Synthese (2021) 199:637–659
https://doi.org/10.1007/s11229-020-02686-0

Counterfactual scepticism and antecedent-contextualism

Alan Hájek[1]

Received: 12 December 2019 / Accepted: 2 May 2020 / Published online: 29 May 2020

Abstract
I have argued for a kind of 'counterfactual scepticism': most counterfactuals ever uttered or thought in human history are false. I briefly rehearse my main arguments. Yet common sense recoils. Ordinary speakers judge most counterfactuals that they utter and think to be true. A common defence of such judgments regards counterfactuals as *context-dependent*: the proposition expressed by a given counterfactual can vary according to the context in which it is uttered. In normal contexts, the counterfactuals that we utter are typically *true*, the defence insists, while granting that there may be more rarefied contexts in which they are false. I give a taxonomy of such contextualist replies. One could be a contextualist about the counterfactual connective, about its antecedent, or about its consequent. I offer some general concerns about all these varieties of contextualism. I then focus especially on *antecedent-contextualism*, as I call it. I firstly raise some high-level objections to it. Then, I look at such a contextualist account due to Sandgren and Steele. I think it has many virtues, but also some problems. I conclude with some avenues for future research.

Keywords Counterfactuals · Counterfactual scepticism · Contextualism · Antecedent-contextualism · Interpretation · Ceteris paribus laws · Sandgren · Steele

1 Introduction

Most counterfactuals are false. I have argued for a kind of 'counterfactual scepticism' (2020). I imagine an extremely long transcript of all the counterfactuals that have ever been and that ever will be uttered or thought, in human history. Most of the entries on the transcript are false—or so I argue. In the next section, I will briefly rehearse my main arguments.

✉ Alan Hájek
hajek.alan@gmail.com

[1] Australian National University, Canberra, Australia

Yet common sense recoils. Ordinary speakers judge most counterfactuals that they utter and think to be true. Here are a couple of recurring examples in the literature:

If the cup (plate) were released, it would fall.

We hold a cup (or a plate) up in the air. We don't in fact release it, but we make a claim about what would happen after its hypothetical release.

If Sophie had gone to the parade, she would have seen Pedro dance.

There was a parade of baseball stars, including Pedro, and at one point he danced. Alas, Sophie did not attend the parade, so she missed his dance. But we judge this counterfactual to be true. And so it goes. I attribute widespread errors to competent speakers!

It is not surprising, then, that a number of philosophers have defended our judgments against counterfactual scepticism. The most common defence that I encounter regards counterfactuals as *context-dependent*: the proposition expressed by a given counterfactual can vary according to the context in which it is uttered. So far, this is compatible with my scepticism—the different propositions expressed could be false, irrespective of the context. ('All tall people have two heads' is context-dependent, since the gradable adjective 'tall' is, but it is false irrespective of the context.) But the defence then continues that in normal contexts, the counterfactuals that we utter are typically *true*, while granting that there may be more rarefied contexts in which they are false. For example, we may say that 'if the cup were released, it would fall' is true when uttered under usual circumstances on the street or in your kitchen, say. To be sure, we might draw attention to anomalous possibilities in which the cup does not fall—as I am fond of doing—creating a context in which the counterfactual is false. Nevertheless, most counterfactuals are uttered in normal contexts like that of the street or the kitchen, and their truth in those contexts assures us that most of the entries on my transcript are true. Or so go contextualist replies to my scepticism.

I will give a taxonomy of such replies. One could be a contextualist about the counterfactual connective, about its antecedent, or about its consequent. I will offer some general concerns about all these varieties of contextualism. I will then focus especially on *antecedent-contextualism*, as I call it. I will firstly raise some high-level objections to it. Then, I will look at an especially well-articulated account of it in more detail, that of Alex Sandgren and Katie Steele's "Levelling Counterfactual Scepticism" (2020). I think it has many virtues, but also some problems. I will present their account, bringing out some issues concerning its formulation, before turning to more specific objections to it. I will conclude with some avenues for future research. But the bottom line remains the same. I maintain that counterfactual scepticism is on the level: it is the true account regarding counterfactual truth values.

2 My main arguments for counterfactual scepticism

2.1 Chancy consequents

Here is a fair coin that will never be tossed. Consider this counterfactual:

If the coin were tossed, it would land *heads*. (Not tails!)

I say that this is false. Coin tossing is a paradigm example of a chancy process. The counterfactual second-guesses the outcome of this hypothetical chance process, when it has no business doing so. Now, let the coin be highly biased to heads, but still with some chance of landing tails—say, a 99% chance of landing heads when tossed, and a 1% chance of landing tails. I say that the counterfactual is still false. It's still a hypothetical chance process, and the counterfactual still has no business second-guessing its outcome. *Chance undermines 'woulds'*.

Science apparently tells us that we live in a chancy world. Even macroscopic events that we take to be deterministic are chancy. If I were to release the cup, there would be some positive chance that it would be lifted by a sudden updraft, or saved from falling by a lunging spectator. Even the chances of its vaporising, or quantum tunnelling to China, would be positive (albeit miniscule). So it is false that if the cup were released, it would fall. If Sophie had attended the parade, there would have been a positive chance of her not seeing Pedro dance—for example, there would have been a positive chance of her getting stuck behind a tall person. And so it is false that if she had attended the parade, she *would* have seen him dance.

This argument for scepticism focused on the consequents. The next focuses on the antecedents.

2.2 Unspecific antecedents

Even if our world is deterministic, this counterfactual is still false:

If the cup were released it would fall.

The antecedent is unspecific: the cup is released *somehow or other*. It does not specify the exact initial conditions of the cup's release. According to statistical mechanics, this macrostate is compatible with many microstates that deterministically evolve to anomalous microstates—for example, ones that correspond to the cup vaporising instead of falling. The counterfactual leaves open which microstate would obtain, yet makes a specific claim about how things would transpire with respect to the cup. It is false in virtue of this mismatch in specificity.

We don't need high-powered physics to make the point. Counterfactuals with unspecific antecedents and comparatively specific consequents are false. Consider:

If Bizet and Verdi had been compatriots, they would have both been *French.* (Not Italian!)

I say that is false. We have an unspecific antecedent coupled with a comparatively specific consequent.

I think that most counterfactuals are like this, just less obviously so. If Sophie had attended the parade (somehow or other), she would have seen Pedro dance (specifically). *False*, I say. A counterfactual cannot second-guess the resolution of an unspecific antecedent. *Unspecificity undermines 'woulds'*.

This argument focused on the antecedents. The next focuses on the relationship between the antecedents and consequents, brought about by either probabilistic consequents or unspecific antecedents (or both).

2.3 'Might nots'

Both chanciness and unspecificity give rise to 'might not' counterfactuals. 'Might not' counterfactuals conflict with their 'would' counterparts. Just try saying out loud this sentence:

> If the cup were released, it might *not* fall; but it would fall if it were released. #

Or this one:

> If Sophie had attended the parade, she might *not* have seen Pedro dance; but she would have seen Pedro dance if she had attended the parade. #

In each case, the first part clashes with the second part. I say that it is the clash of incompatibility. *'Might nots' undermine 'woulds'*.

