the dynamic semantic model and Moss's pragmatic solution is that according to the former (2b) is *false*. If Moss is right (2b) may very well be true, just as (8b) <But that animal was born with stripes> may be true, even if asserted after (the true) <Cleverly disguised mules are not born with stripes>. So if it were to turn out that the second counterfactual in a Heim sequence is *always* false when the sequence is infelicitous, this would tell strongly against Moss's proposal, which lacks the resources to account for why this would be so.[10] While I will not argue that the second sentence of an *infelicitous* Heim sequence is always false, I will argue that it is false in all of the interesting cases. Since Moss's solution cannot account for even this more limited claim, the limited claim's truth would suffice to show that an alternative explanation is needed. Before getting to my defense of the limited claim, however, we should consider one final proposal.

IV. Karen Lewis: the Contextualist Solution

[10] Note that if von Fintel and Gillies are right, the second counterfactual of the Heim sequence is not merely false when the sequence is infelicitous. Unless the defender of the dynamic semantic account can give a reason to think otherwise, it seems that on that view we should expect the second counterfactual to *always* be false. But this makes falsity far too ubiquitous, and indeed, is the basis of another objection to their semantics: Clearly Moss's (9b) <if John had proposed to Mary, he would have been really unhappy> need not be *false* even if the previously uttered (9a) <If John had proposed to Mary and she had said yes he would have been really happy> is true.

Karen Lewis ((2015), (forthcoming)) has a novel solution that evades the problems faced by the alternative accounts. Lewis argues that counterfactuals should be understood as on a par with context sensitive expressions like, for instance, automatic indexicals (*I*, *today*), gradable adjectives (*tall*, *rich*) and absolute gradable adjectives (*flat*). On her picture the semantic value of a counterfactual is sensitive to features of the conversational context such as e.g., the standards of precision and the salience (or non-salience) of possibilities. Formally, the difference between Karen Lewis's semantics and the Lewis-Stalnaker semantics is that while on the classic model possible worlds are ordered just according to how similar they are to the world of assessment, for Karen Lewis worlds are ordered by both similarity and *relevance*. "The only thing we need to add to [the Lewis-Stalnaker semantics] for the time being to get the account off the ground", she writes, "is that the similarity ordering or selection function is influenced by relevant salient possibilities, so that the closest worlds as determined by a given context include relevant salient possibilities [and exclude ignored, non-salient possibilities]" (forthcoming: 22). In a bit more detail:

> …the closest worlds are not just the most similar worlds, but…both similarity and relevance contribute to what counts as a closest world: some worlds that are most similar but aren't relevant are *not* among the closest worlds, and some worlds that are relevant but that are not the most similar are among the closest worlds. Picturesquely, this is how the ordering sources interact: begin with a Lewisian-like similarity metric (it need not be Lewis's *exactly* similarity metric). Relevance takes worlds that are among the most similar worlds and move them farther away, so that they are not among the closest worlds. It also takes worlds that are less similar – as long as they are similar enough (which is vague) – and moves them closer, so that they are among the closest worlds. (forthcoming: 25, Lewis's emphasis)

Her truth conditions are given below.

> For all contexts c, P $\Box\!\!\rightarrow$ Q is true at *w* in c iff all the closest P-worlds to *w* are Q-worlds, where closeness is a function of both similarity and relevance. (forthcoming: 25)

But what exactly is relevance? Karen Lewis characterizes it as an *objective* feature of the conversation, although by this she just means that “speakers are limited in how much they can affect what is relevant and irrelevant”. Not every possibility can be made relevant if brought to salience (i.e., not any less-similar world can become among the closest) and not every possibility can be made irrelevant by being ignored (i.e., some most-similar worlds will be among the closest no matter what). In other words, there are objective constraints on what must, can, and cannot be relevant in a context. It will be important, for what follows, for us to see what kinds of constraints Lewis has in mind. They include the following. “High probability (conditional on the antecedent) macroscopically—described outcomes are always relevant" and “…some possibilities are just determinately irrelevant, like ones that are really dissimilar to the actual world (given the antecedent – some antecedents clearly require more departure than others). (forthcoming: 25-26) In addition, the truth matters irrespective of the conversational participants’ beliefs. So, for instance, “if conversational participants (perhaps justifiably) think that Sophie is extremely shy and is very likely to cower at the back of the crowd at a parade, if the facts are such that they are wrong – Sophie is not shy in this way – the possibility that she has an unblocked view is relevant. Such possibilities cannot be legitimately ignored.” (forthcoming: 25-26)

Now that we have the basic picture let us see how it gets used to solve the Heim sequence problem. Take sequences (1) and (2) again:

(1) a. If Sophie had gone to the parade she would have seen Pedro.
b. If Sophie had gone to the parade and been stuck behind someone tall she would not have seen Pedro.

(2) a. If Sophie had gone to the parade and been stuck behind someone tall she would not have seen Pedro.
b. # But if Sophie had gone to the parade she would have seen Pedro.

