CORE
Metadata, citation and similar papers at core.ac.uk
Provided by PhilPapers

**This is the penultimate draft.
For citations please consult the published version.**

To appear in *The Routledge Handbook of Epistemic Contextualism*

Counterfactuals and Knowledge

Karen S. Lewis

1 Introduction

The standard semantics for counterfactuals comes from Lewis (1973, 1986), Stalnaker (1968, 1981) and Kratzer (1977, 1981). Abstracting away from differences between the accounts, the basic semantics is as follows:

> A would-counterfactual P$\Box\!\!\rightarrow$Q is true (at w) iff all the closest P-worlds (to w) are Q-worlds.

Traditionally, the semantics for counterfactuals is thought to have a limited kind of context-sensitivity. That is, most of the time, there is a consistent way in which the closest P-world(s) are selected (relative to the world of evaluation); what counts as the closest P-world does not vary based on the conversational context. There are some exceptions to this. Lewis discusses Quine's famous case of Caesar in Korea, arguing that in some conversational contexts (1) is true, while in other contexts (2) is true, and this depends on what facts are being held fixed in the context:

(1) If Caesar had been in command in Korea, he would have used the atom bomb.

(2) If Caesar had been in command in Korea, he would have used catapults.

Stalnaker (1981) endorses the idea that in some contexts, different worlds count as equally similar due to negligible differences (whereas those differences might not be negligible in other contexts). For example, in a case in which we have a line in the margin of a book that is actually just less than an inch, and are considering counterfactuals about how long it would be if it was more than an inch long, worlds in which the line is a little more than an inch up to worlds where it is 2 or 3 inches might count as equally similar (suppose the margin is 3 inches wide).

Aside from these sorts of considerations, counterfactuals on these views are not deeply contextualist; people generally speak of their truth values absolutely, not relative to conversational context. But there are several puzzles that plague this basically invariantist semantics for counterfactuals; ones that bear striking similarity to the sort of puzzles used to motivate contextualism for 'know' and other terms. These have motivated Ichikawa (2011) and Lewis (2016, Forthcoming) to argue for a thoroughly contextualist semantics for counterfactuals. In §2, I explore the various motivations for counterfactual contextualism. In §3, I discuss and compare both my and Ichikawa's versions of counterfactual contextualism. Finally, in §4, I examine the relationship between contextualism for counterfactuals and for knowledge, arguing that though a close relationship between a contextualist semantics for counterfactuals and one knowledge is elegant, it faces problems.

2 Motivations for counterfactual contextualism

Counterfactuals enter into several puzzles that, if accepted at face value, appear to have the power to undermine the truth of nearly all contingent counterfactuals. Call this the problem of *counterfactual skepticism*. Most people generally believe that there are true contingent counterfactuals (in fact, lots of them), in that they are accepted as true by native speakers in ordinary conversation, and used theoretically in philosophy, psychology, history, artificial intelligence, and other disciplines. One of the most straightforward ways to resolve this tension is to adopt some version of a contextualist semantics for counterfactuals.[1]

[1]For someone who accepts that most counterfactuals are false and adopts an error theory to explain ordinary judgments, see Hájek (ms). For a pragmatic explanation of the

2.1 Puzzle 1: Mights and woulds

Consider an ordinary contingent counterfactual, one we are normally certain is true. For example, suppose I was holding my favorite mug just now as I stood in the middle of the kitchen. I did not drop it, but we think it true that:

(3) If I had dropped my mug, it would have fallen to the kitchen floor.

A skeptic might come along and warn us *not so fast. Isn't it possible*, she asks, *that I might have very swiftly caught the mug before it reached the floor*? Though I'm not known for my coordination, this is not beyond the realm of what I am physically capable of doing, and it seems to support the truth of the might-counterfactual (4):

(4) If I had dropped my mug, I might have deftly caught it before it fell to the floor.

This directly supports the truth of (5):

(5) If I had dropped my mug, it might not have fallen to the kitchen floor.

To borrow a term from DeRose (1999), (5) inescapably clashes with (3). In fact, it sounds very much like a contradiction.

(6) # If I had dropped my mug, it might not have fallen to the kitchen floor; if I had dropped my mug, it would have fallen to the kitchen floor.

The problem generalizes when we take into consideration the fact that our best physics seems to support indeterministic laws. If indeterministic interpretations of quantum mechanics are correct, then there is some chance (albeit very, very small) of almost anything happening, such as my mug quantum tunneling to China, which supports the truth of (7):

(7) If I had dropped my mug, it might have quantum tunneled to China (and so not fallen to the floor).

---

data, see Moss (2013).

This counterfactual also supports the truth of (5), and so we are again faced with the inescapable clash of (6).

Even if it turns out that deterministic statistical mechanics is the right physics, the problem remains. For the vast majority of counterfactual conditionals, and for virtually all ordinary counterfactual conditionals, the antecedent is underdescribed in terms of the microphysical detail by which it occurs. There are very many precise ways in which I could have dropped my mug (the exact position of my hand, the exact position of the mug in my hand, the strength with which I dropped it, etc.). Statistical mechanics tells us that while the vast majority of these initial conditions lead to the expected outcome, i.e., the mug falling to the floor, at least some initial condition is such that it macroscopically looks just like the ones in which things go normally, but things don't go normally, for example, my mug flies sideways and lands safely on the counter. If statistical mechanics is right, then it is true that:

(8) If I had dropped my mug, it might have flown sideways and landed safely on the counter.