The 'might not' counterfactuals are typically true; so the corresponding 'would' counterfactuals are typically false. Sophie might not have seen Pedro dance: she *might* have got stuck behind a tall person. So it's false that she *would* have seen Pedro dance.

Three of my arguments for counterfactual scepticism are now before you. (I have more!) But there have been various contextualist responses to it.

3 Contextualism about counterfactuals

Contextualism concerning some part of language is the thesis that the very same words taken from that part may express different propositions in different contexts. For example, nearly everyone is a contextualist about indexicals. "I am hungry now" expresses a different proposition in your mouth from mine, and different propositions at different times.

On the classic Stalnaker/Lewis counterfactual semantics, a sentence of the form 'if A were the case, B would be the case' (or equivalent locutions) expresses a proposition determined by three parts:

1. the proposition expressed by its antecedent A;
2. the proposition expressed by its consequent B; and
3. a function, represented by the counterfactual connective '$\Box\!\!\rightarrow$', from those two propositions to another one.

The kinds of contextualism about counterfactuals that are my concern regard one or more of these as *moving* parts in a counterfactual—they may shift across contexts. What I will call *antecedent-contextualism* regards the interpretation of the *antecedent* as context-sensitive. *Consequent-contextualism* regards the interpretation of the *consequent* as context-sensitive. And *connective-contextualism* regards the function in 3 as context-sensitive. Furthermore, we could combine these kinds of contextualism, or not. Accordingly, one may be a contextualist about one, two, or three of these parts,

or a non-contextualist (though I'm not aware of any contextualist about two or more of them).

Antecedent-contextualism will be my main target in this paper, but let me briefly say something about the other kinds of contextualism. I will start with consequent-contextualism, a natural response to my counterfactual scepticism that a number of people have put to me. Against my argument from chancy consequents, it might be tempting to interpret

'if *A* were the case, *B* would be the case' as
'if *A* were the case, *B* would *probably* be the case', or as
'if *A* were the case, *B* would *typically* be the case'.

Then we could point out that the operators 'probably' or 'typically' are context-dependent. See Leitgeb (2012) for a sophisticated development of such an account (the "probably" version).[1]

Let's consider the first of these interpretations (my objections below will apply equally to the second). For example, suppose you say on the street: "If the cup were released, it would fall." We might regard that as true, since if the cup were released, the chance of its falling would be very high—high enough to clear the threshold for 'probable' that is appropriate for street-talk. However, I may then raise the bar by drawing attention to the minute chances of various anomalous things happening. I thereby create a new context in which the counterfactual is false.

Consequent-contextualism may seem to respond to my argument from chanciness, but it does so by changing the meaning of the original counterfactual. My argument from 'might nots' brings this out. I said that a 'might not' counterfactual clashes with the corresponding 'would' counterfactual, regarding this as the clash of incompatibility. Now perhaps you regard it instead as pragmatic (see DeRose 1999). But there is no clash *at all* between a 'might not' counterfactual and the corresponding 'would probably' counterfactual. "If the cup were released it might not fall, but it would *probably* fall" is perfectly felicitous. Indeed, far from being in tension, 'might not' is *scalar implicated* by 'would probably'! Saying that the cup "would *probably* fall" rather than outright "would fall" pragmatically conveys that the cup's *not* falling is a live possibility—that is, it might not fall. (It is similar to how "it will *probably* rain" pragmatically conveys that it is a live possibility that it will *not* rain.)

Furthermore, it is unclear how this kind of contextualism responds to my argument from unspecific antecedents. After all, that argument did not turn on considerations of low chance. See my (2014) for further objections to Leitgeb's account.

[1] Here is a key passage:

> (1) If the match were struck, it would light.
> … it should be unproblematic to qualify the consequent of (1) in either of the following ways, while leaving the meaning of the conditional—completely or at least almost completely—unaffected:
> (2) If the match were struck, it would necessarily [definitely] light
> (3) If the match were struck, it would be very likely to light. (p. 36)

That said, he goes on to reconstruct (3) as a conditional chance statement:

> (4) The conditional chance of the match lighting given that it is struck is very high. (p. 37)

I think that connective-contextualism is more promising. And an especially promising version of such contextualism is due to Karen Lewis in "Elusive Counterfactuals" (2016). Here is one statement of it:

> For all contexts c, P $\square\!\!\rightarrow$ Q is true in c iff all the closest P-worlds are Q-worlds, where closeness is a function of both similarity and relevance [in c, presumably]. (p. 292)

Notice that the antecedent and consequent remain fixed. What varies is the function from P and Q to the proposition expressed by the counterfactual: that function depends on the contextually variable similarity and relevance relations. More specifically, she writes: "Counterfactuals do not just quantify over the most similar worlds, but the worlds that are relevant in the context, or, roughly speaking, the semantic content of a counterfactual is sensitive to the standards of precision in the conversation" (p. 291), and to the "conversational purposes" (p. 293).

Lewis argues that this context-dependence can defuse all three of my arguments for counterfactual scepticism. For example, in normal-standards contexts, low-chance possibilities such as the cup flying upwards or quantum tunnelling are not relevant, so they are not realised among the closest worlds where it is released. However, by drawing attention to these possibilities, I can make them relevant: I raise the standards of precision, shifting us to a more demanding context, and thereby changing the closeness relation to one for which these possibilities *are* among the closest such worlds. Regarding my unspecific antecedents argument, Lewis argues that given assumptions about conversational purposes (e.g. expressing regret over Sophie not being at the parade), certain possibilities are legitimately ignored as irrelevant (e.g. her being stuck behind a tall person). However, they may become relevant if those purposes change. And in her explanation of how 'might' counterfactuals can induce context shifts, Lewis says "in uttering a might-counterfactual (that involves previously ignored possibilities), *the speaker* raises the conversational stakes—*she takes* lower probability events to be relevant" (p. 295, my emphases).

I will not reply to this account in detail here; I will just note some general qualms I have, mindful that it is worthy of an article-length response in itself. Firstly, I am concerned that this kind of contextualism makes counterfactuals too subjective, dependent on what *the speaker* takes to be relevant. It sounds like the more ignorant or unimaginative you are, and hence the less you take to be relevant, the easier it is for your counterfactuals to come out true. Secondly, I find the notion of a "context" to be … well, somewhat elusive. To be sure, we need the notion in the semantics of various lexical items, such as indexicals, demonstratives, and gradable adjectives. (Lewis extends this list.) That said, the notion seems especially hard to pin down for counterfactuals. Suppose you utter the 'cup' counterfactual in a bar to a barman and to a quantum physicist who is jotting down notes on quantum tunnelling. What is the context? One with lower or higher standards? Suppose you don't utter the counterfactual at all; you merely think it in the privacy of your study. What is the context? There is no conversation, hence no standards of precision in the conversation.

To be sure, Lewis could arguably develop her account to help resolve such questions. She could add more details, to settle the standards for conversations whose participants have varied or even conflicting purposes. She could extend her characterisation of

contexts, so that a private thought also has one—perhaps dependent on what else you are thinking and the purpose of the thinking.[2] And yet in both the bar and the study cases, without knowing what the context is, the counterfactual can still be understood perfectly well, I claim. In any case, I think that other problems for contextualism will apply. See in particular my objections in Sect. 8.4 concerning disagreement and retraction. But before we can get to them, we must turn to the kind of contextualism that is my main focus.