Supposing that (1a), (1b) and (2a) are true, (2b) can still be false if the possibility raised to salience by (2a) – that is, that Sophie is both at the parade and stuck behind someone tall – is relevant in the context when (2b) is evaluated. If it is relevant then worlds where Sophie goes to the parade and is stuck behind someone tall become among the closest Sophie-goes-to-the-parade-worlds (even though they were not among the closest when (1a) was evaluated, which is why (1a) could be true). And if worlds where Sophie is stuck behind someone tall and so does not see Pedro are among the closest Sophie-goes-to-the-parade-worlds, then (2b) is false.

Karen Lewis's analysis avoids many of the difficulties faced by the alternative proposals.[11] For example, it can easily account for the felicity of felicitous Heim sequences: if the possibility raised to salience by the first counterfactual of the Heim sequence is not *relevant*, it does not impact the closeness ordering. Nevertheless, I think the contextualist solution is wrong. I now advance two arguments against it.[12]

Argument 1: Counterfactuals are more objective than Lewis's semantics would have it.

One immediate challenge for a contextualist semantics like Karen Lewis's is that counterfactuals like <if Sophie had gone to the parade she would have seen Pedro> seem to have determinate, invariant truth values once the "facts have been fixed", and the similarity

[11] Lewis discusses some ways that her account seems preferable to Moss's in her (forthcoming).
[12] For some additional arguments against contextualism about counterfactuals in general (i.e., not against Karen Lewis's contextualism in particular), see Hájek (ms).

ordering has been established.[13] It seems clear that the truthvalue of a statement like <Agnes is rich> or <Lester is tall> depends on the standards for richness, or for tallness in the context. But counterfactuals like (1a) seem importantly different. Intuitively, that Sophie would have seen Pedro had she gone to the parade is either determined by facts about the world or it isn't. Either way, this seems to be an objective question about the world – even if context helps determine *which* are the relevant facts. The question's answer intuitively does *not* depend on which possibilities happen to be salient in the context. And if that's right Karen Lewis's contextualism is wrong.

To see this more clearly let us compare discourses involving paradigm context sensitive expressions with discourses involving counterfactuals. When we say of a pool table that it is flat, or of a fifth grade child that he is tall, we are in general saying something about the table, or the fifth grader, relative to some standard or comparative class. If it is objectively true that the table is flat (or the fifth grader tall), it is so relative to the standard implicit in the conversational context. Indeed, as we will see in a moment, the speaker is usually able to make explicit the standards or domain or comparative class she has in mind when pressed.

But a counterfactual utterance is different. When I assert that <if Sophie had gone to the parade she would have seen Pedro>, I do not take myself to be asserting what would

[13] Of course, most agree that counterfactuals are context-sensitive in one important sense: which facts count toward the world similarity ordering depends upon the context. Karen Lewis distinguishes her own kind of context-sensitivity from this "ordinary" kind by saying that hers "is context-sensitivity after all the facts are fixed." (2015: 16). In other words, there is additional work for the context to do once the similarity function is (contextually) determined.

have happened relative to a standard of any kind. The difference is evident upon consideration of some examples:[14]

(10) a. He is tall!
  b. [#?] No he is not. Remember the NBA players we saw at the game last night?
  c. $_{\text{Ok}}$ Stop being a smart alec, you know that's not what I meant. I meant that he is tall for a boy his age.

(11) a. It is raining outside
  b. # No it is not, the sky is perfectly clear
  c. $_{\text{Ok}}$ Stop being a smart alec, you know that's not what I meant. I meant that it is raining where *I* am, in Tucson.

(12) a. The table is flat
  b. [#?] No it is not; if you look with a microscope you'll see unevenness…
  c. $_{\text{Ok}}$ What I meant was that for our purposes it is flat. It is flat enough to lay your drink on.

(13) a. All the beer is warm
  b. [#?] Well, not *all* the beer. Surely someone, somewhere has some chilled beer.
  c. $_{\text{ok}}$ That's not what I meant. I meant that all the beer in the house is warm.

(14) a. That man is rich.
  b. [#?] No he is not. Have you been to my neighborhood?
  c. $_{\text{Ok}}$ Okay, relative to my standards he is rich. Relative to your standards he is not.

Compare:

(15) a. If Sophie had gone to the parade she would have seen Pedro.
  b. $_{\text{Ok}}$ But if Sophie had gone to the parade and been stuck behind someone tall she wouldn't have seen Pedro.
  c. # That's not what I meant. I meant that for our purposes she would have seen Pedro./# Okay, relative to my standards she would have seen Pedro but relative to your standards she might not have./#That's not what I meant, I meant that if

[14] In her (2015) Lewis responds to what might appear to be a similar kind of objection to the one given here. The objection she anticipates and addresses involves retraction cases. According to that objection, "a contextualist theory, including the one I have been arguing for, predicts that speakers or other conversational participants should be able to judge as true what was said in an earlier context. Furthermore, they should be able to use propositional anaphora to call *true* the sentence that is predicted to be true by the theory in its context, even though the current context…is not one in which the same words express a truth." (2015: 18) The phenomenon she is referring to is different than the one I am discussing, and, as far as I can tell, nothing that she says in response to that objection helps against the objection I make here.

Sophie had gone to the parade she would have seen Pedro relative to standard (or restricted by domain) ___. [Where in the "___" goes any standard (or domain).]

The problematic replies in (15c) provide good evidence that the speaker who asserts (15a) takes herself to be speaking in absolute terms. She does not intend to assert that <if Sophie had gone to the parade she would have seen Pedro> is true *relative* to some standard.