And again, we have the same clash of woulds and mights. It should be pretty clear that the above sorts of considerations generalize to all or almost all contingent counterfactuals without explicit mention of probability in the consequent.

These sorts of undermining might-counterfactuals don't just come about from extremely low probability events like quantum tunneling or rare feats of amazing coordination. They can often come about when dealing with much more ordinary, merely somewhat low probability events. For example, suppose I had a party to which you were invited, but did not come. Much to your dismay, you missed an extraordinary party. Shockingly, Lady Gaga showed up for about 20 minutes! I tell you that you really should have accepted my invitation, as:

(9) If you had come to the party, you would have seen Lady Gaga.

This is the sort of counterfactual that is commonly accepted as true in conversation. But the skeptic need not look very far to find undermining might-counterfactuals, for it also seems true that:

(10) If you had come to the party, you might have been sick to your stomach and so in the bathroom the whole time Lady Gaga was there.

Or, given your disposition for getting lost in philosophical conversation:

(11) If you had come the party, you might have been out on the deck lost in philosophical conversation and not noticed the commotion inside when Lady Gaga was there.

Such examples are also easy to multiply.

As Hájek (ms) points out, we need not invoke might-counterfactuals to present this sort of problem. Invoking chance (in the case of indeterminism) or indeterminacy (in the case of determinism) directly is enough — as Hájek puts it, *chanciness undermines wouldiness* (or *indeterminacy undermines wouldiness*). Merely pointing out that the dropping of a mug is a chancy event, in that it could turn out that the mug lands on the floor, but it could also turn out that it quantum tunnels to China or is caught in a swift act of coordination is enough, it seems, to undermine our confidence in the truth of counterfactuals like (3). (The same goes for the party case.)

2.2 Puzzle 2: The similarity ordering

The clash between woulds and mights is not the only way to get counterfactual skepticism off the ground. On a Lewisian-style similarity ordering, many of the undermining worlds are among the closest worlds. Lewis's system of weights is as follows:

1. It is of the first importance to avoid big, widespread, diverse violations of law.

2. It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.

3. It is of the third importance to avoid even small, localized, simple violations of law.

4. It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly. ((Lewis, 1986, 47-8))

Provided the antecedent is not true at the actual world, the closest worlds are generally an exact match to the actual world until some time not long

before the antecedent occurs, at which point there is a small violation of law (e.g., a particle in a different place, a neuron fires that didn't actually fire) that brings the antecedent about. Then the laws of the actual world do what they will, and if they bring about the consequent in all the closest worlds, the would-counterfactual is true. If the laws are the indeterministic ones of quantum mechanics, the problem is the clearest. By hypothesis, the very same history, up to and including the same small miracle that brings about the antecedent, will lead to, say, a mug falling to the floor in one world and tunneling to China in another. This is just what it means for the actual laws to be the indeterministic laws of quantum mechanics. But even if the laws are actually the deterministic laws of statistical mechanics, the problem persists. This is because the antecedent is underdescribed, and so there are many equally good small miracles, in the sense that they are equally small violations of the law that will bring about the antecedent. In other words, there is indeterminacy in how the antecedent is brought about. Provided that one of these equally good small miracles brings about the initial conditions that lead to weird things happening — and I don't see why it wouldn't — worlds in which the mug flies sideways and lands safely on the counter are also equally close. Even if one does not ascribe to Lewis's particular system of weights, any natural interpretation of a similarity ordering in the spirit of Lewis, i.e., ones in which the closest worlds are ones in which minor changes are made to incorporate the antecedent without contradiction and nothing more, will predict that these are among the closest worlds.[2] This is not to say that it is impossible to introduce a different similarity ordering; this is just the strategy of Lewis (1986) in introducing the notion of a quasi-miracle and Williams (2008) in invoking typicality. Changing the similarity relation in some way like this is one possible response to the puzzle.[3]

What about the more ordinary cases, like my deftly catching the mug before it falls or your being out on the deck and missing Lady Gaga at my party? Arguably, a natural similarity ordering is going to count some of these

[2]This is equally a problem for Stalnaker, who endorses the uniqueness assumption in his formal semantics, i.e., that there is a unique closest world. In reality, the application of the semantics is often faced with indeterminacy, and ties for equal closeness are represented by the selection function potentially selecting different worlds on different precisifications. For Stalnaker, the problem of counterfactual skepticism as presented by this puzzle is not that most counterfactuals are false, but that most counterfactuals are indeterminate.