4 Antecedent-contextualism

Again, the central idea here is that the counterfactual has a context-sensitive effect on the *antecedent*. Natural versions of the idea qualify the antecedent: it is realised in *normal*, or *typical*, or *ordinary*, or *relevant* circumstances, or what have you, where these qualifications are context-dependent. Understood this way, we may regard this kind of contextualism as strengthening the antecedent with such a qualification:

$$(A \,\&\, Q_C) \,\Box\!\!\rightarrow\, B.$$

And strengthening the antecedent—making it more specific—may be thought to provide relief from my argument from unspecific antecedents. For example, we might say, 'if Sophie had attended the parade *and things had gone normally*, she would have seen Pedro dance' is true; hence, the original 'Sophie' counterfactual is true (on this interpretation). Likewise, it is true when the antecedent is qualified with 'and things had gone typically', 'in ordinary circumstances', 'in relevant circumstances', or what have you. Or so we might say.

Now, I deny that even these strengthened-antecedent counterfactuals are true. Someone getting stuck behind a tall person at a parade is hardly 'abnormal', 'atypical', 'extraordinary' or 'irrelevant'; it happens at nearly all parades. But if you insist that it is abnormal, say, and thus should be ruled out by the antecedent, then I think this only drives home how this reinterpretation weakens the counterfactual, and hence changes its meaning. The reinterpreted counterfactual lacks some entailments that the original counterfactual had. For example, the original counterfactual,

> (SEEN) if Sophie had attended the parade, she would have seen Pedro dance

entails

> if Sophie had attended the parade, she would not have been stuck behind a tall person.[3]

But this is not entailed by

[2] Thanks to an anonymous referee here.

[3] Here I assume for simplicity that her getting stuck behind a tall person entails her not seeing Pedro dance. If not, just change the entailed counterfactual to: 'if Sophie had attended the parade, she would not have been prevented from seeing Pedro dance'.

(NORMALLY-SEEN) if Sophie had attended the parade *and things had gone normally*, she would have seen Pedro dance.

For the latter leaves open the possibility of her attending the parade but things not going normally, notably by her getting stuck behind a tall person. The same goes for the other qualified reinterpretations.

Moreover, in strengthening the antecedent, clashes are lost, be they pragmatic or semantic.

Suppose you say:

(MIGHT NOT SEEN) if Sophie had attended the parade, she *might not* have seen Pedro dance.

We should all agree that this clashes with SEEN. But there is no clash whatsoever between MIGHT NOT SEEN and NORMALLY-SEEN. Sophie might not have seen Pedro dance because things might *not* have gone normally. Indeed, far from clashing, MIGHT NOT SEEN is even conversationally implicated by NORMALLY-SEEN. Drawing attention to what would have been the case for a *normal* attendance suggests that this is the strongest thing that can be said with respect to Sophie's prospects for seeing Pedro dance; all bets are off otherwise. A clash with SEEN vanishes with NORMALLY-SEEN; hence they do not mean the same thing. These points carry over to the other qualified reinterpretations, involving 'typical', 'ordinary', 'relevant', and what have you.

Here's another way to see how strengthening the antecedent weakens the counterfactual. Let B be Q_C in the antecedent-contextualism schema:

$$(A \,\&\, Q_C) \,\Box\!\!\rightarrow\, Q_C.$$

This counterfactual is so weak as to be a logical truth. For example, this is logically true:

If the cup were released and circumstances were normal, circumstances would be normal.

But now consider the unadorned counterfactual of which this is the putative interpretation:

If the cup were released, circumstances would be normal.

I don't think that is true, but in any case *nobody* should think that it is logically true.

Later on, I will consider the reply that the very placement of Q_C in the consequent shifts the context away from C, let's say to K. In that case, the counterfactual really has the form

$$(A \,\&\, Q_K) \,\Box\!\!\rightarrow\, Q_C,$$

which need not be a logical truth. For now, I raise this as a challenge to any form of antecedent-contextualism.

My discussion of antecedent-contextualism has so far taken place at a rather high level of abstraction. I have offered a template for such contextualism, and I have raised a couple of general objections to accounts that fit it. But I should canvass such an account that has been worked out in considerably more detail. I regard it to be one of the most significant responses to my counterfactual scepticism, because it not only objects to my arguments, but more interestingly, it offers a positive view about counterfactuals that has a number of virtues. So I turn to Sandgren and Steele's account.

5 Sandgren and Steele's account

Sandgren and Steele offer a specific antecedent-contextualist response to counterfactual scepticism:

> In summary, we propose that a statement like 'if A were true then B would be true' must be interpreted with respect to a domain of inquiry d, and accordingly represented as '$A_d \,\Box\!\!\rightarrow B$', where A_d is an interpretation of the antecedent as involving a fitting circumstances clause tied to domain d, and where the *box arrow,* $\Box\!\!\rightarrow$, represents counterfactual dependence… our proposal concerns the correct interpretation of the antecedent and not the interpretation, as it were, of the box arrow …[4] (p. 9)

So on their view, the antecedent may be interpreted differently, depending on the context. When a counterfactual is uttered, its context is a *domain of scientific inquiry*. Associated with this domain is a set of *fitting circumstances*, which "amount, roughly, to ideal conditions in which the regularity or regularities pertinent to the domain in question hold in isolation, absent *interfering* factors" (p. 5). Thus, following the schema that I introduced in Sect. 4, we have the qualification Q_C identified with F_D: context C picks out a domain of inquiry D, for which there are fitting circumstances F_D. So our representation of the interpretation of 'if A were the case, B would be the case (relative to domain of inquiry D)' is:

$$(A \,\&\, F_D) \,\Box\!\!\rightarrow\, B.$$

Here's my statement of the heart of their proposal:

> *'If A were the case, B would be the case', stated or thought in a context in which D is the domain of scientific inquiry,*
> *is interpreted as:*
> *'If A were the case and D's fitting circumstances held, B would be the case'.*

Central to this are the notions of "domain of scientific inquiry" and its "fitting circumstances".[5]

[4] I replace their "C" with "B" to conform with the notation of my template, for which I reserve "C" to mark the context.

[5] Notice that their view, as I have presented it, does not mention *levels* of scientific inquiry at all, although their paper does (even in its nice title). All the work seems to be done by the relevant "*domain* of scientific inquiry" (p. 4).

Sandgren and Steele emphasise that they understand the relevant "science" in a very broad sense, extending it to "more homely domains"—what they call "*extra-special* sciences" (p. 11). For example, consider:

(FUN) Were Alice to come to the party, it would be fun,[6]

uttered "on the understanding that Alice is generally very personable and fun-loving" (p. 2). Here, "the relevant domain of inquiry is plausibly one concerning Alice's impact on others in ordinary social milieu" (p. 6). Even in this "extra-special" scientific domain, there are lawful regularities according to Sandgren and Steele; I will say more about the nature of the relevant laws in a moment.