Before moving on, I wish to briefly reply to a common objection. While most I have informally surveyed have agreed with me that each of the variations of (15c) are infelicitous, a few have had the intuition that in ordinary contexts the intended meaning of (15a) in fact is something like, *assuming things would have gone as expected,* or *given ordinary circumstances*, if Sophie had gone to the parade she would have seen Pedro. (Note that this claim is distinct from the claim that what is ordinarily meant by (15a) is that if Sophie had gone to the parade she would *probably* have seen Pedro. This latter possibility will be addressed in the next section of the paper.) It would be futile to attempt to deny that (15a) could ever be used to mean something like this. We do have good reason to deny, though, that this is the ordinary, or the usual meaning.

Suppose that unbeknownst to the speaker of (1a), Sophie in fact went to the parade, was stuck behind someone tall and did not see Pedro. Would we then think that the speaker had spoken falsely when he asserted (1a)? It seems to me that we ordinarily would. Indeed, we would expect the speaker to agree that she was wrong upon finding out what actually happened. This is even clearer, perhaps, when the counterfactual is futureregarding.[15] If, prior to the time of the parade, someone were to assert that if Sophie were to go to the parade she would see Pedro, we'd certainly think the speaker spoke falsely upon finding out that Sophie went to the parade and did not see Pedro. We'd think the speaker had spoken falsely, it seems, regardless of the reason that Sophie did not see Pedro. This is surprising if the proposal presently in consideration is right. For if the meaning of (1a) is that if Sophie had gone to the parade and nothing unexpected had happened then she would have seen Pedro, then its truth is *compatible* with Sophie actually going to the parade, getting stuck behind someone tall and not seeing Pedro. And if that is the case, we should not be inclined to think the utterance clearly false when it turns out that she was actually stuck behind someone tall and so did not see Pedro. The fact that we do think it false is good evidence that one of the above-suggested qualified meanings of (1a) is *not* (1a)'s meaning in ordinary contexts.

<u>Argument 2</u>: Karen Lewis draws the line in the wrong place.

Compare these two sequences:

(2) a. If Sophie had gone to the parade and been stuck behind someone tall, she would not have seen Pedro.
b. # But if Sophie had gone to the parade she would have seen Pedro.

You and I are talking about a particular (dry) match:

(16) a. If this match had been wet and struck it would not have lit
b. $_{\text{Ok}}$Okay, but if this match had been struck it would have lit.

Why is (2b) infelicitous but not (16b)? Both (2a) and (16a) make salient a possibility which is incompatible with the consequent of (2b) and (16b), respectively. Karen Lewis's theory predicts that Heim sequences will be felicitous if the possibility made salient by the

[15] I thank [omitted] for pointing this out.

antecedent of the first counterfactual in the sequence obtains only at worlds that are too far away to be relevant. But attempting to provide an explanation for why worlds at which the match is wet are too far away to be relevant in ordinary contexts brings out something peculiar. That is that, with one exception, it seems that no matter how similar we make the worlds where the match is wet, (16b) remains felicitous. For my purposes here it will be helpful to use discourse (16'), rather than (16), as my example. The only difference is that in (16') the counterfactuals make reference to a specified time, t.

Once again, you and I are talking about a particular match. The match is dry at t:

(16') a. If this match had been wet and struck at t it would not have lit
b. $_{\mathrm{Ok}}$ Okay, but if this match had been struck a t it would have lit.

(16b') is felicitous. As we've seen, Karen Lewis can account for this by maintaining that the worlds in which the match is wet and struck are too far away to be made relevant by the salience of wet-match possibilities. But why are the wet-match worlds too far away? Surely we can devise a scenario where the wet-match worlds are not so far away. Imagine that there are thousands of wet matches in the room, and the match referred to in (16') is the only dry match. Or suppose that at some time shortly before *t* there was a very high probability that the match in question would get wet. Despite the close call, the match remained dry. In both of these cases, it is still felicitous to utter (16b') in response to (16a') (I ask the reader to imagine these contexts and test the felicity of (16b') herself). Since we are referring to a particular dry match, it seems to make no difference how far away the worlds at which the match is wet.

As I indicated before, there is an exception, however. If *t* refers to the time the match first hits the matchbox, let us use t_1 to refer to the moment, some milliseconds later, when the match catches fire. If there is a chance – whether large or small – that if the striking had been initiated the (actually dry) match could have gotten wet *because of* the process of the striking (and prior to t_1), then it would no longer be felicitous to utter (16b') in response to (16a'). For in this case it is unknown if the match would have been wet or dry at t_1 had it been struck at t.

We have found a context, then, in which the worlds at which the match is wet and struck are not too far away to be made relevant. This is a context in which there is known to be a chance that the otherwise dry match could have become wet prior to the flame igniting if the striking had occurred. But notice that in this context we no longer have reason for thinking that the worlds in which the match is struck and wet are further away from the actual world than are the worlds in which the match is struck and dry. For although the match was actually dry at t_1, this cannot be relevant to the closeness ordering if the process of the match striking could have itself somehow been a cause of the match becoming wet. In general we should not count toward closeness match in factors that are causally dependent on the events required for the obtainment of the counterfactual's antecedent. We can see this very clearly if we imagine a scenario in which the objective chance that the match would have become wet had it been struck is very high. Say the chance is 1. In that case we certainly can't count the actual dryness of the match at t_1 as relevant for the similarity ordering, since <if this match had been struck at t it would *not* have lit> should come out as true. And for it to be true it must be the case that the worlds at which the match is wet and struck are at least as close as the worlds at which the match is dry and struck (indeed, they should be closer, since we'd also want to evaluate <if this match had been struck at t it would have been wet at t> as true.).