[3]For arguments against these views see Hawthorne (2005) (against quasi-miracles) and Lewis (2016) (against both).

among more distant worlds, but some among the closest worlds, depending on the case and the facts about the actual world. By hypothesis, I am not a terribly coordinated person. So worlds in which I catch the mug before it falls to the floor are probably not among the closest worlds.[4] Here is a case where the pressure from might-counterfactuals and chanciness comes apart from the pressure from similarity. But other cases are not so clear. Suppose, as I did in the above case, that like most philosophers, you are prone to getting embroiled in philosophical debate and losing track of your surroundings. Given that you weren't at the party, there is no fact of the matter regarding exactly where you would have been standing had you been there at the time Lady Gaga arrived (let's also assume you don't have any peculiarities like always standing in the same spot at parties or always next to the same person). There are many ways in which you could have been present at the party, given that there were many people in the kitchen, many in the living room, and many out on the deck. Different small miracles lead you to be at the party in slightly different ways, and so standing at slightly different spots. (Or, if this example doesn't convince you, there are different times at which you arrive at and leave from the party. Some of these times do not have you overlap with Lady Gaga's surprise appearance.) So even some of the ordinary cases seem to be among the closest worlds on a natural similarity ordering. If this is the case, again we are threatened with counterfactual skepticism.

2.3 Puzzle 3: Clashing would-counterfactuals

Counterfactual skepticism can also be motivated by a third kind of puzzle, one involving only would-counterfactuals, as Ichikawa (2011) presents it. Consider the case of the party again. You didn't come to my party, and so you missed Lady Gaga's surprise appearance. You regret not coming because:

(12) If you had come to the party, you would have seen Lady Gaga.

But of course, it is also true that:

[4]Then again, maybe not. Perhaps the fact that I am uncoordinated just means there are not *many* worlds among the closest in which I catch the mug, if we think of skill as something like corresponding to how many lottery tickets one has. Thanks to Jonathan Ichikawa (p.c.) for suggesting this. In any case, if this is right, it is only more fuel for counterfactual skepticism and the contextualist solution.

(13) If you had come to the party and been distracted by philosophical conversation out on the deck, you would not have seen Lady Gaga.

And this also seems to undermine our confidence in (12):

(14) # If you had come to the party and been distracted by philosophical conversation out on the deck, you would not have seen Lady Gaga. But of course, if you had come to the party, you would have seen Lady Gaga.

This data was first noted in publication by von Fintel (2001), who attributes the observation to Irene Heim in a seminar at MIT. Essentially, the observation is that conditionals like (12) and (13) form a consistent sequence, often called a *Sobel sequence*. But it seems they are only consistent when they are in that order. Reverse the order, and it sounds inconsistent. Both von Fintel and later Gillies (2007) take this data to support the need for a dynamic semantics for counterfactual conditionals.[5] But we can also take this data, as Ichikawa does, as a skeptical puzzle and a motivation for a contextualist account of counterfactuals.[6]

There have been various solutions to these three puzzles proposed in the literature. Hájek (ms) embraces counterfactual skepticism and proposes an error theory to account for our ordinary judgments. DeRose (1999) proffers a solution to the inescapable clash between woulds and mights by adopting a Stalnakerian semantics in which all might-counterfactuals are would-counterfactuals with a wide scope epistemic possibility operator ranging over them, and the clash is explained as pragmatic. Lewis (1986) and Williams (2008) aim to solve the problem from similarity, altering the similarity ordering to relegate the pesky worlds to more distant realms. Gillies (2007) and von Fintel (2001), and Moss (2012) give dynamic semantic and pragmatic accounts, respectively, of the reverse Sobel sequence data (Gillies also offers an account for the clash between mights and woulds). Putting aside other issues with each of these views, one major problem is that they each address only the puzzle they are designed to account for; the solutions cannot be extended to the other puzzles (with the exception of Hájek's error theory, which is an embrace of counterfactual skepticism rather than a solution to it, and

[5]See Moss (2012) for a pragmatic account of this data that maintains the traditional Lewis-Stalnaker semantics.

[6]In fact, I agree with Ichikawa in that I think a contextualist solution is the best account of this data (see Lewis (Forthcoming) for a full defense of this claim and the account).

Gillies' view, which applies to two out three). It is possible that each puzzle represents a distinct phenomenon that warrants a distinct explanation. But at least on the surface, taken together, these seem to be related phenomena. A contextualist semantics for counterfactuals can explain them all, and this is certainly a theoretical virtue.

3 Counterfactual Contextualism

Contextualism for counterfactuals, like contextualism for knowledge, can take many forms. I'll describe my preferred view, which is one version of the view I defend in Lewis (2016). In essence, it is a Lewisian-style variably strict conditional semantics that takes both similarity and relevance to contribute to the closeness ordering. I will then briefly compare it to Ichikawa's 'all cases' version of contextualism.

There are two central components to my preferred version of counterfactual contextualism. First, counterfactuals are sensitive not just to the most similar worlds, but to the worlds that are relevant given the conversational context. Second, the pragmatic effect of both might- and would-counterfactuals can be to expand what possibilities are relevant in the context. Beginning with the semantic component, the truth-conditions of a would-counterfactual are as follows:

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

Mights are the duals of woulds (P$\Diamond\!\!\rightarrow$Q $=_{def}$ $\neg$(P $\Box\!\!\rightarrow\neg$Q)) so:

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

Both similarity and relevance contribute to what counts as a closest world: worlds that are most similar might not be among the closest because they are simply not relevant to the conversation, and worlds that are not among the most similar might be among the closest because they are relevant in the context. Picturesquely speaking, relevance can take worlds that are among

the most similar and move them farther away, and take worlds that are less similar, though not *too* dissimilar, and move them to the closest sphere.