Notice how important this broad understanding of 'science' is to their answer to counterfactual scepticism. A narrower understanding would not speak to the majority of counterfactuals that people utter. Fundamental physics gives us no guidance regarding Alice's party proclivities or Sophie's parade-viewing prospects; neither does biology, sociology, economics, or other 'special sciences'. Yet I think that my imagined transcript consists predominantly of counterfactuals like the ones about Alice or Sophie—high-level counterfactuals concerning highly circumscribed domains. Our everyday conversations abound with counterfactuals about specific people or non-repeatable macroscopic events. In order to answer my scepticism, 'scientific inquiry' needs to include such domains; the 'sciences' often need to be 'extra-special'.

What about "fitting circumstances"? Here Sandgren and Steele appeal to *ceteris paribus* (*cp*) *laws*. For example, 'shield volcanoes erupt effusively' is a *cp* law of volcanology. It does not apply to all shield volcanoes—for example, not to pyroclastic volcanoes—but it informatively captures a general tendency among shield volcanoes. Sandgren and Steele believe that undergirding a counterfactual uttered in a given scientific domain of inquiry, there are regularities in that domain captured by *cp* laws. One cannot give non-trivial necessary and sufficient conditions for the holding of such laws—doing so would render them *laws*, simpliciter, which they are not. But one can give *sufficient* conditions for the holding of such laws, and *fitting circumstances* are such conditions.

Notice an important choice point here, and the wise choice that they have made, given their project of answering counterfactual scepticism. One could instead regard *cp* laws as *weaker* than strict laws: stating what happens probably or typically, but not what is guaranteed to happen. Appealing to *cp* laws understood that way would play into the hands of a sceptic like me. For example:

[6] As an anonymous referee has pointed out, this example, like some others in this paper, would be classified by linguists as a *future less vivid conditional*. It is about the future, with the antecedent still realisable, but there is an implicature that it is more likely to be false than true. (See Iatridou 2000). On the other hand, I think that all of the examples in this paper would count as "counterfactuals" according to the *Stanford Encyclopedia of Philosophy's* classification. (See Starr 2019, §1.1.) This is not merely a terminological issue if the distinction matters to the semantics, as some think it does (e.g. Iatridou). However, since neither Sandgren and Steele nor I offer semantics for such conditionals, I believe that my discussion can proceed without resolving this semantic issue. All that matters for my purposes are that my arguments go through for the falsehood of the conditionals, and against Sandgren and Steele's interpretation of them—particularly that it weakens them so as to change their meaning.

- 'Shield volcanoes (only) *probably* erupt effusively' does not entail of any particular volcano that

  if it were a shield volcano, it would erupt effusively. ()

  Chance undermines 'woulds'!
- The unspecific antecedent (a shield volcano of some kind or other) and comparatively specific consequent (erupt effusively, and no other way) yield a false counterfactual. Unspecificity undermines 'woulds'!
- And *cp* laws understood this weak way support and perhaps even entail 'might not' counterfactuals: if it were a shield volcano, it *might not* erupt effusively (e.g. it might be pyroclastic). 'Might nots' undermine 'woulds'!

Instead of understanding *cp* laws as *weak* laws, Sandgren and Steele understand them as *restricted* laws—ones that hold in suitably narrow circumstances. These are "fitting circumstances". But when these circumstances hold, the law holds deterministically. I'm not sure what are fitting circumstances for shield volcanoes to erupt effusively, but once they are in place, such volcanoes are *guaranteed* to erupt effusively. So

- there is no chance of their not doing so;
- strengthening  's antecedent with a conjunct that fitting circumstances hold is specific enough to entail the consequent; and
- so strengthened, the corresponding 'might not' counterfactual is false.

Sandgren and Steele can even agree with another part of my view about counterfactuals: roughly, that the only true ones are those for which there is nomic entailment between antecedent and consequent. But we differ on how the counterfactuals should be interpreted. I interpret them literally: we should take them at face value. And we differ on the nature of the underlying laws that secure the entailment: I take them to be strict laws, whereas Sandgren and Steele take them often to be *cp* laws.

It should be clear how their view is supposed to answer counterfactual scepticism. Take a counterfactual that we normally regard as true: "If the cup were released, it would fall". The domain of inquiry is something like *the normal behaviour of macroscopic objects*. In this domain, fitting circumstances are ones in which the *cp* laws governing such objects hold—among other things, free from quantum mechanical or statistical mechanical anomalies. So the counterfactual is interpreted as something like: "If the cup were released *and circumstances were fitting for the normal behaviour of macroscopic objects*, it would fall". And that is supposed to be true. However, if I am in a philosophy seminar in which I have just drawn attention to such recherché possibilities from microphysics, the context is different: the domain of inquiry, and with it the fitting circumstances, have changed. Now, what previously were anomalies no longer count as such—fitting circumstances include them. The counterfactual is interpreted as something like: "If the cup were released *and* it were to obey the laws of quantum mechanics and statistical mechanics, it would fall." And that is false—those laws give positive chance to the cup not falling.

More generally, on Sandgren and Steele's view,

> interpreting a counterfactual requires identifying a domain of inquiry, which will make a difference to the content of the antecedent. This means that the *truth*

> *conditions* of counterfactual claims depend on the pertinent domain of inquiry. In turn, the truth *values* of counterfactuals depend on what *lawful regularities* actually hold. (p. 6)

To summarise: when we utter counterfactuals in normal contexts, their antecedents are interpreted so as to rule out confounding circumstances. My sceptical arguments draw attention to just such circumstances. In doing so, they create a *new* context in which the counterfactuals are now interpreted with new antecedents in which those circumstances are no longer ruled out; and so the counterfactuals come out false.

Now, there is a suppressed premise that the counterfactuals uttered in normal contexts are not only interpreted in particular ways, but furthermore that *so-interpreted, they mostly come out true*. After all, without this premise, this contextualism is compatible with my scepticism. It could be that the counterfactuals are false even after they are interpreted à la Sandgren and Steele. Perhaps people are systematically wrong about the regularities that pertain to the operative domains of inquiry. For example, perhaps we are wrong about Alice's partying dispositions, and the *cp* law that in fact holds in the relevant social domain entails that parties that she attends are not fun. More generally, folk science is notoriously mistaken in many ways—and presumably folk extra-special science, too—and they could lead us to assert false counterfactuals systematically, contextualism notwithstanding.

That said, I am happy to spot Sandgren and Steele this premise. Let's agree that people are generally right about such regularities, and that counterfactuals *so-interpreted* are mostly true. My disagreement will mainly concern this interpretation—or rather, *re*interpretation.

6 Virtues of the account

Their account has a number of virtues. It is based on the important idea of looking at the *function* of counterfactuals in our conversation and thinking. It seeks to preserve common sense, yet is grounded in science. As they put it, their proposal "suggests a method of harmonising constraints stemming from philosophy of language with those stemming from philosophy of science" (p. 27). They are referring to my sociological observation in my (2020) that those who are sceptical about my counterfactual scepticism tend to be philosophers of language, while those who are convinced by it tend to be philosophers of science and other naturalistically-inclined philosophers. The former often hold a view of semantics and of mental states according to which most of what we say or think must be true. Here, principles of charity and other interpretivist constraints come to the fore. On this view, I must be misinterpreting our counterfactual talk, since I attribute so much falsehood to it. I retort that science—mostly unknown by the folk—teaches us that we live in a chancy world, that chance undermines 'woulds', and that it should come as no surprise if the folk are mistaken about the counterfactuals that they utter and think. Here I tend to win over philosophers of science and other naturalistically-inclined philosophers.