Here's a second case to illustrate the point. Suppose Smith tosses a fair, six-sided die and it lands on six. Now we evaluate the counterfactual <if Jones had tossed the die instead of Smith, it would have landed on six>. Is this counterfactual true? It is not. The reason it is not is that the worlds at which the die lands 1, 2, 3, 4 or 5 are equidistant from the actual world as is the world in which the die lands 6, despite the fact that the die landed on 6 in the actual world. Since a new die toss can cause a new toss outcome, we cannot count the actual outcome as a factor relevant to world-closeness. And just as the worlds at which Jones tosses the die and it lands on 2, say, are just as close to the actual world as the worlds at which Jones tosses the die and it lands 6, analogously, supposing that there is a chance that the match would have become wet prior to t_1 if struck, the worlds at which the match is struck and dry at t_1 are no closer than the worlds at which the match is struck and wet at t_1.[16]

I have argued that unless there is a chance that if the match were struck, it might first become wet (perhaps in the process of the striking), it appears that it does not matter how similar the wet-match world: it remains the case that (16b') is a felicitous response to (16a'). Furthermore, I have argued that in the event that there is a chance that if the match had been struck it could have become wet prior to t1, we no longer have reason to think that the met-match worlds are any further away than the dry-match worlds. This suggests

[16] In an effort to make the two cases analogous we may stipulate that the chance that if the match had been struck at t it would have gotten wet prior to t1 is 0.5 – although doing so is unnecessary: the chance that the match would have gotten wet is not relevant to the similarity ordering unless similarity is a function of likelihood. And there are good reasons to reject the idea that similarity is a function of likelihood (for one such reason see footnote 18).

that, contra Karen Lewis, the best explanation for the felicity of (16b') is *not* that worlds where the match is wet and struck are vastly dissimilar from the actual world. Nor does there seem to be a good candidate for a different constraint one could impose on the relevance function to rule out worlds at which the match is wet in the contexts in which (16b') is felicitous. None of the alternative constraints Karen Lewis discusses can do the job. So how do we rule out match-is-wet-and-struck-worlds to account for the felicity of (16b')? How about the old fashioned way: the nearest worlds where the match is struck and wet are simply *less* similar to the actual world than the nearest worlds where the match is struck and dry. And it is only the most similar antecedent worlds that are relevant for the evaluation of a counterfactual.

V. Deflating the Problem

The kinds of difficulties faced by the proposals on offer suggest a different solution to the puzzle. Compare our original Sobel and Heim sequences (sequences (1) and (2), respectively) with the sequences about the match:

(17) a. If this match had been struck at t it would have lit
b. But if this match had been wet and struck at t it would not have lit.

(16') a. If this match had been wet and struck at t it would not have lit
b. $_{\text{Ok}}$ Yes, but it's still the case that if this match had been struck at t it would have lit.

We just saw that we cannot account for the felicity of (16b') by insisting that the worlds where the match is wet and struck are extremely dissimilar from the actual world: (16b) is still felicitous when we fill in background details that make the wet-match-worlds not very dissimilar from the actual world, at all. But then what does explain why (16b), but not (2b), is felicitous? Here is an answer no one has given: the worlds where Sophie goes to the parade and is stuck behind someone tall are *just as similar* to the actual world as are the worlds where she goes to the parade and sees Pedro.[17] (Of course, it may be *unlikely* that Sophie would have been stuck behind someone tall, but similarity is not a function of likelihood.[18]) In contrast, while the worlds at which the match was wet and struck could be very similar to the actual world indeed, assuming there is not some chance that the match will become wet in the process of the striking (in which case (16b') is also infelicitous), they are *less* similar than the worlds at which the match was dry and struck (and lit).

If this is right, it means that (1a) <If Sophie had gone to the parade she would have seen Pedro> was never true (at least given the classic semantics). It was never true because it is not the case that all the closest antecedent worlds are consequent worlds. And if (1a) is not true then (2b) is not true, either. This explains (2b)'s infelicity. But if (1a) is not true, why did it seem true? I suggest that (1a) seemed true simply because the possibility that Sophie could have gone to the parade and been stuck behind someone tall (or been in the bathroom, or been distracted by her phone, etc.) did not occur to us. Someone needed to raise the possibility that, for instance, Sophie might have been stuck behind someone tall to make us aware of the possibility that she might not have seen Pedro.

[17] Karen Lewis also points out that "on a natural interpretation of a Lewis-style similarity ordering…the baseball player tripping and falling, and Sophie getting stuck behind someone tall at the parade count as among the closest worlds." (2015: 7-8) She uses this against the classic Lewisian semantics and in support of her contextualist account.