What notion of conversational relevance is at play here? It is relevance both to the purpose of asserting a counterfactual and to conversational purposes more generally. Counterfactuals are often used for making predictions, expressing regret, or making dispositional claims. This means that the actual world is always relevant if it is an antecedent world. (This corresponds to Ichikawa's invocation of Lewis's *rule of actuality*.) High probability outcomes (conditional on the antecedent) are also always relevant. It doesn't matter whether the conversational participants are aware of the probabilities or the facts that make them so; high probability outcomes cannot legitimately be ignored. Conversely, low probability outcomes can be ignored when they are not otherwise relevant to conversational purposes. *How* high counts as high (or how low counts as low) depends on the standards of precision operant in the conversation. Conversations about scientific experiments will have different standards from casual conversation, and conversations about quantum physics experiments will have different standards from conversations about biology experiments. Conversational participants have limited control over these measures of relevance. Once the conversational purposes and standards of precision are in play, they cannot legitimately ignore the actual world or high probability outcomes, even if they are completely ignorant of the actual world or the relevant probabilities.

The conversational participants can shift the conversational purposes or standards of precision by explicitly or indirectly introducing previously unconsidered possibilities. Two (of many) ways of so doing are to utter a might-counterfactual or would-counterfactual that includes a previously unconsidered possibility. For example, suppose we are engaged in casual conversation and have not yet considered quantum physics. In this case, (3), repeated here as (15), is true when initially uttered in the conversation:

(15) If I had dropped my mug, it would have fallen to the kitchen floor.

But if you say (16) or (17), you introduce possibilities previously legitimately ignored in the conversation:

(16) If you had dropped your mug, it might have quantum tunneled to China (and so not fallen to the floor).

(17) If you had dropped your mug and it had quantum tunneled to China, it would not have fallen to the floor.

In the context in which (16) is asserted, it is technically false, since we have been legitimately ignoring worlds in which quantum events occur (and so none of the closest worlds are quantum worlds). But we tend to interpret speakers charitably, and it is clear that you are trying to shift conversational standards. So the context is (at least temporarily) shifted to include quantum worlds, making (16) true in the newly accommodated context. (17) is straightforwardly true, since even in the casual context the closest (i.e. most similar and relevant) mug-dropping worlds in which quantum tunneling occurs are presumably worlds in which the cup does not fall to my kitchen floor. Whether such assertions *permanently* change the conversational context depends on the conversational participants. I can stand my ground and refuse to accommodate, maintaining that quantum outcomes are simply irrelevant to what we are talking about. Or I can acquiesce, accepting the shift in conversational standards, in which case (15) is not true in the new context. (If the conversational participants disagree about whether accommodation should take place, disagreement or negotiation can ensue.) This is similar in spirit to Ichikawa's version of Lewis's *rule of attention*. Something being salient is not enough for it to become relevant; it must be taken seriously. To this I'd also like to add another point. Another way in which the conversational context cannot be permanently changed by making something salient is if the possibility made salient is very dissimilar; perhaps one way to understand this rule is that when considering counterfactuals one should never take seriously very dissimilar possibilities unless the antecedent requires it. So, for example, suppose, in a casual context, I am holding a reliable dry match that I never strike and truly say:

(18) If I had struck this match just now, it would have lit.

(19) is also true, since the most relevant similar worlds in which the match is soaking wet and struck, it does not light:

(19) If I had struck this match just now and it had been soaked overnight, it wouldn't have lit.

But it doesn't induce any change to the context. If (18) is uttered subsequently, it is still true, since given that the match in question is actually dry (and perhaps assuming some other plausible facts about the lack of proximity of water or the unlikelihood of people soaking matches around these

parts), worlds in which it was soaked overnight are just too dissimilar to be relevant (even though we just mentioned them). In other words, accommodation cannot legitimately occur. Similarly, (20) cannot be made true by accommodation in the same context:

(20) If I had struck this match just now, it might have been soaked overnight.

We are now in a position to see how contextualism addresses all of the skeptical puzzles presented in §2. First, it can explain the inescapable clash between woulds and mights, while vindicating our intuition that many would-counterfactuals are true. Would-counterfactuals like (15) are true *in many contexts*, particularly the casual contexts of ordinary conversation (but also potentially in historical, psychological, philosophical, or scientific contexts). In many contexts, it is legitimate to ignore relatively low probability outcomes like my swift catching of the mug, quantum tunneling, or statistically unlikely flying to the counter. But when an undermining might-counterfactual is raised, as long as the conversational shift it brings about is accommodated, what counts as relevant changes, and (15) is not true in this newer, more precise context. The same goes when we are face to face with the chanciness of things. Contrary to what Hájek argues, on this view, it is not that chanciness undermines wouldiness; rather it is that chanciness, when raised to salience and accommodated in the conversational context, undermines wouldiness (a much less catchy slogan, admittedly). *Mutatis Mutandis* for indeterminacy.