Sandgren and Steele offer a way of reconciling these opposing camps. They offer an interpretation of counterfactual talk and thought according to which they can mostly

come out true; yet the interpretation essentially appeals to science, in particular the *cp* laws of the relevant domain of inquiry. Their account thus promises to explain much of our patterns of assertion and assent of counterfactuals, with good scientific credentials. Relatedly, it strives to give contextualism an objective basis, while various other accounts make context too beholden to what subjects think, independently of the facts. For example, they argue that Lewis's (2016) account has this defect; I offered my own argument for this in Sect. 3. (However, we will revisit this point in Sect. 8.2.)

Another virtue of the account is that it can be ecumenical regarding the analysis of counterfactuals themselves. Continuing my quoting of them from above, "[s]ince our proposal concerns the correct interpretation of the antecedent and not the interpretation, as it were, of the box arrow, it can be plugged into a wide range of accounts of the semantics of counterfactuals" (p. 9). They do not give truth conditions for counterfactuals, and as such they can help themselves to various such conditions. This is in contrast to Karen Lewis's account: while it is in the spirit of those of Stalnaker and David Lewis, it is revisionary, giving alternative truth conditions.

A significant point of Sandgren and Steele's account concerns the relationship between counterfactuals and laws. Much as for David Lewis it's an important priority to hold fixed the *strict* laws, on their view it's an important priority to hold fixed the *ceteris paribus* laws—and which ones are operative is context-sensitive. Their view does not need anything like Lewisian best-system laws. Their laws can be much more localised—regularities about parties attended by Alice, and so on. As I have mentioned, their view resonates with my view that the counterfactuals that are true are those for which there is nomic entailment from antecedent to consequent. It's just that on their view the entailment holds for interpreted counterfactuals, and the laws undergirding the nomic entailment may be *cp* as well as strict.

Sandgren and Steele's account also avoids some of the major costs of various other accounts. Their view upholds valid inference patterns that some other accounts apparently violate. For example, they argue that various "near-miss" accounts violate agglomeration and modus ponens, while their own upholds them—at least regarding the *reinterpreted* counterfactuals.

Of most interest to me is the way that their view avoids counterfactual scepticism. I regard their account especially as a reply to my argument from unspecific antecedents. I interpret counterfactuals at face value, and observe disparities in specificity between their antecedent and consequents. Sandgren and Steele interpret counterfactuals so that their antecedents become much more specific—so much so that they nomically entail their consequents.

So there is much to say in favour of their account. Not so much, however, for me to give up my counterfactual scepticism!

7 Unpacking the account further

Let's return to the key notions of "the operative domain of scientific inquiry" and "fitting circumstances", and their relationship to counterfactuals.

7.1 The operative domain of scientific inquiry

This presupposes that there is exactly *one* operative domain of scientific inquiry. But there might be more than one. A given counterfactual can cross different domains. "Were Alice to come to the party, it would be fun, and Schrödinger's equation would still hold". What is "the" domain of inquiry? Which *cp* laws are relevant? We might say that the most fundamental domain trumps—in this case, the microphysics of quantum mechanics. But what if the domains can't be compared with respect to fundamentality? "If the rainforests on the Amazon islands were destroyed, thousands of Brazilians would have to move." Is "the" operative domain here rainforest ecology or island biogeography?

Even focusing just on Alice's social skills, there is a version of this problem of multiple candidate domains. Suppose that while generally the life of the party, she tends to get stressed at Christmas parties, and as a result those parties are less fun; but she tends to relax at parties that Bob attends. This happens to be a Christmas party that Bob attends. What is "the" domain of inquiry, and "the" associated lawful regularity under which we should subsume this party? We have a reference class problem. This matters, because it's unclear how the counterfactual is to be interpreted by Sandgren and Steele's lights. Note that my literal interpretation of counterfactuals does not face such a problem.

There might also be *no* operative domain of scientific inquiry. It's arguably a bit of a stretch to call a party's social prospects, or Sophie's parade-viewing prospects, areas of "scientific inquiry", however "extra-special". The more localised the subject matter, the less worthy it is of the honorific "science". Similarly, the more localised the subject matter, the less 'law-like' any corresponding regularities seem to be. Yet many of the counterfactuals that we utter are *highly* localised—specifically about *this* party, *that* parade, or what have you. And sometimes we just shoot the breeze with our counterfactuals. They have truth values, but not necessarily in virtue of their place in an inquiry, however localised. To repeat, Sandgren and Steele say that "interpreting a counterfactual requires identifying a domain of inquiry" (p. 6). But what if there is no such domain to be identified—does the counterfactual have no interpretation? Again, I say that the literal one is all the interpretation it needs.

7.2 Fitting circumstances

So much for the domain. What about its "fitting circumstances"? The previous concern that there may be multiple operative domains of scientific inquiry becomes: there may be multiple sets of fitting circumstances. So there may be multiple candidates for what should be conjoined to the original antecedent. Which determines the truth value of the original counterfactual? The previous concern that there may be no operative domain of scientific inquiry becomes: there may be no associated fitting circumstances.

Sandgren and Steele later write of "reconceiving the ceteris paribus clause as implicitly referring to *a* fitting, privileged set of circumstances in which the purported regularity *does* in fact hold" (p. 13). Note the italics on "*a*": what matters is there being *a* fitting set of circumstances, not that it is *unique* ("*the* fitting set"). They pro-

vide *sufficient* conditions for the purported regularity to hold, but there will be various such sufficient conditions, some stronger than others. (At an extreme, a contradiction provides sufficient conditions for anything.) But now the problem of multiple interpretations is exacerbated, and with it the threat of different truth values, depending on which interpretation is adopted. And the fact that the conditions are merely sufficient, not necessary and sufficient, should alert us that the reinterpreted counterfactual is not equivalent to the original.

8 More specific objections

So far I have raised some concerns at a relatively high level, about the formulation of Sandgren and Steele's view. They might be regarded as invitations for them to provide further details, or to tweak their view in minor ways that keep its spirit intact. I now turn to some more specific objections.

8.1 Weakening counterfactuals

My general arguments against antecedent-contextualism of Sect. 4 come into play against Sandgren and Steele's particular version of it. Their contextualism weakens counterfactuals, losing entailments that they originally have. Consider again:

(FUN) Were Alice to come to the party, it would be fun.

Sandgren and Steele interpret this as

(FITTING-FUN) 'Were Alice to come to the party *and fitting circumstances for the operative domain of inquiry held*, it would be fun.'

Suppose that in the actual world, Alice comes to the party, there is a fire, and it is not fun. I say that this falsifies FUN. But it does not falsify FITTING-FUN, which makes no commitment to what happens if fitting circumstances *do not* hold (as is the case in my story). FUN entails that my story is false; FITTING-FUN does not entail this.

Moreover, the following 'might not' counterfactual conflicts at least pragmatically, and I say semantically, with FUN:

Were Alice to come to the party, it *might not* be fun.