[18] There are a handful of philosophers who have argued for probabilistic truth conditions, e.g., a counterfactual is true if and only if the consequent is true at a sufficiently high proportion of the closest antecedent worlds (Bennett 2003) and, a counterfactual is true if and only if the conditional probability of the consequent given the antecedent is sufficiently high (Leitgeb (2012a), (2012b)). Even on these accounts, though, *more likely* does not in general mean *more similar* – it is only once a certain threshold of unlikeliness is reached that the (extremely unlikely) possibilities are ruled out. And there are reasons to reject these sorts of accounts, anyway. One simple reason is that they entail that (sufficiently) likely events are always the events that would have occurred, counterfactually. But this seems wrong. *Sometimes* (often, even) the extremely unlikely happens.

There is good independent reason to think that the worlds at which Sophie went to the parade but was stuck behind someone tall are just as close as the worlds at which she went to the parade and saw Pedro. There is, I suggest, a test that can be used to (defeasibly) determine if a particular set of antecedent-worlds are among the closest antecedent-worlds. Before introducing this test it will be helpful to consider one last case.

> A (ordinarily dressed) child is playing on the top of the jungle gym and trips and almost falls. Two observers have the following exchange:

(18) a. It is a good thing that child didn't fall. If she had she would have broken a bone!
b. (#?) Yes, but if the child had been wearing a full-body padded suit and fallen, the child would not have broken a bone.
c. $_{ok}$Uhh, okay….but if that child had fallen, she would have broken a bone./$_{ok}$Okay, but if *that* child had fallen, she would have broken a bone./$_{ok}$Okay, but if that child, *just as she is*, had fallen, she would have broken a bone.

This is a very clear-cut case. The worlds where the child is wearing the padded body suit and falls are less similar to the actual world than worlds where she is wearing no padded suit and falls (note that this is the case whether 9 out of the 10 kids in the park are wearing padded suits, or whether there had earlier been a 0.9 chance that, that kid would be one of those chosen to wear the suit. All that matters is that in fact, that child is not wearing it). The various appropriate replies to (18b) (shown in (18c)) are suggestive of a test, which can be dubbed the *Identify Something in the World Test*.

> **The "Identify Something in the World" Test**: If some world, w1, is less similar to the actual world than is another world, w2, then there should be something one can 'point to' in the actual world in virtue of which this is so.

It is possible for the person who asserted (18a) to reply to (18b) by emphasizing that he is not talking about a scenario in which the child is wearing a padded suit. He is talking about *that* child, just as she is right now. The speaker can point to something about the world — the child, and what the child is wearing – to make it clear that the worlds where the child is wearing a protective suit are not relevant in the conversation. They are less similar worlds, and they are less similar in virtue of the fact that the child is actually dressed normally.

Sequence (17) passes the Identify Something in the World test as well.

(17) a. [Pointing to a match]: If this match had been struck at t it would have lit
b. If this match had been wet and struck at t it would not have lit.

In response to (17b) the first speaker might say "that's true, but if this match had been struck it would have lit", and that would be just fine. But if she wanted to, she could also choose to emphasize that she is talking about *this* (dry) match. She can do that by emphasizing 'this', or by adding a 'just as is' clause: "that is true, but if this match *as it is now* had been struck, it would have lit". We can identify something in the world — in this case, the dryness of the match — to explain why worlds that are different in that respect are less close. Worlds where the match is not dry are less similar to the actual world in virtue of the fact that in the actual world the match is dry.[19]

What about sequences (1) and (2)? What can be pointed to, in the actual world, which plausibly makes it such that the worlds where Sophie is stuck behind someone tall are less similar to the actual world than worlds where she sees Pedro? I say that there is nothing to point to. In reply to (1b), the speaker who asserted (1a) *cannot* helpfully point to Sophie and say "no, I'm talking about *that* Sophie"; "no, I'm talking about Sophie as she actually is" (unless she is tall enough that the possibility that she is stuck behind someone

[19] The *Identify Something in the World* Test only works in one direction: as we saw earlier, two worlds can be equidistant from the actual world even if only one of the two shares a given feature with the actual world which can be "pointed" to.

tall can be ruled out – in which case (2b) would not be infelicitous). Nor can the speaker point to anything about the parade: "no, I'm talking about *that* parade (as it actually is)" (unless it is known that attendees stood in single file, in which case (2b) would not be infelicitous). If there is nothing that can be pointed to, we conclude, by the *Identify Something in the World* test, that the worlds where Sophie attends the parade but is stuck behind someone tall are no further away than the worlds where she attends the parade and sees Pedro.

I do not think it a coincidence that the sequences usually used to illustrate the problem of Heim sequences are sequences (1) and (2) (the counterfactuals in these sequences are not among those regularly used as examples in discussions of other counterfactual-related topics). This is not because I think that it has been widely recognized that worlds where Sophie is stuck behind someone tall are no further away than worlds where she sees Pedro. To the contrary, just about everyone in the literature has taken it for granted that (1a) is true: and for (1a) to be true, all nearest Sophie–goes–to–parade–worlds must be Sophie–sees–Pedro–worlds.[20] But it is the very fact that, as it happens, *not* all nearest Sophie–goes–to–parade–worlds are Sophie–sees–Pedro–worlds that accounts for the infelicity of (2b).

