Chancy-but-true Counterfactuals

Joshua Edward Pearson
pearsonj@mit.edu

Abstract. Taking inspiration from recent developments in epistemology, I outline and defend a theory of counterfactuals that (i) avoids counterfactual skepticism; (ii) maintains a tight connection between the truth of a counterfactual $A \mathrel{\Box\!\!\rightarrow} C$ and the relevant chance of C conditional on A; yet (iii) validates attractive principle like AGGLOMERATION, by which $A \mathrel{\Box\!\!\rightarrow} C_1$ and $A \mathrel{\Box\!\!\rightarrow} C_2$ jointly entail $A \mathrel{\Box\!\!\rightarrow} C_1 \& C_2$.

Were I to drop my mug, it would hit my desk. At the same time, there are various bizarre scenarios in which dropping my mug does not result in it hitting my desk — such as the scenario in which my mug quantum tunnels through it. And according to our best scientific theories, the chance of such scenarios occurring, conditional on me dropping my mug, is above zero. This raises a puzzle. How can it be that my mug *would* hit my desk were I to drop it, if there's some chance it quantum tunnels through my desk instead? These claims seem in tension.

'Pessimists' argue that this tension is nothing short of inconsistency: it's simply false that my mug would hit my desk if there's some chance it would quantum tunnel through it. And since our world is thoroughly chancy — for any apparently mundane situation there is a non-zero chance of some bizarre event occurring — Pessimists can argue from this tension to 'Counterfactual skepticism': the view that most of the counterfactuals we assert in ordinary life are false (Hajek 2014).

In contrast, 'Optimists' hope to hold onto the truth of ordinary counterfactuals even in the face of these claims about chance.[1] On these views, the claim that there is some chance the mug quantum tunnels through my desk

[1] E.g. (D. Lewis 1979), (Hawthorne 2005), (Williams 2008), (K. Lewis 2016) and (Boylan 2023).

is not inconsistent with the claim that it would, in fact, hit my desk. Though the specifics of these optimist views differ, one feature they all tend to share is that any connection between counterfactual truth and chance is severed. That is, these views tend to give truth conditions for the counterfactual $A \mathrel{\Box\!\!\rightarrow} C$ that are completely independent from the chance of C conditional on A. This is hard to swallow: even if the truth of $A \mathrel{\Box\!\!\rightarrow} C$ is compatible with a *minor* chance that not-C conditional on A, it is still plausibly not compatible with a *substantial* chance that not-C conditional on A.

I'm an optimist. But I'm also an optimistic optimist. Here, I hope to defend a non-skeptical theory of counterfactuals that embraces the thorough chanciness of our world without severing counterfactual truth from chance. The most obvious way to do this would be to endorse 'Counterfactual Lockeanism' — the view that $A \mathrel{\Box\!\!\rightarrow} C$ is true iff the chance of C given A is sufficiently high. However, along with other many philosophers, I am dissatisfied with Counterfactual Lockeanism and particularly it's consequence that the principle 'Agglomeration' fails — which says that if $A \mathrel{\Box\!\!\rightarrow} C_1$ and $A \mathrel{\Box\!\!\rightarrow} C_2$ are are both true then $A \mathrel{\Box\!\!\rightarrow} C_1 \& C_2$ is also true.[2] By drawing inspiration from recent work in epistemology,[3] I'll defend an alternative chance-based yet optimistic view of counterfactuals that avoids this pitfall.

§1 demonstrates why most optimistic theories must sever counterfactual truth from chance; §2 sketches my alternative theory to Counterfactual Lockeanism; §3 responds to objections. In §4, I conclude by considering an alternative reading of this paper. Some have suggested to me that the resources I use might be better used to construct a theory of when counterfactuals are knowable or assertable, rather than when they are true. I agree that such an account of counterfactual knowledge or assertability would be attractive. However, I argue that it would be difficult to explain how such an account of counterfactual knowability or assertability could be correct unless we also accept the theory of counterfactual truth I defend here.

[2]E.g. (Hawthorne 2005), (Hajek 2014), (K. Lewis 2016) and (Boylan 2023).

[3]In particular, (Lin and Kelly 2012), (Leitgeb 2014), (Holguín 2022), (Dorst and Mandelkern 2022), (Goodman and Salow fc) and (Pearson ms).

1 Severing Counterfactual Truth from Chance

As I said above, I'm an optimist. I endorse:

Optimism
Many counterfactuals asserted in ordinary discourse are true.

I also said that most views defended so far that endorse Optimism sever counterfactual truth from chance. But what does this 'severance' amount to, and why are optimists often forced to it?

As I'm understanding it, a view severs counterfactual truth from chance to the extent that it invalidates principles of the following form, where *Ch* is a measure of objective chance:

Chance-Truth Link
If $Ch(C \mid A) \geqslant T$ then $\neg(A \mathrel{\Box\!\!\to} \neg C)$.

That is, if the chance that C conditional on A is sufficiently high, then the counterfactual *were A the case, C wouldn't be the case* is false. Alternatively: if it's true that *were A the case,* $\neg C$ *would be the case*, then the chance that C conditional on A must be sufficiently low. Which chance function is the relevant one? I'll follow (Hajek ms) and assume that the relevant chance function is that of the actual world, shortly before, but not too shorty before, it was settled whether A.[4]

Chance-Truth Link is intuitively plausible when T is fairly high — even just $T = \frac{1}{2}$ will do. Suppose I don't flip this fair coin, and claim "had I flipped it, it would have landed heads." This claims seems false, and Chance-Truth Link explains why: it's false to say it would land heads, as the relevant conditional chance of it landing tails is significantly high.

Optimist views tend to sever counterfactual truth from chance because they are forced in invalidate Chance-Truth Link, at least when $T < 1$. Why's

[4]Strictly speaking, this means I should index the chance function *Ch* to the antecedent of the counterfactual being considered, A. I'll (harmlessly, I hope) ignore this complication here but plan to pay closer attention to it in further research.

that? This is best illustrated through the following paradox, based on one discussed in epistemology by (Dorr, Goodman, and Hawthorne 2014).[5] For simplicity, I'll assume that $T = \frac{1}{2}$ in Chance-Truth Link, but the paradox can arise even if T is set to be higher.

Consider the following principle:

Possibility Preservation

If $A \boxright C$ and $\neg(A \boxright \neg B)$ then $(A\&B) \boxright C$.

Possibility Preservation is intuitively plausible. Suppose that, were I to drop my mug, it would hit my desk. Suppose further that it's false that, were I to drop my mug, I wouldn't curse.[6] It seems to follow that were I to drop my mug and curse, it would hit my desk.[7]

The paradox is that Optimism, Chance-Truth Link and Possibility Preservation cannot all be true. To see this, consider the following case — a counterfactual analogue of a case considered by (Dorr, Goodman, and Hawthorne 2014):

> **Leafy.** It's the first day of fall and Leafy the maple leaf has just shed. I assert: "Had Leafy not shed today, it still would have shed by the first day of Spring." As it turns out, the physics of leaf-shedding works roughly as follows: for any particular morning on which a leaf has not yet shed, there is a 0.5 chance it still won't have shed the morning after.

[5] Leitgeb (2013) also outlines what I take to be, in essence, the same paradox concerning counterfactuals. What he calls 'Rational Monotonicity' I have below called 'Possibility Preservation'. I outline with this version of the paradox simply because I find it easier to grasp than Leitgeb's.