The second puzzle challenged the truth of most contingent counterfactuals based on the fact that at least many of the undermining worlds are among the closest on a natural similarity ordering. The contextualist view addresses this problem in that the closeness ordering is no longer merely a matter of similarity, but a combination of similarity and relevance. Quantum worlds, for example, may be among the most similar, but they are often not among the most relevant, and thus will often not be among the closest.

Finally, the third puzzle concerned sequences of clashing would-counterfactuals, or reverse Sobel sequences. On the present view, (12), repeated here as (21) is true in the context in which it is asserted, since it is a context that legitimately ignores the relatively low probability possibilities in which you come to my party and don't see Lady Gaga anyway.

(21) If you had come to the party, you would have seen Lady Gaga.

(13), repeated below as (22), is true in the same context, since, given the facts about the situation, all the most similar-relevant worlds in which you are at the party and distracted by philosophical conversation out on the deck are worlds in which you miss Lady Gaga.

(22) If you had come to the party and been distracted by philosophical conversation out on the deck, you would not have seen Lady Gaga.

But (22) also makes this previously unconsidered possibility salient in the conversation. And given that it is not too distant — you are a philosopher, after all, and there were many other philosophers at the actual party — it is very likely to be taken seriously, and can no longer be legitimately ignored. So (21) is not true in the new context.

In this way contextualism explains why many would-counterfactuals are often true in the first place, but also why they seem to clash with other counterfactuals that apparently undermine their truth. They do clash with these other counterfactuals, but not in a way that undermines their truth once and for all. Counterfactual skepticism is avoided; many counterfactuals that we think are true are in fact true. The caveat is that they are not true in every context, though they are true in many contexts.

The other version of contextualism defended in print is the 'all cases' version defended by Ichikawa (2011). On his view, would-counterfactuals are contextually restricted strict conditionals:

> *If A were the case, C would be the case* is true just in case all of the A possibilities are C possibilities (Psst!—except those possibilities we're properly ignoring) (p.296)

While Ichikawa only addresses the third of the three skeptical puzzles discussed above, his version of contextualism also has the potential to solve all three in a similar way. The problem with the similarity relation doesn't arise because a similarity ordering isn't invoked at all; as long as the right possibilities are properly ignored, many counterfactuals will come out as true in many contexts. He could also define might-counterfactuals as the duals of woulds, so that they are simply contextually restricted existential quantifiers over possibilities. Might-counterfactuals, like woulds, introduce new possibilities, thereby changing the context, in much the same way I've described above, explaining the clash between woulds and mights.

The central difference between my version of contextualism and Ichikawa's is the logic of counterfactuals; since his is a strict conditional account and mine a variably strict conditional account, they differ in which rules of inference they validate. Let's begin with what the two views have in common. Substitution of equivalents is valid in both frameworks. That is, if the context does not shift when logical equivalents are substituted for each other, then the substitution will be valid. Clearly, if all A possibilities are C possibilities and the B possibilities are just the A possibilities, then all B possibilities are C possibilities (and the same for D possibilities instead of C possibilities, where C and D are logically equivalent). Similarly, if the closest A-worlds are all C-worlds, and the A-worlds just are the B-worlds, then the closest B-worlds are all C-worlds (and similarly for substituting the logically equivalent D for C). But the validity is also limited in its applicability, since very often substituting a logical equivalent will place us in a different context, and so it won't be true that the A possibilities range over the same domain as the B possibilities. This is because logical equivalents can raise different possibilities to salience. For example, consider (23) vs. (24), where ϕ_1 through ϕ_n are all the possible ways in which a mug can fall described on a microphysical level:

(23) If I had dropped my mug, it would have fallen to the kitchen floor.

(24) If I had dropped my mug, it would have fallen to the kitchen floor in manner ϕ_1 or manner ϕ_2... or ϕ_n.

In this case, (24) makes salient microphysical detail, and thus puts us in a much more precise context than (23). Quantum events and other low probability outcomes are going to be relevant in the context of (24) while they are not in that of (23).

Both theories can endorse agglomeration as a reasonable inference, in the sense that whenever the premises are truly asserted, the conclusion is true in the same context:

Agglomeration: $A \mathrel{\Box\!\!\rightarrow} B,\ A \mathrel{\Box\!\!\rightarrow} C \vdash A \mathrel{\Box\!\!\rightarrow} (B \,\&\, C)$

This is good, since it is an overwhelmingly intuitive principle. However, neither account validates the principle, since counterexamples come in the following form. Consider a fair lottery with a million tickets that is in fact

never played. The premises are each of the form *If the lottery had been played, ticket 1 would not have won*, *If the lottery had been played, ticket 2 would not have won*, and so on for each of the million tickets. Now consider these relative to a context in which low probability outcomes, such as those that have a one in a million chance of occurring, are legitimately ignored. Each premise then comes out true. But apply the principle of agglomeration, and the conclusion is certainly false at any context: *If the lottery had been played, ticket 1 would not have won and ticket 2 would not have won... and ticket 1 000 000 would not have won.* Hence, agglomeration is not a valid principle for a contextualist account of counterfactuals. But it is important to note that these counterexamples can only be constructed for contexts in which the premises are *not asserted*. On both accounts under consideration, if these million premises (or even a small subset of them) were asserted, they would land us in such a context in which low probability outcomes *do* matter and the premises wouldn't be true. So a version of agglomeration *qua* reasonable inference is still applicable.