This solution to the puzzle, according to which neither (1a) nor (2b) is true, avoids the problems faced by the alternative proposals. As we've seen, it can explain why felicitous sequences like (9) and (16) are felicitous, while (2) is not, without committing to

[20] Karen Lewis is, as far as I know, the only one who recognizes the possibility that the worlds where Sophie is stuck behind someone tall could be as similar to the actual world as worlds where Sophie sees Pedro (forthcoming). But she still assumes that (1a) is true: on her view worlds that are among the most similar can be 'pushed back' if not relevant, so (1a) can be true even if the worlds where Sophie is stuck behind someone tall are no less similar than worlds where Sophie sees Pedro.

counterfactuals being context sensitive in the way that weather predicates or gradable adjectives are. It can also explain why (16b') is felicitous even if the worlds in which the match is wet are relatively close to the actual world (as long as they are not *as* close as the dry-match worlds). And, it can easily account for why it is only contexts in which the wet-match worlds appear to be just as close as the dry-match worlds that (16b') is infelicitous.

For Moss, the infelicity of counterfactuals like (2b) is accounted for on purely pragmatic grounds. For all that Moss's account says, (2b) can be true. But, as I argued a moment ago, (2b) is not true. It is not true because not all of the closest Sophie-goes-to-parade worlds are worlds where Sophie sees Pedro. This is not particular to this example. We've seen that in other cases (for example in the match case) when the worlds made salient by the first counterfactual in the Heim sequence (i.e. A&not-C worlds) are not as similar as the most similar A&C worlds, the sequence is felicitous.[21]

If I am right, philosophers have been wrong to take Heim sequences as a challenge to the classic semantics. When used correctly, the classic semantics either rules two identical counterfactuals both true or both not true in a fixed context, regardless of where each occurs in a sequence. But there is a serious difficulty for my proposal. Its truth entails that most ordinary counterfactuals are not true. Consider the match scenario, again. We know that worlds where the match is wet and struck are further away from the actual world than worlds where the match is dry and struck, and so these worlds are irrelevant to the

[21] This is not to say that the second counterfactual of a Heim sequence will *never* be infelicitous if it is true: I do not want to rule out the possibility that a counterfactual can sound infelicitous not because it is false but because the people in the conversation do not *know* if it is false. It is undoubtedly possible for there to be vague or ambiguous contexts in which it is impossible to discern which worlds count as among the most similar (in that context) and which do not. And in a context like that, I would expect Mossian-type pragmatic considerations to be relevant: if it cannot be ruled out that there are A&not-C-worlds as close as the nearest A&C-worlds, the counterfactual is unassertable, regardless of its truth-value.

evaluation of <if the match had been struck it would have lit>. But there may be other worlds, not yet salient, which are just as close to the actual world and where the match is struck but does not light. For instance, if someone had made salient the possibility that a gust of wind, or some indeterministic quantum event, could have occurred just as the match was being struck to make the match not light, this Heim sequence might be infelicitous:

(19) a. If the match had been struck but indeterministic quantum event *x* had occurred, it wouldn't have lit.
b. #But if that match had been struck it would have lit.

If it is a real (even if unlikely) possibility that quantum event *x* could have occurred and made the match not light, (19b) is infelicitous (in response to (19a), the person asserting (19b) should have at least weakened his assertion with something like 'probably'). And, as I have argued, if it is infelicitous, this is (generally[22]) because it is not true. It is not true because there are worlds just as close as the match–is–struck–and–lights–worlds where the match is struck and does not light. But if that is so then the ordinary counterfactual <if the match had been struck it would have lit> is not true, even when taken by itself. And indeed, nor are most other counterfactuals, since as Hawthorne (2005), Hájek (manuscript), Lewis (2015) and others have pointed out, there is usually some world among the closest antecedent worlds where the consequent does not obtain for some reason or other.

How serious of a cost is this? Many would take it to be very serious. Many have gone to great lengths to avoid accepting the conclusion that most ordinary counterfactuals

[22] See footnote 21.

are not true.[23] I find the strong aversion to the claim somewhat surprising. We should not be so averse to thinking that most ordinary counterfactuals are not true for at least two reasons. The first is that people are almost always willing to modify their counterfactual assertion with "it is likely that" or "it is probable that", when pressed. In most cases it is not even necessary to bring a problematic possibility to salience to get a speaker to qualify his utterance in this way. A simple request for greater precision is usually enough. For instance, if someone says to me "if the match had been struck it would have lit", it will in most cases be enough for me to reply with "do you mean to say that if the match had been struck it would have *definitely* lit?" The natural response to this, I submit, is to retreat to something weaker: "Well, what I mean to say is that if the match had been struck it would have very likely lit", for instance. And if the speaker is not willing to retreat in this way, it is probably just for lack of imagination. If I confront her with reasons why it is possible that the match would not have lit – for instance, a gust have wind could have blown through the open window at just that moment! – it is only reasonable that she qualify her claim, then. (Of course, if there is no legitimate way for the match to have been struck but not lit – that is, if there are no antecedent worlds, among the closest, where the match does not light — then the unqualified counterfactual is true.)