[6] Intuitively, the falsity of this counterfactual is equivalent to the truth of the following might counterfactual: *were I to drop my mug, I might curse*. But assuming this would be to assume the controversial principle Duality, an issue I'd rather not get into here.

[7] Further, as (Boylan and Schultheis 2021) observe, Possibility Preservation is a natural weakening of the contested Antecedent Strengthening principle. (See (D. Lewis 1973), (Fintel 2001), (Moss 2012) and (K. Lewis 2016) for considerations for/against Antecedent Strengthening.) Antecedent Strengthening says that, if $A \boxright C$ is true, then strengthening the antecedent with *any other* proposition, B, will preserve truth — that is, $A\&B \boxright C$ must also be true. In contrast, Possibility Preservation only says that truth is preserved by such a strengthening if B is possibly true by A's lights; that is, if it's not the case that $A \boxright \neg B$.

If there are any true ordinary counterfactuals despite the thorough chanciness of our world, the asserted counterfactual *had Leafy not shed today, it would have shed by Spring* can be true in **Leafy**. Optimism thereby commits us to the truth of:

(1) Had Leafy not shed today, it would have shed by Spring.

By the description of the case, we know that had Leafy not shed today, there would have been a 0.5 objective chance it wouldn't have shed tomorrow, either. Hence by Chance-Truth Link:

(2) Not-(Had Leafy not shed today, it would have shed tomorrow).

But now, by Possibility Preservation, (1) and (2) together entail:

(3) Had Leafy failed to shed both today and tomorrow, it would have shed by Spring.

So far, no problem. (3) is plausible. The problem is that we can repeat this reasoning over and over. By the description of the case, conditional on Leafy not shedding today or tomorrow, there's a 0.5 chance it wouldn't have shed on the day after, either. Hence, by Chance-Truth Link:

(4) Not-(Had Leafy failed to shed both today and tomorrow, it would have shed the day after).

So by again applying Possibility Preservation to (3) and (4) we can infer:

(5) Had Leafy failed to shed today, tomorrow, and the day after, it would have shed by Spring.

Perhaps you can now see where this is going. By constantly reapplying the above reasoning, we can eventually infer the contradictory:

(6) Had leafy failed to shed today, tomorrow, the day after, the day after that, ...and on the first day of Spring, it would have shed by Spring.

Upshot: If we like Optimism, we have to give up one of Chance-Truth Link or Possibility Preservation.

The majority of endorsed optimist views entail Possibility Preservation. For example, consider the influential and widely endorsed similarity accounts of counterfactuals from (D. Lewis 1973) and (Stalnaker 1968), by which $A \boxright C$ is true iff all of the closest A-worlds are C-worlds. This means that if $\neg(A \boxright \neg B)$, then some of the closest A-worlds are B-worlds. So if $A \boxright C$ and $\neg(A \boxright \neg B)$, all of the closest $A\&B$-worlds are C worlds, giving us $(A\&B) \boxright C$, as per Possibility Preservation.[8] The majority of optimists must therefore reject Chance-Truth Link, thus severing counterfactual truth from chance.

Perhaps it can eventually be argued that this severance is a necessary cost for anyone who wants to endorse Optimism. But I think you should be quite hesitant to pay this cost. For if you give up Chance-Truth Link for both Optimism and Possibility Preservation, you are committed to the following absurd consequence regarding **Leafy**. Consider counterfactuals of the following form:

(7) Had Leafy not shed within k days from today, Leafy would have shed within $k+1$ days from today.

Such counterfactuals sound just as false as the claim that, had I flipped this fair coin, it would have landed heads.[9] Yet endorsing both Optimism and Possibility Preservation means you are committed to there being some k for which the corresponding counterfactual of form (7) is true. For if all

[8](Boylan and Schultheis 2021) avoid validating Possibility Preservation by denying that similarity to the actual world is always a comparable notion: sometimes w is no more similar to the actual world than w' is, nor vice versa, nor are they equally as similar. Their motivating counterexample to Possibility Preservation is very different to **Leafy**; I leave it for further research to explore how my work interacts with theirs.

[9]You might be more inclined to think that counterfactuals like (7), rather than being false, are instead simply unknowable. While I am slightly sympathetic to this reaction, I wish to direct readers who are also sympathetic to it to my conclusion, §4. There, I note that we can construct an epistemic variant of paradox of this section, which only depends on claims like (7) being unknowable rather than false. Yet, I argue, this version of the paradox is most satisfactorily resolved if we accept the theory I am going to defend in §2.

such counterfactuals were false, these false counterfactuals plus Possibility Preservation and Optimism would be enough to derive the absurd (6).[10]

I therefore contend that, before we reject Chance-Truth Link, a far more thorough investigation of optimistic views which instead reject Possibility Preservation is called for. The next two sections provide such an investigation. I argue that, in fact, there are highly attractive optimistic views that preserve Chance-Truth Link at the cost of Possibility Preservation.

2 How to Have Your Chancy Cake and Eat It Too

2.1 Counterfactual Lockeanism

How might an optimist like me hope to preserve Chance-Truth Link? An obvious route is to endorse:[11]

Counterfactual Lockeanism
$A \mathrel{\Box\!\!\rightarrow} C$ iff $Ch(C \mid A) \geqslant T$. $(\frac{1}{2} \leqslant T < 1)$.

For Counterfactual Lockeanism entails Chance-Truth Link.[12] In line with the above paradox, it invalidates Possibility Preservation.[13] Counterfactual Lockeanism therefore constitutes a natural resolution to the above paradox that avoids severing counterfactual truth and chance. So why don't optimists like it?

[10]There are ways for optimists who like Possibility Preservation to soften the blow of accepting the truth of (7). For instance, on a view like that in (K. Lewis 2016), it might be that while such true counterfactuals exist, they are elusive in that trying to assert them will inevitably shift the context, taking us to a context where (7) expresses a distinct counterfactual that's false. I'll leave exploring this kind of response for future research.

[11](Leitgeb 2012) endorses something like Counterfactual Lockeanism, at least as an account of when a counterfactual is *approximately* true.

[12]Assuming, as is plausible, that $A \mathrel{\Box\!\!\rightarrow} C$ entails $\neg(A \mathrel{\Box\!\!\rightarrow} \neg C)$. (The reverse direction, a form of the "conditional excluded middle" principle, is more controversial.)

[13]A toy model: suppose coin c is fair, A is the proposition *c is flipped three times*, B the proposition *The first flip lands heads* and C the proposition *Some flip lands tails*. If $T = \frac{7}{8}$, then Counterfactual Lockeanism predicts a counterexample to Poss-Str as $Ch(C \mid A) \geqslant \frac{7}{8}$, meaning $A \mathrel{\Box\!\!\rightarrow} C$, yet $Ch(\neg B \mid A) < \frac{7}{8}$, meaning $\neg(A \mathrel{\Box\!\!\rightarrow} \neg B)$, and $Ch(C \mid A\&B) < \frac{7}{8}$, meaning $\neg((A\&B) \mathrel{\Box\!\!\rightarrow} C)$.

The most commonly cited reason concerns the following principle:[14]

Agglomeration
If $A \mathrel{\Box\!\!\rightarrow} C_1$ and $A \mathrel{\Box\!\!\rightarrow} C_2$ then $A \mathrel{\Box\!\!\rightarrow} C_1 \& C_2$.