The two frameworks differ on antecedent strengthening, contraposition, and transitivity.

Antecedent strengthening: $A \mathrel{\Box\!\!\rightarrow} C \vdash (A \wedge B) \mathrel{\Box\!\!\rightarrow} C$
Contraposition: $A \mathrel{\Box\!\!\rightarrow} C \vdash \neg C \mathrel{\Box\!\!\rightarrow} \neg A$
Transitivity: $A \mathrel{\Box\!\!\rightarrow} B$, $B \mathrel{\Box\!\!\rightarrow} C \vdash A \mathrel{\Box\!\!\rightarrow} C$

Ichikawa's strict conditional semantics validates all of these; apparent counterexamples are due to shifts in context. By contrast, my semantics invalidates these rules of inference, as it is a variably strict conditional semantics. Apparent counterexamples to antecedent strengthening and contraposition would generally not be able to be explained away anyway, since for the single premise rules there is no context shift between premise and conclusion. Unlike Ichikawa's account, on my view would-counterfactuals generally induce a context-shift, if they induce one at all, after, and not before, their semantics is calculated.[7] Therefore the standard counterexamples in the literature to these rules are also genuine counterexamples on my account.[8]

[7]An exception to this generalization is when the consequent includes much more precise considerations than the sort previously under consideration in the conversation, but this is only applicable here when these rules are combined with substitution of equivalents.

[8]Brogaard & Salerno (2008) also argue that apparent counterexamples to antecedent strengthening, contraposition, and transitivity are due to shifts in context. It's not clear

Are there benefits to one account or the other? On the one hand, Ichikawa's semantics is simpler. One might think that the ideal account is a strict conditional one; historically we only moved away from such an account to a variably strict one because of the counterexamples. If the alleged counterexamples can be explained away by appealing to context-sensitivity, this is a very appealing picture. On the other hand, there may be good reasons for wanting a semantic account of the invalidity of antecedent strengthening rather than a pragmatic one. For example, against background assumptions that the match in question is extremely reliable and there is no water for miles around, it seems (25a) and (25b) are genuinely true together at the same context and are not simply consistent because of a context shift like the in case of *Everything in the fridge is edible* and *Not everything in the fridge is edible*.

(25) a. If I had struck this match, it would have lit.

b. If I had struck this match and it was wet, it would not have lit.

For Ichikawa to account for this, the conversational participants have to (at least temporarily) take seriously worlds in which the match is wet, even if it has already been explicitly established in the conversation that the match wouldn't have been wet, e.g. the following is perfectly coherent: *There is no water anywhere, so if I had struck this match, there is no way it would have been wet. So (25a). Still, (25b) is true.* Now, it could be that there are very subtle shifts in context like this when it comes to sequences of counterfactuals, but one might think that the story of the two match examples is a lot more straightforward than that, especially when at the same time conversational participants would likely *not* agree to *If the match had been struck, it might have been wet*, even if this possibility only has to be considered temporarily.

On the other hand, validating transitivity seems like an advantage over my account, since as far as I can tell all the traditional counterexamples do involve an intuitive context shift between the premises and conclusion. Ichikawa's semantics nicely explains that data. So the logical considerations, though pertinent, do not without further consideration seem to point decisively in either direction.[9]

whether they are using this as an argument in favor of a contextualist strict conditional analysis or not.

[9]For a discussion of the metaphysical consequences of embracing either kind of contextualism for the notion of chance, see Emery (2015).

4 Counterfactuals and Knowledge

What is the relationship, if any, between contextualism for counterfactuals and for knowledge? They have very similar motivating skeptical puzzles. Both the first and third puzzles presented in §2 above closely parallel standard arguments when it comes to skepticism about knowledge. As Ichikawa emphasizes, clashing would-counterfactuals like (14) closely resemble the "abominable conjunctions" of DeRose (1995), such as:

(26) I don't know that I am not a brain in a vat, but I do know that I have hands.

This is similarly true for the clashes between woulds and mights. The possibility that I am a brain in a vat seems to undermine my knowledge that I have hands in the same way the pesky possibilities raised by the skeptics in §2.1 and §2.3 undermine the relevant counterfactuals. And Hawthorne (2004) (p. 5, fn. 10) points out that the lottery puzzle for knowledge parallels the second puzzle for counterfactuals, the pressure from similarity.

Since at least the puzzles that motivate the contextualist views are so similar, are there any other connections between contextualism for knowledge and counterfactuals?[10] In my work on the subject, I have made no claims either way about a connection between contextualism for counterfactuals and for knowledge. But Ichikawa argues that the domain over which counterfactuals range is the very same domain over which knowledge claims range. Call this *identity contextualism* for short.

Ichikawa endorses a Lewis-style contextualism for knowledge, essentially:

> S knows that p if and only if S's evidence eliminates all the ¬p-cases, where 'all' is a context-sensitive restricted quantifier.