Speakers' willingness to qualify their counterfactual assertions in this manner when pressed could suggest one of a couple of different things. It could mean that 'A>C' should be taken to mean something weaker than that all nearest A-worlds are C-worlds. Maybe 'A>C' means that *most* nearest A–worlds are C–worlds, or that almost all nearest A–worlds

[23] See, for example, Lewis (1986), Bennett (2003), Williams (2008), Ichikawa (2011) and Lewis (2015), among many others.

are C–worlds. The second possibility is that 'A>C' *does* mean that all nearest A–worlds are C–worlds, but that when we use the counterfactual (unmodified) we are simply being imprecise: if speaking precisely we'd say that 'A>probably–C'.[24] There are several reasons to go with the second option over the first. One is that one's credence in a given counterfactual A>C is generally not the same as one's credence in its weaker counterpart. I might have credence of 0.8 that if Sophie had gone to the parade she would have seen Pedro, but credence very near 1.0 that if Sophie had gone to the parade she would have probably seen Pedro. This suggests that the two do not mean the same thing.[25] Another reason to deny that 'A>C' means that almost all A–worlds are C–worlds was already given in footnote 16: treating 'A>C' in this way commits us to accepting that any event sufficiently likely to have happened had the antecedent obtained, *would* have happened, had the antecedent obtained (where by "sufficiently likely" I mean above the threshold to make the counterfactual true on this view, whatever that threshold is). But since the extremely likely does not always happen in the actual world, there is little justification for thinking that the extremely likely would have always happened counterfactually.

A much better explanation for why speakers are usually willing to retreat to a more modest "A>*probably*–C" when pressed is that when in a thoughtful mood, speakers can come to realize that "A>C" is just too strong. If the speaker wants to communicate what she (can, upon reflection, come to realize she) means more precisely, she must say something weaker. The fact that speakers are usually happy to admit that they meant

[24] Hájek endorses something like this in his (ms).

[25] Note that if David Lewis's semantics is right, my credence in <If Sophie had gone to the parade she would have seen Pedro> should actually probably be much lower than 0.8, since it seems quite likely that there is at least one nearest Sophie-goes-to-the-parade-world where Sophie does not see Pedro (in which case, on David Lewis's semantics, the counterfactual is false). Stalnaker (1981) and Edgington (1995), (2014) have objected to the Lewisian semantics on these sorts of grounds.

something a bit weaker than what is captured by '*all* nearest A–worlds are C–worlds', is one reason not to be reluctant to accept the possibility that most unqualified counterfactuals are not true. A second reason not to be averse to the possibility is that imprecision is utterly ubiquitous in ordinary language.[26] When I say that I am 5'6" tall I do not mean that I am precisely 5'6'' tall. I am speaking loosely. What I really mean, which I will readily admit to if pressed, is that my height is somewhere in the vicinity of 5'6''. I am *about* 5'6'' tall. To take another ordinary example, suppose that someone is baking some bread (in addition to cooking some other food). Jasper walks in, smells the bread and asserts "that is the smell of bread". This gets the point across but is probably not quite right. Bread likely smells subtly different than what Jasper perceives. What Jasper smells is actually a mixture of bread and a host of other things combined. And what of the chemist who says that water is H_2O? Does she say something true? Maybe not, technically. Here is Paul Teller (forthcoming):

> Water is never absolutely 100% a collection of H_20 molecules. There are always some impurities and some dissociated molecules. When one closely approaches the (in practice unobtainable) goal of 100% H_20 the properties of the substance become significantly different from the water of familiar experience. So, more carefully, water is MOSTLY H_20. (Teller's emphasis)

The examples are endless. It is not so radical to hold that people regularly speak loosely when asserting counterfactuals if people regularly speak loosely in general.

In addition to not always speaking precisely we often speak more confidently than what might reasonably be called for. A large proportion of what we say can, and perhaps should, be appropriately weakened with "probably": "He's at the park" can be "he's probably at the park" (he said he was going to the park, but he could have lied or changed his mind); "the car is in the lot" can be "the car is probably in the lot" (it could have been

[26] On this see Teller (2011), Braun and Sider (2007), and Elgin (2004).

towed or stolen), etc. Frequently we leave the 'probably' out although we'd usually be happy to put it in if pressed. It is no surprise, then, that we do the same when speaking counterfactually.[27]

But why should people regularly speak loosely when uttering counterfactuals if to speak a truth one must be more precise? Why not be a bit more careful? For many of the same reasons, I suggest, that Jasper is not more careful when he asserts that the smell he smells is the smell of bread. For one, Jasper might not in that moment recognize that there are likely to be other smells adulterating the otherwise pure bread smell. And even if he were to realize that bread probably does not smell *exactly* like *that*, he would risk misleading his listeners if he were to say something more precise. If, just to be safe, Jasper were to say "that is approximately the smell of bread" or "that is a mixture of smells, the most salient of which is bread", he would risk implicating that he knows more than he does. For instance, he might implicate that he recognizes other smells that are presently diluting the bread smell, or, that he is able to distinguish between pure-bread smell and impure-bread smell. Yet (we can suppose) neither of these things is true. The irony is that if he had expressed something closer to the truth in the first place – e.g., that what he smells is approximately the smell of bread – he would have risked implicating falsehoods.