Counterfactual Lockeanism invalidates Agglomeration. For instance, let $T = 0.8$ and consider the proposition *I threw a birthday party* (B). Assume that the chance my sister came (S), conditional on B, is 0.8. Assume also that the chance my mother came (M), conditional on B, is 0.8. Assume finally that my sister and mother are at least not *completely* dependent upon one-another: conditional on B, there is some chance one came without the other. Given these assumptions, $Ch(S \mid B) = 0.8$, $Ch(M \mid B) = 0.8$, yet $Ch(S\&M \mid B) < 0.8$. Counterfactual Lockeanism therefore predicts that $B \mathrel{\Box\!\!\rightarrow} S$ and $B \mathrel{\Box\!\!\rightarrow} M$ are true yet $(B \mathrel{\Box\!\!\rightarrow} S\&M)$ is false, contra Agglomeration.

This is a problem as Agglomeration strikes most as extremely plausible. Note that Agglomeration is even more plausible than analogous 'closure' principles discussed in epistemology which claim that one is justified in believing p and in believing q only if one is justified in believing $p\&q$. Many do not find it too hard (though I personally have trouble) to get into the frame of mind whereby one can justifiably believe, of any particular invitee, that they will come to the party, yet at the same time be justifiably unsure as to whether all invitees will come to the party.[15] It is much harder to get into the analogous frame of mind for counterfactuals. Suppose I didn't throw a party, but that, for any invitee X, had I thrown the party, X would have come. If that's really the case — for every invitee X, they would have come had I thrown the party — how could it nevertheless be that some invitee might not have come?

Nevertheless, some I have spoken to in conversation are willing to bite the bullet on failures of Agglomeration. So, let me now outline one further issue with Counterfactual Lockeanism which, to my knowledge, has not

[14] See, for instance, (Hawthorne 2005), (Hajek 2014), (K. Lewis 2016) and (Boylan 2023).
[15] As (Makinson 1965) famously argues.

yet been discussed in the literature.[16]

Consider the following natural weakening of Possibility Preservation:

Would Preservation
If $A \mathrel{\Box\!\!\rightarrow} C$ and $A \mathrel{\Box\!\!\rightarrow} B$, then $A\&B \mathrel{\Box\!\!\rightarrow} C$.

Would Preservation is even more plausible than Possibility Preservation: if it's both true that *were I to drop this mug, it would hit the desk*, and that *were I to drop this mug, I would curse*, it surely follows that *were I to drop this mug and curse, it would hit the desk*. Even if Possibility Preservation is false — and so strengthening the antecedent of $A \mathrel{\Box\!\!\rightarrow} C$ with a proposition that is *possible* by A's lights may not be truth preserving — surely strengthening the antecedent with a proposition that is *necessary* by A's lights is still truth preserving. Would Preservation is thereby a natural weakening of Possibility Preservation.

The problem: Counterfactual Lockeanism invalidates Would Preservation. For example, let A be the proposition *Fair coin c is flipped three times*, B be *c lands heads at least once* and C be *c lands tails at least once*. Then, if $T = \frac{7}{8}$, since $Ch(C \mid A) = Ch(B \mid A) = \frac{7}{8}$, we have both $A \mathrel{\Box\!\!\rightarrow} B$ and $A \mathrel{\Box\!\!\rightarrow} C$. But $Ch(C \mid A\&B) = \frac{6}{7}$, which is less than T, and so we get $\neg(A\&B \mathrel{\Box\!\!\rightarrow} C)$, contra Would Preservation. So Counterfactual Lockeanism has problems even if we are happy to bite the bullet and reject Agglomeration.

2.2 Filtered Counterfactual Lockeanism

The combined failure of both Agglomeration and Would Preservation is worrying enough that I think Counterfactual Lockeanism should be rejected. Still, I deny that these reasons should force us to reject *all* optimistic views which attempt to preserve a strong connection between counterfactual truth and chance. Indeed, I'll now sketch an alternative optimistic

[16] Though the problem is not entirely new: an analogous issue is raised by (Shear and Fitelson 2018) and (Goodman and Salow fc) concerning the Lockean view of belief by which a belief in p is justified iff it is sufficiently likely.

view, inspired by recent work in formal epistemology,[17] that can preserve a connection between counterfactual truth and chance without giving up on attractive principles like Agglomeration and Would Preservation.

I'll split this subsection into three further sub-subsections. §2.2.1 will sketch the general strategy I'm pursuing. §2.2.2 will then outline a particular implantation of that strategy — taking the theory of belief in (Goodman and Salow fc) as my point of inspiration — resulting in a theory I call 'Filtered Counterfactual Lockeanism'. §2.2.3 will finish by outlining how Filtered Counterfactual Lockeanism can produce an attractive model of the **Leafy** case discussed in §1.

2.2.1 The General Strategy

The general strategy is best outlined by briefly taking a step back from counterfactuals to instead consider a general structural fact about sets of propositions. Call a set of propositions Γ 'closed under conjunction' if and only if, whenever two propositions A_1 and A_2 are both members of Γ, so is their conjunction — $A_1 \& A_2$. Next, say that a proposition in Γ, S, is the 'strongest proposition in Γ' if and only if all propositions in Γ are entailed by S. Importantly, it turns out that if Γ contains a strongest proposition S, it must be closed under conjunction. For if A_1 is in Γ and A_2 is in Γ, they will both be entailed by S, which means that S in turn entails $A_1 \& A_2$, and so $A_1 \& A_2$ must be in Γ, too.[18]

Returning to counterfactuals, we can use this observation to help devise a strategy for constructing a theory of counterfactuals that both (a) satisfies Agglomeration, yet (b) maintains a strong link between counterfactual truth and chance.

[17] Especially (Lin and Kelly 2012), (Leitgeb 2017), (Holguín 2022), (Dorst and Mandelkern 2022), (Goodman and Salow fc) and (Pearson ms).

[18] Indeed, it also turns out that any set which is closed under conjunction must also contain a strongest proposition. For suppose Γ does *not* contain a strongest proposition, then there must exist two distinct propositions in Γ, A_1 and A_2, such that no single proposition in Γ entails both of them. This would be false if Γ were closed under conjunction, for then $A_1 \& A_2$ — which entails both A_1 and A_2 — would be in Γ, too.

For (a), consider the proposition A which may be the antecedent of a counterfactual, and let Γ_A be the set of propositions C such that the counterfactual $A \boxright C$ is true. If we know that for any A, Γ_A is closed under conjunction, then we know that AGGLOMERATION is valid.[19] And as we've already observed, Γ_A is closed under conjunction if it contains a strongest proposition. Call such a proposition '$A^{\square}$' — spoken 'A squared.' $A^{\square}$ is the strongest proposition that would be true, were A true. We can guarantee AGGLOMERATION is valid so long as we can, for any antecedent A, identify $A^{\square}$. That'll be our strategy for (a).[20]

But what about (b)? That is, how do we maintain a strong link between counterfactual truth and chance? The idea now is to use facts about the chances, and in particular facts about the relevant chances conditional on A, $Ch(\cdot \mid A)$, to determine $A^{\square}$. This will, broadly, ensure that the relevant chances play a significant role in determining the truth of counterfactuals. For example, suppose we identify $A^{\square}$ in a way that guarantees that $Ch(A^{\square} \mid A) \geqslant T$. Then we'll easily be able to validate CHANCE-TRUTH LINK — for $A \boxright C$ is true only if $A^{\square}$ entails C, and so from $Ch(A^{\square} \mid A) \geqslant T$ it follows (from the axioms of probability) that $Ch(C \mid A) \geqslant T$, as per CHANCE-TRUTH LINK.