(Ichikawa's more developed version includes reference to the basis of the evidence, but this complication need not concern us for present purposes.) The identity claim is that the 'all cases' in the above definition for knowledge and the 'all cases' in the definition for counterfactuals are identical in any

[10]For considerations of space, I put aside here the very interesting question of whether there is better or different linguistic evidence for contextualism for one or the other, as well as the question of what alternative options there are for each, e.g. subject sensitive invariantism is an option for an account of knowledge claims whereas it is not an option for an account of counterfactuals.

given context. Ichikawa offers three reasons to support this: 1) The domains seem to shift together, i.e., the same possibilities can be ignored when it comes to knowledge and counterfactuals, and the same possibilities when introduced into the domain undermine knowledge claims and counterfactuals alike. 2) Theoretical simplicity and 3) Treating the domains as identical allows the dissolution of counterexamples to treating safety and sensitivity as necessary conditions on knowledge.

In support of the first point, Ichikawa offers the case of Blanche the professor of abstract science:

> Blanche is, and has long been, Professor of Abstract Science. This, as Melissa is well aware, won't change any time soon. So (27) is true:
>
> (27) Melissa knows that Blanche will be Professor of Abstract Science tomorrow.
>
> In the unlikely event that Ida, the principal, abandoned her position tonight, Blanche would leave her post as Professor of Abstract Science and take her place:
>
> (28) If Ida resigned tonight, Blanche would be principal tomorrow. (p.299, my numbering of examples)

As Ichikawa points out, the domains for both (27) and (28) in their natural, non-skeptical contexts do not include remote possibilities in which Blanche resigns tonight. Moreover, if the possibility that she resigns tonight is explicitly introduced into the domain, both (27) and (28) are undermined. Ichikawa writes, "I am unaware of any data suggesting that the two domains should come apart; consideration of particular cases like this one provides some reason for thinking that they will shift together". (p. 300)[11]

[11]This example seems like an odd choice in support of the view, since (27) and (28) can't range over the same domain and both be true (unless (28) is trivially true, which is not what I think Ichikawa has in mind). For (28) to be non-trivially true, there are some possibilities in the domain over which it ranges in which Ida resigns and Blanche is principal tomorrow and so not Professor of Abstract Science. For (27) to be true, there can be no such cases in the domain, since presumably Melissa's evidence doesn't rule out cases in which Ida resigns tonight; rather those cases are properly ignored in the natural, non-skeptical context in which (27) is true. This is just to comment on the oddity of the case in support of identity contextualism, not an objection to identity contextualism

In support of the third point, since safety and sensitivity are generally formulated in terms of counterfactuals, and since they are contrapositives of each other, on Ichikawa's semantics for counterfactuals, they are equivalent, necessary conditions on knowledge:

Sensitivity: S knows that p only if $\neg p \,\Box\!\!\rightarrow \neg B(p)$ [on the basis of evidence E]
Safety: S knows that p only if B(p) [on the basis of evidence E] $\Box\!\!\rightarrow p$

In brief, some of the advantages of adopting identity contextualism is that it solves the alleged problem of sensitivity implying the failure of single-premise closure for knowledge claims, because the context shifts between premise and conclusion in the apparent counterexamples. It also resolves various counterexamples that have nothing to do with the failure of closure. For example, consider, as Ichikawa does, Ernie Sosa's garbage chute case.

> **Garbage chute:** Anna throws a trash bag down the chute of her high rise condo. She knows that the bag will soon be in the basement. But in the (incredibly unlikely) case that the bag gets caught in the chute on the way down, she would still believe that the bag would soon be in the basement (based on the same evidence). Thus Anna's belief is insensitive, but seems to constitute knowledge anyhow.

The counterfactual that establishes sensitivity is *If the bag were not to arrive shortly in the basement, Anna would not believe that the bag would arrive shortly in the basement.* On Ichikawa's identity contextualism, this is *true* in the same context as *Anna knows the bag will arrive shortly in the basement*, since in that context there are no worlds in the domain in which the bag fails to arrive in the basement. Of course, we hear the counterfactual as false, but this is because as soon as it is asserted, it shifts the context so at to include worlds in which the bag doesn't arrive in the basement. In this context, the counterfactual is false, but so is the knowledge claim. It is Ichikawa's

itself. Since cases in which Ida resigns are not part of the natural context of (27), there is nothing in Ichikawa's view that requires that these two have the same domain.

contention that all alleged counterexamples can be explained away in this way.[12]

Is identity contextualism right? It certainly is theoretically desirable; we get one contextualism for knowledge and counterfactuals, and it offers an elegant explanation of how sensitivity and safety are necessary conditions on knowledge. But I worry it can't be right. Knowledge claims deal in epistemic possibilities, would-counterfactuals (by most people's lights) in metaphysical ones. By identifying the domains over which they each range, either counterfactuals are more epistemic or knowledge claims more metaphysical than is plausible. Consider the following case. Suppose Kristy and Mary Anne are discussing whether they should take a particular vase off the table before they move the table (since if they move the table with the vase on it, it might fall off). The vase is actually extremely fragile, but they do not know what the vase is made out of, or whether or not it is fragile. Suppose further that the table is 5 feet off the ground, and the floor is marble. In the natural, non-skeptical context the three following sentences are all intuitively true:

(29) If the vase were to drop, it would break.