[27] Of course the difference between non-counterfactual unqualified assertions and counterfactual unqualified assertions is that the former are still *true* if the speaker is correct (even if he spoke too confidently) whereas if I am right, counterfactual utterances are usually *not* true if not appropriately qualified. But this is just a consequence of the kind of strange creatures counterfactuals are. In the case of non-counterfactual propositions like "he is in the park", the world (or something in it) makes the proposition true if he is actually in the park, even if the speaker was not in an epistemic position to appropriately assert the sentence. In contrast, there is nothing in the actual world to make the unqualified counterfactual true if, given the way the world is, there are multiple, equally good candidates for how it could have been. But given that ordinary speakers are not in general privy to these differences, we should not expect them to be more concerned about not qualifying their counterfactual assertions than they are about not qualifying their non-counterfactual assertions.

The same is true for the person who asserts a counterfactual. If I were to get into the habit of always adding "probably" before any counterfactual consequent, I would likely inadvertently communicate much that I do not mean. If I were to say, "if that match had been struck it probably would have lit" my listeners might assume, and *not* unreasonably, that I had some particular reason to think that the match may not have lit if struck. Maybe I know something about the match, or its environment, that they do not know. My listeners are likely to assume that I take the odds that the match would have lit, had it been struck, to be lower than I actually take the odds to be. Since speakers generally do not weaken their counterfactual – or, for that matter, their non-counterfactual – assertions in this way, to do so could suggest some degree of uncertainty that may not in fact be felt. For these reasons, and because it can be tiresome and unnecessary for speakers to be precise, and because the unlikely, unexpected ways one's utterance could be made false are often not at the forefront of one's mind when one speaks, it is not at all surprising that 'A>C' is regularly said in place of 'A>probably-C'. And if it is something slightly different than A>C that is (upon reflection) intended, it should not so much matter if A>C is not true.

VI. Conclusion

I have argued that we should not be averse to accepting that most unqualified counterfactuals are not true. They are close to something that is true, and that is good enough. As for the Heim sequence "problem", it is not really a problem at all. Once we are willing to accept that most unqualified counterfactuals are not true, we can recognize infelicitous Heim sequences for what they are: evidence that we've come across a counterfactual that should be qualified. If this is right, there is no need to revise or reject the classic semantics in response to Heim sequences. Nevertheless, it might prove worthwhile to shift some attention away from trying to understand unqualified counterfactuals first and foremost, and toward trying to advance our understanding of the qualified ones.[28]

Work Cited

Barker, S. (1999), "Counterfactuals, Probabilistic Counterfactuals, and Causation," *Mind* 108: 427-469.

Bennett, J. (2003), *A Philosophical Guide to Counterfactuals*, Oxford University Press.

Braun, D. & Sider, T. (2007) "Vague, So Untrue," *Nous* 41 (2): 133-156.

Edgington, D. (1995) "On Conditionals," *Mind*, 104: 235-329.

Edgington, D. (2014) "Indicative Conditionals", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2014/entries/conditionals/>.

Elgin, C. (2004) "True Enough", *Philosophical Issues*, 14: 113-121.

Gillies, T. (2007) "Counterfactual Scorekeeping," *Linguistics and Philosophy*, vol. 30: 329-360.

Hájek, A. (manuscript), *Most Counterfactuals are False*. ANU, monograph in progress.

Hawthorne, J. (2004) *Knowledge and Lotteries*, Oxford: Clarendon Press.

Hawthorne, J. (2005) "Chance and Counterfactuals", *Philosophy and Phenomenological Research*, Vol. LXX, No. 2.

Ichikawa, J. (2011) "Quantifiers, Knowledge, and Counterfactuals," *Philosophy and Phenomenological Research*, Vol. LXXXII No. 2.

[28] There is much work to be done if we are to understand the semantics of probabilistic counterfactuals. For an introduction to some of the unsolved puzzles regarding how counterfactuals interact with probability operators, see Barker (1999).

Leitgeb, H. (2012a) A Probabilistic Semantics for Counterfactuals: Part A," *The Review of Symbolic Logic* 5: 1, xx-xx.

Leitgeb, H. (2012b) A Probabilistic Semantics for Counterfactuals: Part B," *The Review of Symbolic Logic* 5: 1, xx-xx.

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

Lewis, D. (1986) Counterfactual Dependence and Time's Arrow. In *Philosophical Papers Volume II*, chap. 17. Oxford University Press.

Lewis, K. (2015). "Elusive Counterfactuals", *Nous* 50 (2): 286-313.

Lewis, K. (forthcoming) "Counterfactual Discourse In Context", *Nous*.

Moss, S. (2012), "On the Pragmatics of Counterfactuals", *Nous* 46:3, 561-586.

Stalnaker, R. (1968). "A Theory of Conditionals," In Harper et al. (1981), 41-55.

Stalnaker, R. (1981). "A Defense of Conditional Excluded Middle," in Harper, Stalnaker and Pearce (eds.), pp. 87-104.

Teller, P. (2011), "Two Models of Truth,", *Analysis* 71:3, 465-472.

Teller, P. (forthcoming), "Truth and Fiction in Science," *Science et Avenir.*

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

Williams, R. (2008) "Chances, Counterfactuals and Similarity," *Philosophy and Phenomenological Research* 77(2). 385-420.