But not every way of using facts about $Ch(\cdot \mid A)$ to determine $A^{\square}$ will result in a good theory of counterfactuals. For instance, the theory which identifies $A^{\square}$ with the set of worlds to which $Ch(\cdot \mid A)$ assigns a chance of $\frac{1}{\pi}$ to is not remotely plausible.[21] So, to assess the credibility of the suggested strategy, we need to look at a concrete implementation of it. This what I turn

[19]For if $A \boxright C_1$ and $A \boxright C_2$ are both true, then both C_1 and C_2 are members of Γ_A. So if Γ_A is closed under conjunction, $C_1 \& C_2$ must also be a member of Γ_A, and so $A \boxright C_1 \& C_2$ must be true, too.

[20]What about WOULD PRESERVATION? In fact, this general strategy does not guarantee WOULD PRESERVATION, and some implementations of this strategy invalidates it (such as, for instance, an implementation that takes the theory of belief defended in (Pearson ms) as its point of inspiration). This is one reason why I take the counterfactual analogue of Goodman and Salow's theory to be particularly attractive here, since it happens to also validate WOULD PRESERVATION — see fn. 25 for details.

[21]In most cases, such a theory will say that $A^{\square}$ is empty, giving the result that almost all counterfactuals are (trivially) true!

to now.

2.2.2 Filtered Counterfactual Lockeanism

There are many ways we could implement the general strategy. One can see last 12 years of probabilistic theories of belief as offering different ways to implement this strategy in the context of trying to have a probabilistic theory of belief that allows for beliefs to be closed under conjunction — see, for instance, (Lin and Kelly 2012), (Leitgeb 2014), (Goodman and Salow fc) and (Pearson ms). Each of these views, adapted appropriately to constitute a theory of counterfactuals, will look slightly different. I won't have time to explore all of them here. Instead, I'll focus on one I take to be particularly promising: a counterfactual analogue of the approach defended by Goodman and Salow (fc).

Going forward, I'll be assuming a standard possible-worlds framework for propositions, in which a proposition A is equivalent to the set of worlds in which A is true. In turn, conjunctions, disjunctions, and negations of propositions can be respectively interpreted in terms of intersection, union and set complements, in the standard ways. Further, proposition A entails proposition B just in case A is a subset of B.

On this view, $A^{\Box}$ is identified with the smallest set of worlds S such that (i) S contains all worlds that are, according to $Ch(\cdot \mid A)$, at least as likely as some world in S, and (ii) S is itself sufficiently likely according to $Ch(\cdot \mid A)$. Intuitively, we can think of $A^{\Box}$ being constructed procedurally as follows. First, look at the set of possible worlds — call it 'W' — and identify in W the world, or worlds, that $Ch(\cdot \mid A)$ assigns the highest probability to. Put these worlds into a set S. Now ask: is S itself sufficiently likely by the lights of $Ch(\cdot \mid A)$? That is, for some specified threshold T ($\frac{1}{2} < T \leqslant 1$), is it true that $Ch(S \mid A) \geqslant T$? If so, identify S with $A^{\Box}$. If not, go back to W, identify the *second* most likely world/worlds, and add them to S, too. Now ask again: is it true that $Ch(S \mid A) \geqslant T$? If so, identify S with $A^{\Box}$. If not, repeat the process again, adding in the *third* most likely worlds to S, and so on, until

$Ch(S \mid A) \geqslant T$, in which case identify S with $A^{\Box}$.

Metaphorically, you can see condition (i) as claiming that any world which survives a "filtering" process will be members of $A^{\Box}$.[22] Imagine a filter whose mesh filters out a world iff it is sufficiently small (that is, if its probability is sufficiently low). If a world w is not filtered through, then it, along with any world of equal of greater size (probability), will also survive the filtering process and are thus members of $A^{\Box}$. Condition (ii) then tells us that we want this filtering process to leave us with a set of worlds that is itself sufficiently weighty (that is, their combined probability is sufficiently high).

We need one last condition before the view is complete. To motivate it, suppose that for some particular proposition A, the actual world, call it '@', is highly improbable by A's lights — that is, $Ch(@ \mid A)$ is low. For all I've said so far, it's possible that both (a) @ is a member of A, so A is true, yet (b) @ — due to being improbable by A's lights — is not a member of $A^{\Box}$. This is undesirable, as such a situation is one in which the counterfactual variant of modus ponens will fail: both A and $(A \mathrel{\Box\!\!\rightarrow} A^{\Box})$ are true, yet $A^{\Box}$ is false.

To amend this, we need to add a third condition (iii), saying that in cases where A is true and so contains @, for a set S to be identified with $A^{\Box}$, S must also contain @.[23] In other words, condition (iii) tells us that, when @ is in A, then if (i) and (ii) together don't by themselves admit @ into $A^{\Box}$, then @ must be added to $A^{\Box}$ (along with, thanks to the filtering process , any world at least probable as @).[24]

[22] I borrow the metaphor from (Dorst and Mandelkern 2022), who defend a similar constraint with respect to their theory of rational guessing. (Holguín 2022) also defends a similar condition in his theory of belief which he calls "cogency."

[23] Interestingly, an analogous condition is required for Goodman and Salow when they extend their probabilistic theory of *belief* into a theory of *knowledge*, in order to account for the factivity of knowledge. See (Goodman and Salow 2021).

[24] In more technical terms, condition (iii) is like the condition often called "Weak Centering" in similarity semantics for counterfactuals. Another option would be to maintain that conditionals with true antecedents aren't properly classed as "counterfactuals". (Hajek 2014) appears sympathetic to this idea, but does not commit to it. See (D. Lewis 1973) for a discussion sympathetic to the idea that a theory of counterfactuals should, in fact, account for such cases.

Note that there is another issue pertaining to counterfactuals with true antecedents I

Putting everything together, here is the theory I am proposing:

Filtered Counterfactual Lockeanism
$A \mathrel{\Box\!\!\rightarrow} C$ is true iff C is entailed by $A^{\Box}$, where $A^{\Box}$ is the smallest set S which satisfies the following conditions:

(i) For all $w \in S$, if $Ch(w' \mid A) \geqslant Ch(w \mid A)$, $w' \in S$.

(ii) $Ch(S \mid A) \geqslant T$; $\frac{1}{2} < T \leqslant 1$

(iii) If $@ \in A$, $@ \in S$.