(30) Kristy doesn't know that if the vase were to drop, it would break.

(31) Kristy doesn't know that the vase is fragile.

For (29) to be true on Ichikawa's semantics, all the worlds in the domain are either ¬drop-worlds or they are drop&break-worlds. In this case, it is trivial that Kristy's evidence rules out all drop&¬break-worlds, since there are none in the domain. Since by the contextualist definition of 'know', Kristy's evidence ruling out all the drop&¬break-worlds in the domain is a sufficient condition for knowledge, it follows that Kristy knows that if the vase were to drop, it would break (contra our intuition about (30)). Similarly, assuming that for the vase to be non-fragile means that it can drop without breaking, it is trivial that Kristy's evidence rules out all worlds in which the vase is not fragile, since there are none in the domain. So, contra (31), Kristy does know that the vase is fragile. But this is absurd. The whole set up for the case is that neither Kristy nor Mary Anne know anything about the compositions or dispositions of the vase vis-a-vis its fragility. Now take the case where the domain is such that it makes (30) and (31) true, that is,

[12]See Ichikawa (2011) for much more detail on sensitivity and safety.

there are some worlds in the domain (which Kristy's evidence cannot rule out) in which the vase is not fragile, i.e., it drops but does not break. In this case, (29) is straightforwardly false if it ranges over this same domain.[13]

While contextualism may be the right semantic strategy for counterfactuals and knowledge alike (though I want to make no claims about the latter), I have cast doubt on whether the contextualist semantics for the two range over the same domain of possibilities in a given context. It is possible that a more sophisticated contextualism for knowledge or for counterfactuals (or both) could resolve these worries.[14]

References

Brogaard, Berit & Joe Salerno. 2008. Counterfactuals and Context. *Analysis* 68(297). 39–46.

DeRose, Keith. 1995. Solving the skeptical problem. *Philosophical Review* 104(1). 1–52.

DeRose, Keith. 1999. Can It Be That It Would Have Been Even Though It Might Not Have Been? *Noûs, Supplement: Philosophical Persepctives* 33(13). 385–413.

Emery, Nina. 2015. The Metaphysical Consequences of Counterfactual Skepticism. *Philosophy and Phenomenological Research* (Advanced online access: http://dx.doi.org/10.1111/phpr.12254).

von Fintel, Kai. 2001. Counterfactuals in a Dynamic Context. In *Ken Hale: A Life in Language*, MIT Press.

Gillies, Thony. 2007. Counterfactual Scorekeeping. *Linguistics and Philosophy* 30. 329–360.

[13]Moving to Ichikawa (2011)'s more sophisticated version of contextualism with basing doesn't help here. On this version, S knows that p just in case, for some evidence E, (i) S believes that p on the basis of E, and (ii) all E cases are p cases. (p.301) The counterexample need only be adjusted to include that Kristy believes that if the vase were to drop, it would break (or that the vase is fragile) based on some intuitively irrelevant or inconclusive evidence, e.g. that it is a vase. Since there are no drop&¬break-worlds in the domain, any evidence, no matter how irrelevant, will trivially eliminate any such worlds.

[14]See Ichikawa (Forthcoming) for further discussion.

Hájek, Alan. ms. Most Counterfactuals are False. ANU, monograph in progress.

Hawthorne, John. 2004. *Knowledge and Lotteries*. Oxford University Press.

Hawthorne, John. 2005. Chance and Counterfactuals. *Philosophy and Phenomenological Research* 70(2). 396–405.

Ichikawa, Jonathan. 2011. Quantifiers, Knowledge, and Counterfactuals. *Philosophy and Phenomenological Research* LXXXII(2). 287–313.

Ichikawa, Jonathan Jenkins. Forthcoming. *Contextualizing Knowledge*. Oxford University Press.

Kratzer, Angelika. 1977. What 'must' and 'can' must and can mean. *Linguistics and Philosophy* 1(3). 337–355.

Kratzer, Angelika. 1981. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic* 10(2). 201–216.

Lewis, David. 1973. *Counterfactuals*. Blackwell.

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

Lewis, Karen. 2016. Elusive Counterfactuals. *Noûs* 50(2). 286–313.

Lewis, Karen. Forthcoming. Counterfactual Discourse in Context. *Noûs* .

Moss, Sarah. 2012. On the Pragmatics of Counterfactuals. *Noûs* 46. 561–86.

Moss, Sarah. 2013. Subjunctive Credences and Semantic Humility. *Philosophy and Phenomenological Research* LXXXVII(2). 251–278.

Stalnaker, Robert. 1968. A Theory of Conditionals. In N. Rescher (ed.), *Studies in logical theory*, Oxford University Press.

Stalnaker, Robert. 1981. A Defense of Conditional Excluded Middle. In Harper et al. (ed.), *Ifs: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht: D. Reidel.

Williams, Robert G. 2008. Chances, Counterfactuals, and Similarity. *Philosophy and Phenomenological Research* 77(2). 385–420.