But it does not conflict at all with FITTING-FUN. Indeed, it is even conversationally implicated by FITTING-FUN. The qualification about fitting circumstances holding invites one to imagine their *not* holding. It would be uncooperative for Gricean reasons to add this qualification if it made no difference—if the party would be fun either way. So the hearer of this utterance could reasonably infer that it *does* make a difference: the party with Alice might not be fun, namely if fitting circumstances did *not* hold. Thus, FITTING-FUN does not capture the true strength of FUN, which needs to be interpreted literally.[7]

[7] In the background are broader issues concerning standards of precision, loose talk, and the extent to which strict interpretations of the things that we say are appropriate (see e.g. D. Lewis 1979, 1989; Bolinger and

Here's another argument that Sandgren and Steele's interpretation weakens counterfactuals. Suppose we are in a domain of inquiry D, and we utter a contingent counterfactual with antecedent A, and consequent 'D's fitting circumstances hold':

> If A were the case, D's fitting circumstances would hold. (*)

Offhand, it seems that Sandgren and Steele interpret this as:

> If A were the case and D's fitting circumstances held, D's fitting circumstances would hold.

That is a logical truth, since the antecedent entails the consequent. So offhand, (*) is interpreted by them as a logical truth. For example, suppose our domain of inquiry is Alice's social impact, and our contingent counterfactual is:

> If Alice were to come to the party, the Alice-social-impact domain of inquiry's fitting circumstances would hold. (**)

Offhand, it seems that Sandgren and Steele interpret this as:

> If Alice were to come to the party and the Alice-social-impact domain of inquiry's fitting circumstances held, the Alice-social-impact domain of inquiry's fitting circumstances would hold.

This is a logical truth. I don't think that (**) is true, but in any case *nobody* should think that it is logically true. A substantive, contingent counterfactual has been weakened to the point of vacuity.

Now, they may be able to answer this objection.[8] Perhaps drawing attention to D itself *changes* the context, so that it is no longer the operative domain. In that case, the fitting circumstances are no longer those of D, but those of the new domain—say, D_2. The interpretation of (*) has the form:

$$(A \,\&\, F_{D_2} \,\Box\!\!\rightarrow\, F_D),$$

which is not a logical truth.

At first that might seem a little odd: wouldn't drawing attention to D, or to its fitting circumstances, only reinforce D's status as the operative domain? Not necessarily. Compare: it may be true that when you first uttered "if the cup were released, it would fall", your context was that of street talk (and according to the contextualist, the counterfactual is *true* in that context). But now suppose I draw attention to that *being* the context:

> if the cup were released *in the context of street talk*, it would fall.

I am making a *philosophical* move, thereby shifting the context away from street talk to *philosophy* (and according to the contextualist, the counterfactual may well be *false* in that context). So perhaps, like a chameleon that changes colour when attention is

Footnote 7 continued
Sandgren 2020). I can happily allow laxer interpretations of various other parts of our language; all I insist on here is that *counterfactuals* require strict interpretations, for the reasons I have given.

[8] In what follows I am grateful for discussion with Alex Sandgren.

drawn to it, the context that is operative in a counterfactual's antecedent changes when attention is drawn to it.

So perhaps this is just another invitation to Sandgren and Steele to say more about how context determines a domain of inquiry, and how it may be shifted—in particular, when a counterfactual itself draws attention to that very domain.

The next objections I consider to be more serious.

8.2 Underminers outside the domain of inquiry

Regarding the counterfactual, "Were Alice to come to the party, it would be fun", Sandgren and Steele write:

> the relevant domain of inquiry is plausibly one concerning Alice's impact on others in ordinary social milieu. Here, circumstances in which the party is ruined by a house fire or a plumbing problem, are arguably not fitting, *despite, perhaps, being relatively common events*… This serves to highlight one point …: *fitting circumstances are not primarily identified by their high probability*…. (pp. 6–7, my italics.)

Consistently with this quote and the story, suppose that in fact the chance of the party being ruined by a house fire on the night in question is extremely high, perhaps even 1. "Were Alice to come to the party, it would be fun" really seems *false* to me, and not because of my radical views; I think *that* judgment is commonsensical. After all, we are supposing that the party is almost guaranteed *not* to be fun, despite Alice's presence. But the counterfactual comes out *true* on Sandgren and Steele's view.

The problem here is that for any putative domain of inquiry, it might be highly unlikely that circumstances would be fitting for that domain. Defeaters from outside that domain might encroach on the truth of counterfactuals uttered in the domain. The truth of the 'Alice' counterfactual is hostage to the whole set-up, not just the facts about Alice's place in her social milieu. And so it goes with counterfactuals in general. Conjoining to a counterfactual's antecedent that the operative domain's circumstances are fitting typically makes a difference to the truth conditions of the counterfactual: it rules out such defeaters (even probable ones), while the original counterfactual did not (especially probable ones).

Or consider:

> Were Alice to come to the party, it *might not* be fun. ☹

I insist that's true. I could go on to list various reasons, coming from various domains, why it might not be fun: there might be a house fire, a plumbing problem, Alice might be in a bad mood, her social interactions with others might turn sour, a freakish statistical mechanical anomaly might lead to everyone at the party suffocating, a freakish quantum mechanical anomaly might lead to Alice quantum tunnelling away. Never mind the reasons, and the operative domain—the 'might not' just depends on the *existence* of such reasons, without any need to spell them out.[9]

[9] Thanks here to Justin D'Ambrosio.

Now, Sandgren and Steele agree with me that 'might nots' undermine 'woulds'. In order to reconcile the truth of ☹ with that of the original 'would' counterfactual, they could insist that the 'might not' consequent changes the context. For example, perhaps it prompts us to become more imaginative at coming up with ways in which a party with Alice could fail to be fun. But then my objection that Karen Lewis's account is too subjective comes back to target Sandgren and Steele, despite their claiming an advantage on that score. Again, it sounds like the more ignorant or unimaginative you are, the easier it is for your counterfactuals to come out true. Moreover, this only compounds the problem, emphasised earlier, that there could be multiple operative domains of inquiry. Our new context may open the floodgates to the domains of combustion science, fluid dynamics, individual psychology, social psychology, statistical mechanics, quantum mechanics, and more. Will "the" operative domain of scientific inquiry please stand up?! But never mind our new context; these undermining possibilities were there all along, whether or not we attended to them. They rendered false the original counterfactual, "Were Alice to come to the party, it would be fun", without any help from us.

Consider another example that Sandgren and Steele discuss: "Were you to drink heavily tonight, you would be hungover tomorrow morning" (p. 21). We imagine that in fact you do drink heavily tonight, but as it happens, you die before tomorrow morning. In their example, the sun expands and engulfs the earth. The example need not be so far-fetched. We may suppose that you die for some other reason—for example, from alcohol poisoning! I think that the counterfactual is *false* in this story: its antecedent is true, its consequent is false. Nothing saves the counterfactual—not the operative domain of inquiry, not the fitting circumstances.

There is an instructive analogy to future predictions, expressed using 'wills'. This should come as no surprise, given that it is widely thought that *will* and *would* share morphology (see e.g. Abusch 1997; Copley 2009). For example, suppose you drink heavily, and you say:

I will be hungover tomorrow.