It is hopefully clear from the discussion in §2.2.1 why Filtered Counterfactual Lockeanism will validate Agglomeration and Chance-Truth Link. That Filtered Counterfactual Lockeanism validates Would Preservation is less obvious; I've put the details in the following footnote.[25]

won't have time to discuss in detail here. In particular, a common intuition is that if a counterfactual has both a true antecedent and a true consequent, then the counterfactual itself should be counted as true. This condition will fail on my proposed theory since the consequent, even if in fact true, might not be entailed by $A^{\Box}$. I'll quickly note three possible responses. (A) contend, as mentioned just above, that conditionals with true antecedents aren't counterfactuals; (B) reject, along with a list of influential philosophers, that counterfactuals with true antecedents and consequents needn't themselves be true (e.g. (Bennett 1974), (Kratzer 1986) (Leitgeb 2013), (Hajek 2014), and (Williamson 2020) — even Lewis (1973, p. 29) expresses some sympathy with giving up such a principle), or (C) adapt the theory so that counterfactuals with true antecedents are treated as a special case, and in such special cases the truth conditions for counterfactuals are equivalent to the material conditional.

[25]To see that Would Preservation is validated, assume both $A \mathrel{\Box\!\!\rightarrow} B$ and $A \mathrel{\Box\!\!\rightarrow} C$ hold. It follows that $A^{\Box} \subseteq B$ and $A^{\Box} \subseteq C$. Would Preservation is valid if $A\&B \mathrel{\Box\!\!\rightarrow} C$, that is, if $(A\&B)^{\Box} \subseteq C$. Since we know $A^{\Box} \subseteq C$, it's enough to show that $(A\&B)^{\Box} \subseteq A^{\Box}$ as the subset relation is transitive.

$(A\&B)^{\Box}$ can only fail to be a subset of $A^{\Box}$ if there contains some world in $(A\&B)^{\Box}$ that is not in $A^{\Box}$. Given condition (i) of Filtered Counterfactual Lockeanism, this can happen only if for some $w \in A^{\Box}$ and some $w' \in W$, (1) $Ch(w \mid A) > Ch(w' \mid A)$ yet (2) $Ch(w' \mid A\&B) \geqslant Ch(w \mid A\&B)$, in which case it could be that $w' \in (A\&B)^{\Box}$ even if $\neg(w' \in A^{\Box})$.

(1) and (2) cannot both be true if $w \in A\&B$, assuming that the conditional chances are defined in the standard way using the ratio formula. For then, if two worlds are consistent with the proposition being conditioned on, facts about how their probabilities compare to one-another will be preserved. Further, our assumptions rule out that $\neg(w \in A\&B)$. For since $w \in A^{\Box}$, $w \in A$, and so it would have to follow that $\neg(w \in B)$, contradicting the above

2.2.3 A Filtered Model of Leafy

My description of Filtered Counterfactual Lockeanism has so far been extremely abstract. To help see that it delivers an attractive theory of counterfactuals, I'll now use it to outline an attractive model of **Leafy**.

First, we need a set of worlds W. Where $\mathbf{n}$ is the world in which Leafy sheds n days after today, let $W = \{\mathbf{0}, \mathbf{1}, \mathbf{2}, ...\}$. To help with readability, let $shed(n)$ denote the subset of W in which Leafy sheds at most n days after today. In other words, $shed(n)$ is the proposition that Leafy will shed in at least n days.

Next, we need to define a chance function, Ch, over W. Recall that, in **Leafy**, the chance for any leaf that hasn't yet shed that it will have shed by the next day is 0.5. Given this stipulation it's natural to define Ch such that $Ch(\{\mathbf{n}\}) = \frac{1}{2^{1+n}}$, meaning the chance Leafy sheds today (world **0**) is $\frac{1}{2}$, that it sheds tomorrow (world **1**) is $\frac{1}{4}$, and so on.

Finally, we need to set a threshold T. I'll unrealistically assume that $T = \frac{7}{8}$ — this will make more counterfactuals true than is plausible, but it will also make it easier to illustrate the relevant structural features of Filtered Counterfactual Lockeanism, such as that Possibility Preservation is invalidated (which is essential for solving the paradox considered in §1)

Now we can see Filtered Counterfactual Lockeanism in action. Consider, again:

(1) Had Leafy not shed today, it would have shed by Spring.

Supposing that k is the number of days before the first day of Spring, and letting '$\neg shed(n)$' denote all the worlds in which Leafy sheds after n days, (1) translates to:

(8) $\neg shed(0) \boxright shed(k)$

stated fact that $A^{\Box} \subseteq B$. Hence (1) and (2) can't both be true, and so $(A\&B)^{\Box} \subseteq A^{\Box}$, as required.

To assess whether (8) holds by Filtered Counterfactual Lockeanism, we need to identify $(\neg shed(0))^{\Box}$ and see whether it entails (that is, is a subset of) $shed(k)$. Indeed, according to Filtered Counterfactual Lockeanism, $(\neg shed(0))^{\Box} = \{\mathbf{1}, \mathbf{2}, \mathbf{3}\}$, since $\{\mathbf{1}, \mathbf{2}, \mathbf{3}\}$ is (i) the smallest set containing any world at least as likely, according to $Ch(\cdot \mid \neg shed(0))$, as it contains, and (ii) is such that $Ch(\{\mathbf{1}, \mathbf{2}, \mathbf{3}\} \mid \neg shed(0)) = \frac{7}{8} \geqslant T$. [26] Further, recall that in **Leafy**, it is specified that Leafy has in fact shed today — that is, **0** is the actual world — meaning condition (iii) does not require us to include the actual world into $(\neg shed(0))^{\Box}$. Since $\{\mathbf{1}, \mathbf{2}, \mathbf{3}\}$ is a subset of $shed(k)$, our theory desirably predicts (8).

So, Filtered Counterfactual Lockeanism both satisfies Optimism (at least with respect to **Leafy**) and validates Chance-Truth Link. The reason it doesn't fall prey to the paradox in §1 is that Filtered Counterfactual Lockeanism invalidates Possibility Preservation. To see this, observe that, with $T = \frac{7}{8}$, Filtered Counterfactual Lockeanism predicts the truth of:

(9) $\neg shed(0) \Box\!\!\rightarrow shed(3)$

That is, *if Leafy hadn't shed today, it still would have shed within 3 days*. This is because, as we have seen, $\{\mathbf{1}, \mathbf{2}, \mathbf{3}\} = (\neg shed(0))^{\Box}$, which is a subset of $shed(3)$. And since $(\neg\, shed(0))^{\Box}$ is *not* a subset of $shed(2)$, Filtered Counterfactual Lockeanism also predicts the truth of:

(10) $\neg[\neg shed(0) \Box\!\!\rightarrow shed(2)]$

That is, the counterfactual *If Leafy hadn't shed today, it still would have shed within two days* is false. If Possibility Preservation were valid, then it would follow from (9) and (10) that:

(11) $[\neg shed(0) \& \neg shed(2)] \Box\!\!\rightarrow shed(3)$.

[26] To see this, note that conditional on $(\neg\{\mathbf{0}\})$, the chances of $\{\mathbf{1}\}$, $\{\mathbf{2}\}$, $\{\mathbf{3}\}$ and so on, update to, respectively, $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, and so on. (Note that I am understanding $Ch(C \mid A)$ as equivalent to $\frac{Ch(A\&C)}{Ch(A)}$.) Then the combined updated probability of $\{\mathbf{1}\}$, $\{\mathbf{2}\}$, $\{\mathbf{3}\}$ is equal to $\frac{7}{8}$.

That is, *If leafy hadn't shed today, nor within 2 days, it still would have shed within 3 days*. Yet (11) is false by FILTERED COUNTERFACTUAL LOCKEANISM. For $(\neg shed(0)\&\neg shed(2))$ is equivalent to $\neg shed(2)$, and $(\neg shed(2))^{\Box} = \{\mathbf{2}, \mathbf{3}, \mathbf{4}\}$,[27] which is not a subset of $shed(3)$, meaning (11) is false, contra POSSIBILITY PRESERVATION.

I contend that these failures of POSSIBILITY PRESERVATION are, moreover, intuitively plausible (modulo our implausibly low value for T). You might, for instance, be on the one hand inclined to both assert that *If leafy hadn't shed today, it would have shed by Spring*, and accept the possibility that Leafy wouldn't have shed until mid winter had Leafy not shed today. Yet you might still, on the other hand, find the counterfactual *Had leafy not shed until mid winter, it would have shed by Spring* to be notably less assertable. FILTERED COUNTERFACTUAL LOCKEANISM explains this, since it predicts this latter counterfactual can be false, even if the former counterfactual and possibility claim are both true, contra to POSSIBILITY PRESERVATION.

Summarizing, not only does FILTERED COUNTERFACTUAL LOCKEANISM provides an optimistic view of counterfactuals without severing counterfactual truth from chance, it moreover validates AGGLOMERATION and WOULD PRESERVATION, and it provides an attractive model of **Leafy**. Mission complete?

3 Problems

Not quite. As it stands, FILTERED COUNTERFACTUAL LOCKEANISM faces various objections. In this section, I'll respond to these objections, developing the view when required.

[27] To see this, note that, similarly to the calculations in fn. 26, conditional on $(\neg\{\mathbf{0}, \mathbf{1}\})$, the chances of $\{\mathbf{2}\}$, $\{\mathbf{3}\}$, $\{\mathbf{4}\}$ and so on, update to, respectively, $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, and so on.

3.1 Problems with Possible Worlds

The model of **Leafy** in §2.2.3 raises some questions as to how I am understanding the term "world". In particular, I assumed without argument when defining W that there is just one world for every day Leafy might shed. But there are obviously far, far more possible worlds than this.

This causes a problem. For there are, presumably, and infinite number of possible worlds. And that means the chance of any particular possible world is presumably 0 — just like how the chance that a randomly selected number from the unit interval will equal $\frac{1}{\pi}$ is 0.[28] If that's right, Filtered Counterfactual Lockeanism tells us that, for any A, $A^{\Box}$ is identical to W. This would mean almost all counterfactuals — except for the tautological ones — are false. That's counterfactual skepticism.

My preferred approach to this problem is to again take a hint from recent literature on formal theories of belief. (Lin and Kelly 2012), (Leitgeb 2017), (Holguín 2022) and (Goodman and Salow fc) all hold that belief is *partition-sensitive*. Where W is the set of all possible worlds, let Q be a partition of W: a set of subsets of W that are mutually exclusive and jointly exhaustive. These theories of belief then say that whether a belief in p is rational depends on which partition is 'salient'.[29]

We can make an analogous move here. We can accept that, when *all* distinctions among worlds are salient, something like counterfactual skepticism may turn out to be true, for the reasons sketched above. But it is hardly ever true that, in fact, all distinctions *are* salient. We are often ignoring many distinctions between worlds. When considering **Leafy**, for example, it's natural to ignore a distinction between two worlds when, in both of them, Leafy

[28] Which means, in turn, that a chance of 0 does not line up neatly with *impossibility*. See (Williamson 2007) for discussion and an argument that adding infinitesimals into the codomain of the chance function won't necessarily help realign chance 0 with impossibility.

[29] This literature is, however, divided on how this notion of 'salience' is to be interpreted. (Leitgeb 2017) takes it to be salience for the relevant agent — making his view closer to the 'subject-sensitive invariantism' of (Hawthorne 2003). (Goodman and Salow 2021) take it to be salience for those ascribing to the agent rational beliefs — making their view closer to the contextualism of, for example, (D. Lewis 1996).

sheds on the same day. This suggests that when assessing counterfactuals concerning **Leafy** we do so relative to a partition Q which lumps worlds w and w' from W into the same cell iff Leafy sheds on the same day in both w and w'. If that's right, then we can preserve essentially the same model of **Leafy** as outlined in §2, substituting where appropriate 'W' for 'Q', and interpreting **n** not as the *world* in which Leafy sheds in n days from today, but rather *the set of worlds* in which Leafy sheds in n days from today.

Pulling this move means we need to modify the our precise statement of Filtered Counterfactual Lockeanism as follows.[30] Let W be the set of all possible worlds and Q be a partition of W. The value of $A^{\Box}$, and in turn the truth-value of a counterfactual $A \mathrel{\Box\!\!\rightarrow} C$, will then depend on an additional parameter, Q, as follows:

Filtered Counterfactual Lockeanism$_2$
$A \mathrel{\Box\!\!\rightarrow}_Q C$ is true iff C is entailed by $A^{\Box}_Q$, where $A^{\Box}_Q$ is the smallest union of members of Q, S, such that

(i) For all $c_i, c_j \in Q$, if $c_i \subseteq S$ and $Ch(c_i \mid A) \leqslant Ch(c_j \mid A)$, then $c_j \subseteq S$.

(ii) $Ch(S \mid A) \geqslant T$.

(iii) If $@ \in A$, then where $c_@$ is the cell of Q that $@$ is a member of, $c_@ \subseteq S$.

Where size of a union of cells of Q, S, is measured by the number of cells in that union, i.e. the number of cells that are subsets of S.

We can, again, understand this intuitively in procedural terms. Start with the cell of Q that $Ch(\cdot \mid A)$ assigns the highest probability to; say it's c_1. Add all of these worlds from c_1 into S — satisfying condition (i) — and check whether S is at least T-likely (that is, check whether (ii) is satisfied). If not, add in all the worlds from the cell that is assigned the next highest probability, say c_2, and check again. Repeat until S is at least T-likely. Finally, if A is in fact true, ensure further that all the worlds from cell of Q which

[30] Again, I'm here mimicking (Goodman and Salow fc) when they transform their theory into a partition-sensitive one.

contains the actual world, say $c_@$, are also contained in S (meaning that condition (iii) is satisfied) as well as the worlds from cells more probable than $c_@$ (meaning (i) continues to be satified).

FILTERED COUNTERFACTUAL LOCKEANISM$_2$ validates CHANCE-TRUTH LINK, and though it does not validate AGGLOMERATION nor WOULD PRESERVATION as they are written in §2.1, it will validate an appropriately modified version which clarifies that the new additional parameter on $\Box\!\!\rightarrow$, Q, remains fixed.[31] I will often drop the additional parameter Q in my discussion below; unless I specify otherwise, assume that Q is held fixed.

There is further motivation to upgrade FILTERED COUNTERFACTUAL LOCKEANISM to FILTERED COUNTERFACTUAL LOCKEANISM$_2$ — it helps us respond to the next objection.

3.2 Problems with AGGLOMERATION

Consider the following example from Hajek (2014).[32] Suppose we have a coin that is extremely biased — 99.99% — towards heads. One might have

[31] In particular, FILTERED COUNTERFACTUAL LOCKEANISM$_2$ will validate

AGGLOMERATION$_Q$
If $A \Box\!\!\rightarrow_Q C_1$ and $A \Box\!\!\rightarrow_Q C_2$ then $A \Box\!\!\rightarrow_Q C_1\&C_2$.

and

WOULD PRESERVATION$_Q$
If $A \Box\!\!\rightarrow_Q C$ and $A \Box\!\!\rightarrow B_Q$, then $A\&B \Box\!\!\rightarrow_Q C$.