In fact, you're not hungover tomorrow (because you die first). I think that what you said is *false*. It doesn't get you off the hook if you say (in heaven?!): "But what I really meant was 'I will be hungover *assuming fitting circumstances*'—and as it happened, circumstances were not fitting". No, your prediction was false—period. This analogy is especially pressing when the antecedent of the original counterfactual is realised, so that the counterfactual effectively makes a prediction of what will happen. In fact, you do drink heavily. In fact, you are not hungover the next day. You don't get off the hook: what you said is false—period.

Relatedly, when pressed on an utterance of a counterfactual, it is often appropriate to hedge. You say: "If Alice were to come to the party, it would be fun". I press you: "Really? What if there were a house fire or a plumbing problem, or what if Alice were in a bad mood …?" You then hedge: "Alright—I should have said 'if she were to come to the party *under normal circumstances*…'" This seems to display your recognition that your original statement, without the qualification, was hyperbolic.

Hedging gets you off the hook that you were previously on.[10] There are stronger and weaker qualifications that you can add to the antecedent of a counterfactual, much as there are stronger and weaker qualifications that you can attach to a 'will' statement. But the counterfactual itself does not have a qualification already built in as part of its meaning, much as the 'will' statement does not.

8.3 More questions about the reinterpretation

So I wonder why counterfactuals require reinterpretation, but predictions about the future do not, especially when they coincide in cases of true antecedents. Indeed, I wonder what other parts of our language require such reinterpretation. Are counterfactuals loners on Sandgren and Steele's view? If so, that cries out for explanation: why them, and nothing else?

A related point: granted, it's a cost of my view that I attribute so much falsehood. But I think it is a cost of Sandgren and Steele's view that it requires so much reinterpretation! Again, "interpreting a counterfactual requires identifying a domain of inquiry" (p. 6). I think this portrays understanding a counterfactual as harder than it is—especially when multiple candidates for a domain of scientific inquiry present themselves, or none. This seems too demanding of competent speakers and listeners, including children, who may have little idea what such a domain is. My literal interpretation requires us just to understand its antecedent, consequent, and the counterfactual connective.

To be sure, the counterfactuals that we utter may *convey* further information—for example by conversational implicature. That further information may indeed be captured along the lines of Sandgren and Steele's interpretation (and understanding such implicatures may be harder than understanding the counterfactuals themselves). But that is a matter of pragmatics rather than semantics, and the *truth values* of counterfactuals is a matter of semantics. After all, that's the subject of our debate about counterfactual scepticism. Compare: in a letter of recommendation for a philosophy position, a referee says: "Bloggs is punctual and has nice handwriting." This conveys the further information that Bloggs is a poor philosopher. But that is an implicature of what was said—a matter of pragmatics rather than semantics.

And why do counterfactuals require Sandgren and Steele's particular interpretation—adding an extra conjunct that the circumstances are fitting to the antecedent—rather than some other? I think a more plausible interpretation would add that extra conjunct *to the consequent*. We may interpret 'if p were the case, then q would be the case' as: 'if p were the case, then circumstances would be fitting and (for that reason) q.' For example, we may interpret 'if the cup were released, it would fall' as: 'if the cup were released, circumstances would be fitting for the behaviour of macroscopic objects (including no quantum mechanical or statistical mechanical anomalies) and (for that reason) it would fall'. In explicitly claiming that the consequent would hold, the counterfactual implicitly claims that circumstances amenable to that would also hold—this is one of its entailments. To that extent, the counterfactual

[10] To be sure, the original statement may well have been perfectly reasonable to say, when you said it. But that's a matter of pragmatics: a little hyperbole may well be acceptable in a normal conversation.

sticks its neck out (too far, I say). Now, since this is an entailment of the counterfactual rather than something that needs to be added to its literal content, I am happy to stay with that content. But at least I think this interpretation gets the strength of the original counterfactual right; indeed, it is a way of seeing just how strong its commitments are (too strong to be true, I say). By contrast, Sandgren and Steele's interpretation weakens the original counterfactual, strengthening the antecedent while holding fixed the consequent. As such, the counterfactual's neck is pulled in, losing commitments that it had.

8.4 Disagreement, retraction

On the street, you say:

> If the cup were released, it would fall. (↓)

Sitting in the philosophy classroom that overlooks the street, I overhear you. I consider surprising possibilities in which the cup is lifted by a sudden updraft, and I beg to differ:

> *No*, it's not the case that if the cup were released, it would fall.

This seems like a paradigm case of *disagreement.* I take myself to be *correcting* you. You then walk into the classroom, and I draw your attention to these surprising possibilities. Wise person that you are, you follow my lead:

> "*No*, it's not the case that if the cup were released, it would fall", you now say.

This seems like a paradigm case of *retraction*. You're disagreeing with your former self.[11]

But I take it that according to Sandgren and Steele's contextualism, there is no genuine disagreement, nor genuine retraction: the domain of inquiry has changed from the street to the classroom. In their words:

> … it often seems reasonable for speakers to reconsider a would-counterfactual's truth when it is pointed out that there are relevant possibilities in which the antecedent holds and the consequent does not. The contextualist's line is that there is a subtle shift in context in these cases such that apparently contradictory statements are both true. On our specific proposal, what is going on is that the interpreter is prompted to entertain a subtle shift in domain. (p. 7)

I am not convinced that that's what's going on here. In both contexts, the domain of inquiry could be the behaviour of macroscopic objects. It's just that in the classroom, *we conduct that inquiry more carefully*. Sudden updrafts lifting objects are familiar processes in the macroscopic world; they are fitting circumstances for this domain. We only need to be reminded of them.

Now suppose that instead, in the classroom I draw attention to weird quantum mechanical and statistical mechanical possibilities that undermine ↓ (as I am also wont to do). Again, I disagree with what you said on the street: "*No*, it's false that if I

[11] This argument from disagreement and retraction is reminiscent of and inspired by MacFarlane's (2005, 2014) anti-contextualism arguments regarding expressions for which he favours relativism.

were to release the cup, it would fall." And when you join me in the classroom, you utter the same sentence, expressing your retraction. Now, Sandgren and Steele may say that I have changed the domain of inquiry, and that once you have entered the classroom, you accommodate that change. But rather than disagreeing with or retracting what you said on the street, why don't we say that we have simply *changed the topic*? Why don't we say: "Remember what was said on the street? It was *true*, because the domain of inquiry then was the normal behaviour of macroscopic objects"?

Suppose that I am in Taipei, where it is hot, and I call my wife in Canberra, where it is freezing. I say: "It is hot here". She replies: "*Yes*, I know—I saw that on the internet. But it's not hot *here*; in fact, it's freezing." And when I return to Canberra the following day, I say: "It's not hot here. But remember what I said yesterday? It was true." There is no disagreement, and no retraction. It is obvious that the topic changed from the weather in Taipei to the weather in Canberra. I take this pattern to be standard for context-sensitive language such as "here". I think it has no parallel for counterfactuals. Just try it:

> *No*, it's false that if I were to release the cup, it would fall. But remember what was said on the street? It was *true*. #

I take this to be strong evidence that counterfactuals are not context-sensitive in the way that Sandgren and Steele posit.