Note that one interesting issue introducing question-sensitivity introduces is the matter of whether an antecedent can cross-cut the salient question. If so, it may be that more principles about counterfactauls will be invalidated — though AGGLOMERATION and FILTERED COUNTERFACTUAL LOCKEANISM$_2$ will survive. (See (Goodman and Salow fc, §8) for discussion of similar concerns regarding their theory of belief.) For my purposes here, I'll just note that it is natural to assume that when a counterfactual $A \Box\!\!\rightarrow C$ is asserted, the salient question is one for which A is a partial answer to. For, intuitively, if the antecedent of a counterfactual does not ignore certain distinctions, then those distinctions must be salient, and so a partition which compresses those distinctions cannot be the salient one.

[32] See also (Boylan 2023).

thought that optimists like me, who want to preserve a connection between counterfactual truth and chance, would be committed to the truth of:

(1H) If I were to flip the coin 100 million times, it would land heads on the first toss.

Since the conditional chance of the consequent of (1H), given the antecedent, is so high. Further, these same motivations for accepting (1H) extend to any counterfactual claiming the coin would land on heads on the *n*th flip:

(NH) If I were to flip the coin 100 million times, it would land heads on the *N*th toss.

Yet, if we also endorse Agglomeration, it follows that:

(∀H) If I were to flip the coin 100 million times, it would land heads every time.

But (∀H) has a very *im*probable consequent given the antecedent; indeed, the contrary counterfactual, (∃T), sounds impeccable:

(∃T) If I were to flip the coin 100 million times, it would land tails sooner or later.

So, it can seem like those optimists who want to preserve a strong connection between counterfactual truth and chance are in a bit of a bind. In particular, it seems as though they committed to the truth of all of the above, mutually inconsistent counterfactuals.

The partition-sensitivity of Filtered Counterfactual Lockeanism$_2$ helps escape that bind. The general strategy is to maintain that Agglomeration holds, but that the problematic consequences just outlined do not go through since the relevant counterfactuals are not true under a single partition.

First, consider (1H) above. To me, (1H) sounds *false*. But how can this be, given the high chance of the consequent conditional on the antecedent? The reason, I claim, is that (1H) is naturally read relative to a fine-grained partition, call it 'Q_{seq}', which distinguishes between all the different one-million sequences of heads/tails the coin might produce. And two of those sequences may be equally likely, yet for one of them the first flip landed on heads, and the other the first flip landed on tails. If that's right, FILTERED COUNTERFACTUAL LOCKEANISM$_2$ says either both of these cells in the partition are included in *(The coin is flipped 100 million times)*$^{\Box}_{Q_{seq}}$, or neither of them are, making it impossible for (1H) to be true relative to Q_{seq}.

However, some of my informants report being able to hear (1H) as true. What explains this? Well, often the *first* event in a sequence of events is especially salient. When this happens, I suspect (1H) is being read as relative to a different, more coarse-grained partition — call it 'Q_{first}' — that only distinguishes worlds depending on how the first flip lands. Then (1H) comes out as true as the cell corresponding to the first flip landing heads is far more likely than the competing cell in which it does not. This explanation is evidenced by the fact that it is much harder to hear as true analogous counterfactuals to (1H) about flips other than the first.[33] For example, it is very difficult to hear the following as true:

(251kH) If I were to flip the coin 100 million times, the 251,000th flip would land heads.

Note further that there are similar counterfactuals to (1H) which are easier to hear a true reading of. Consider:

(H) If I were to flip the coin, it would land heads.

The reason (H) can sound good, I contend, is that the change in the antecedent from (1H) makes it again natural to assess the counterfactual relative to Q_{first}, which only distinguishes worlds depending on how the first flip went.[34]

[33]Thanks to Kevin Dorst for discussion.

[34]Of course, a true reading of (H) is resisted if you are inclined to think "Well, it still *might* land on tails". On my view, a conversation on which such an assertion is accepted will have to be one in which the threshold T either starts out very high, or is increased to being sufficiently high on when the coin landing tails is accepted in the conversation as a counterfactual possibility.

($\forall$H), by contrast, is either naturally read relative to the very fine-grained partition I suggested for (1H), Q_{seq}, or alternatively is read relative to the partition which only distinguishes worlds based on whether every flip landed on heads or not — call it '$Q_{\exists T}$'. With respect to either partition, Filtered Counterfactual Lockeanism$_2$ correctly predicts ($\forall$H) to be false.

In contrast, whether ($\exists T$) is predicted to be true or not very much depends on whether it is read as relative to Q_{seq} or $Q_{\exists T}$. If the former, ($\exists T$) is false (for the all-heads sequence is more probable than any other); if the latter, ($\exists T$) is true (for it is far more likely to get some tails rather than all heads). I take this to be an advantage of Filtered Counterfactual Lockeanism$_2$: I can get into two minds as to whether ($\exists T$) sounds true or not, and Filtered Counterfactual Lockeanism$_2$ explains why.

All this being said, I shouldn't pretend that positing counterfactuals to be partition-sensitive raises no important questions. It certainly does. In particular, it would be nice if we had an account of how and when certain partitions become salient for the assessment of a counterfactual. Unfortunately, I suspect that no clean account is available. Recall, for instance, that I suggested it's the wording of the antecedent which makes Q_{seq} the natural relevant partition when interpreting (1H). But the antecedents cannot be the whole story. For ($\exists T$) has the same antecedent as (1H), yet, I claim, has a true reading as it can be read as relative to a different partition, $Q_{\exists T}$. So I therefore think the wording of the consequent *also* makes a difference concerning how counterfactuals are assessed.

Thankfully, it turns out that giving a theory of when certain partitions are or aren't available for assessing a counterfactual is beyond the scope of the current paper. (*Phew!*) But it is certainly an important area of research for a fully developed version of this theory.

3.3 Problems with Stronger Chance-Truth Principles

Filtered Counterfactual Lockeanism$_2$ validates Chance-Truth Link. But there are stronger principles connecting counterfactual truth and chance that those inclined to accept Chance-Truth Link might also be inclined to accept. Consider:

Chance-Truth Link$^+$
If $Ch(C_1 \mid A) = Ch(C_2 \mid A)$ then $(A \;\Box\!\!\rightarrow C_1)$ iff $(A \;\Box\!\!\rightarrow C_2)$.

Intuitively, this principle claims that if C_1 and C_2 have equal chance given A, then the counterfactuals $A \;\Box\!\!\rightarrow C_1$ and $A \;\Box\!\!\rightarrow C_2$ are either both true, or both false. You might have thought that any theory motivated to endorse Chance-Truth Link should likewise be motivated to endorse Chance-Truth Link$^+$.