Moreover, why should you accommodate my change of context when you enter the classroom. Why don't you just stand your ground?: "Thanks, but I'd rather stick to talking about the normal behaviour of macroscopic objects." And if you do accommodate the new domain of inquiry, why isn't it easy for you to switch back to the previous domain?: "Thanks, I enjoyed that little detour. Now, let's get back to what I was talking about: the normal behaviour of macroscopic objects! And as I said, if I were to release the cup, it would fall—just as that domain of inquiry tells us." But I claim that it is *not* easy.[12]

I have a straightforward explanation of such disagreement and retraction phenomena. No wonder I disagree with what you said on the street: I think it's *false*. No wonder you retract it when you join me in the classroom: you come to agree with me that it's *false*.

9 Avenues for future research?

So I invite Sandgren and Steele to respond to these thoughts, and I will be delighted if they prompt further developments of their view.

For example, the objection from disagreement and retraction may be regarded as an invitation for them to say more about what each of those consist in. I have been

[12] K. Lewis (2016) distinguishes between "forced" and "natural context shifts", and she acknowledges that in "forced shifts, those induced by the skeptic, if the skeptical possibility is accommodated, it is difficult to downshift to a less strict context" (p. 304). So she would presumably agree with my "*not* easy" claim here. However, she also offers two cases in which such "downshifting" seems to be felicitous. I think that "forced shifts" may draw our attention to nomic possibilities that are and always were relevant—but this is too big a topic to pursue adequately here.

assuming that they involve disparate attitudes to a fixed proposition. My concern was that according to Sandgren and Steele's contextualism, the content is meant to change, and so disagreement and retraction do not get a foothold. Instead, we have two people talking past each other, or one person expressing a different proposition from their younger self. But perhaps there's no fixed proposition that one of us believes and the other does not, or that one time slice of a single person believes and the other does not. For example, the disagreement could be over what the context *should be*, perhaps in order to further other conversational goals. It might thus be metalinguistic, something to be negotiated. I don't think that's right—the "*No*" in my imagined dialogues attaches to the first-order contents—but it might be worth exploring.

Regarding the problem of multiple operative domains of scientific inquiry for a given counterfactual, one option might be to *supervaluate*. Here's one way this might go. Consider each of the candidates for the domain (and its associated *cp* laws), and determine whether the counterfactual comes out true on that domain. If it comes out true on all of them (supertrue), it's true simpliciter; if it comes out false on all of them (superfalse), it's false simpliciter. If it comes out true on some and false on others, the counterfactual is indeterminate.[13] But this no longer promises to be as decisive a response to scepticism, for now the threat is that most counterfactuals are *indeterminate*. To be sure, that is not to claim that most counterfactuals are false, as I do; still, it does entail that most counterfactuals are *not true*. This still flies in the face of common sense. Just try telling a person on the street that it is *not true* that if they were to release a cup, it would fall, and see what looks you get!

Another option might be to *subvaluate*. If the counterfactual comes out true on *any* of the candidates for the domain, it is true simpliciter. Now the worry is this will yield too many true counterfactuals, even by the lights of common sense. So these options do not seem to find the 'sweet spot' of rendering true the counterfactuals that we intuitively regard as such, but no more. Sandgren and Steele may well be able to do better.

I also raised the problem of there being *no* operative domain of scientific inquiry for a given counterfactual—the domain of inquiry is too specific to count as scientific, or we are just shooting the breeze, and not really conducting an inquiry at all. In that case they might allow that there are no fitting circumstances, and hence nothing to conjoin to the counterfactual's antecedent. Then we are just left with the literal interpretation of the counterfactual, the one that I prefer. However, this is music to my sceptical ears. I will simply deploy my argument from unspecific antecedents, and insist that the counterfactual is false. Again, perhaps Sandgren and Steele can do better. More generally, they might like to say more about what determines context, and its updating.

This prompts me to return to my aerial view of counterfactual contextualism in the first half of this paper, having spent most of the second half in the trenches with Sandgren and Steele. I often find contextualism about various matters to be a bit of a magic wand. Somehow—we are not told how—context works its magic to make sure that everything in our semantics works out as we wish. I am no Popperian, but I start to wonder what it would take to falsify such a theory. On the other hand, laying down

[13] This is somewhat analogous to Stalnaker's (1981) supervaluational approach to conditionals, when there are multiple candidates for the selection function that selects the closest antecedent-world.

rules on what determines a context and what context contributes to truth conditions is a difficult and risky enterprise, in most cases where contextualism is invoked. I expect that to be especially so for counterfactuals, as they are especially hard to pin down. Now the worry is that the resulting theory is not only falsifiable, but *false* (like the counterfactuals themselves!). But it's an enterprise well worth undertaking.

Acknowledgements I thank particularly Hannes Leitgeb, Cory Nichols, Alex Sandgren, Katie Steele, Jeremy Strasser, the audience at The Fourth Taiwan Philosophical Logic Colloquium (TPLC 2018), National Taiwan University, and two anonymous referees for *Synthese* for very helpful comments. Special thanks to Justin D'Ambrosio for many valuable conversations and feedback.

References

Abusch, D. (1997). Sequence of tense and temporal de re. *Linguistics and Philosophy, 20*(1), 1–50.
Bolinger, R. J. & Sandgren, A. (2020). Strictly speaking. *Analysis*, *80*(1), 3–11. https://doi.org/10.1093/analys/anz017.
Copley, B. (2009). *The semantics of the future*. New York: Routledge.
DeRose, K. (1999). Can it be that it would have been even though it might not have been? *Philosophical Perspectives, 13*(Epistemology), 385–413.
Hájek, A. (2014). Probabilities of counterfactuals and counterfactual probabilities. *Journal of Applied Logic, 12*(3), 235–251.
Hájek, A. (2020). *Most counterfactuals are false*.
Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, *31*(2), 231–270.
Leitgeb, H. (2012). A probabilistic semantics for counterfactuals. *The Review of Symbolic Logic, 5*(1), 26–121.
Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8*(3), 339–359.
Lewis, D. (1989). Dispositional theories of value. *Aristotelian Society Supplementary, 63*(1), 113–137.
Lewis, K. (2016). Elusive counterfactuals. *Noûs, 50*(2), 286–313.
MacFarlane, J. (2005). The assessment sensitivity of knowledge attributions. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 1, pp. 197–233). Oxford: Oxford University Press.
MacFarlane, J. (2014). *Assessment sensitivity: Relative truth and its applications*. Oxford: Oxford University Press.
Sandgren, A., & Steele, K. (2020). *Levelling counterfactual scepticism*. Retrieved April 17, 2020 from https://alexandersandgren.files.wordpress.com/2020/04/levelling-counterfactual-scepticism-17-04-2020.pdf.
Stalnaker, R. C. (1981). A defense of conditional excluded middle. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs* (pp. 87–104). Dordrecht: D. Reidel.
Starr, W. (2019). Counterfactuals. In E. N. Zalta (Ed.) *The Stanford encyclopedia of philosophy* (Fall 2019 Edition). https://plato.stanford.edu/archives/fall2019/entries/counterfactuals/.