The problem: Filtered Counterfactual Lockeanism$_2$ invalidates Chance-Truth Link$^+$. This can be seen by reexamining the model of **Leafy** given in §2. There, we observed that the following counterfactual is true:

(9) $\neg shed(0) \;\Box\!\!\rightarrow shed(3)$

That is, *if Leafy hadn't shed today, it would have shed within three days*. However, now consider the proposition that Leafy shed on some day *other* than the third from today. This proposition is equivalent to the union of all the cells of Q, apart from the cell containing the worlds in which leafy shed in exactly three days time: $\mathbf{1} \cup \mathbf{2} \cup \mathbf{4} \cup ....$ Call this proposition '$\neg\mathbf{3}$'. Notably, following counterfactual is *false* according to Filtered Counterfactual Lockeanism$_2$:

(12) $\neg shed(0) \;\Box\!\!\rightarrow \neg\mathbf{3}$

That is, the counterfactual *had Leafy not shed today, it wouldn't have shed in exactly three days time* is false. For $(\neg shed(0))^{\Box} = \mathbf{1} \cup \mathbf{2} \cup \mathbf{3}$, which is not a subset of $\neg\mathbf{3}$. Crucially, all of this is despite the fact that $Ch(\mathbf{1} \cup \mathbf{2} \cup \mathbf{3} \mid \neg shed(0))) = Ch(\neg\mathbf{3} \mid \neg shed(0))$. Hence Chance-Truth Link$^+$ is *false* according to Filtered Counterfactual Lockeanism$_2$.[35] Should we amend the theory in light of this?

[35]Thanks to Caspar Hare for discussion.

I don't think so. To see why, consider the following example.[36] Suppose that a horse race, between A, B, C and D was canceled yesterday. Still, we are discussing who we think would have won had the horse race taken place. It is common knowledge that, conditional on the horse race taking place, A winning has a 60% chance, B a 20% chance, and C and D both a 10% chance. Now consider the following counterfactuals:

(13) a. Had the horse race taken place, Horse A would have won.

b. Had the horse race taken place, either Horse A or Horse B would have won.

c. Had the horse race taken place, one of Horses A, C or D would have won.

If you're like me, you'll have the following reaction to these statements. (13a) can sound ok, so long as we are in a lax context where T is set as no greater than $\frac{6}{10}$. (13b) sounds better than (13a), and in particular will sound good whenever (13a) sounds good and sometimes even when (13a) sounds bad (e.g. when T is somewhere between $\frac{6}{10}$ and $\frac{8}{10}$).

There is, however, something *distinctively* infelicitous about (13c), regardless of what value T has in the relevant context. In particular, it's natural to object to (13c) as follows: *why does the consequent of (13c) not list horse B as a possible winner?* After all, B is *more likely* to have won than either C or D had the horse race taken place. So, if one is committed to either C or D potentially winning, it seems one should too be committed to B potentially winning. And that means (13c) is false: it's false to say that A, C or D *would* have won, since if C or D *might have* won, B too surely *might* have won.

Taking this data at face value, then, there is at least some context in which (13b) is true, yet (13c) *false*. And that means we have a counterexample to Chance-Truth Link$^+$: for it is just as likely, conditional on the horse race

[36] Here, again, analogies to epistemology appear. This is just the counterfactual version of a famous case from Jeremy Goodman, discussed in the seminal article by Hawthorne, Rothschild and Spectre (2016) on weak belief.

taking place, that A or B would win, as it is that A or C or D would win.[37]

Further, going back to **Leafy**, I think Filtered Counterfactual Lockeanism$_2$ makes the right predictions even there. For consider (9) again, written in English below, which Filtered Counterfactual Lockeanism$_2$ predicted to be false:

(14) Had Leafy not shed today, it wouldn't have shed in exactly three days time.

I contend that (14) would be a remarkably odd thing to assert. As far as I can tell, there are only two ways two charitably interpret it as indicating something true. On the first, (14) is interpreted as indicating that the speaker has some kind special knowledge of what could possibly have happened in three days from now that rules out Leafy shedding on that day. But such knowledge would be impossible given our stipulations about the relevant chances in **Leafy**, and the fact that the antecedent on (14) is in fact false. Not even a time traveler — who could know, despite the objective chances, when a leaf *will* in fact shed — could obtain such special knowledge. Rather, for this kind of special knowledge we'd need something like a *possible worlds* traveler.[38]

The second way to charitably interpret the assertion would be as the speaker indicating they believe that Leafy wouldn't have shed in exactly three days, since they believe Leafy would have shed within two days. But if the asserter specifies that they do not believe this, we are left with a remarkably odd asertion:

(15) Had Leafy not shed today, it wouldn't have shed in exactly three days time. But I don't say that because I think it would have shed within the first two days — it might have taken longer than three days to shed!

[37]Note, further, that by setting $T = 80\%$, and letting our relevant partition Q be $\{$ A wins, B wins, C wins, D wins$\}$, Filtered Counterfactual Lockeanism$_2$ naturally allows for a context in which (13b) is true but (13c) false.

[38]An idea which is clearly incoherent, though playfully considered in the movie *Everything Everywhere All At Once*.

(15) sounds strange for the same reason (13c) did: if we accept that Leafy might have taken *longer* than 3 days to shed, it must surely also be accepted that Leafy might have taken *exactly* three days to shed.[39]

So, on reflection, I think FILTERED COUNTERFACTUAL LOCKEANISM$_2$ gives exactly the right result regarding CHANCE-TRUTH LINK$^+$. CHANCE-TRUTH LINK$^+$ is just false.[40]

3.4 Problems with Inference

I've saved the hardest problem for last. A good test of a theory of counterfactuals concerns the patterns of inference it licenses. Indeed, I have used this test to in-part motivate FILTERED COUNTERFACTUAL LOCKEANISM$_2$, as it validates AGGLOMERATION and WOULD PRESERVATION.

The problem is that other plausible principles do not play out so nicely. Consider:

WOULD CONSERVATISM
If $A \mathrel{\Box\!\!\rightarrow} B$ and $(A\&B) \mathrel{\Box\!\!\rightarrow} C$, then $A \mathrel{\Box\!\!\rightarrow} C$

WOULD CONSERVATISM can be seen as a parallel to WOULD PRESERVATION. WOULD PRESERVATION tells us that, upon strengthening the antecedent of a true counterfactual with a proposition that already would be true by that antecedent's lights, we do not thereby *invalidate* any previously valid counterfactual inferences. In contrast, WOULD CONSERVATISM tells us that, upon

[39]It might be possible to hear (14) as sounding ok if there is something especially salient about 3 days time. For instance, something like "Had Leafy not shed today, it wouldn't have shed on your birthday" can sound better, even with the added qualifications from (15). But this is again, I think, explained by a shift in the partition: the counterfactual is now interpreted with respect to a partition which only distinguishes worlds in which Leafy sheds on your birthday from those in which it doesn't.

[40]Note that it is no accident FILTERED COUNTERFACTUAL LOCKEANISM$_2$ makes the predictions outlined in this subsection. Condition (i) of FILTERED COUNTERFACTUAL LOCKEANISM$_2$ — the filtering condition — mimics analogous conditions concerning weak belief/guessing defended by (Holguín 2022) and (Dorst and Mandelkern 2022). And these conditions in that context are in turn motivated by considering the felicity of ascriptions of belief analogous to the the counterfactuals stated in (13).

strengthening the antecedent of a true counterfactual with a with proposition that already would be true by that antecedent's lights, we do not thereby *validate* any previously invalid counterfactual inferences. Whereas Would Preservation preserves previously valid inferences, Would Conservatism is conservative in that it prevents new inferences from being licensed.

For example, suppose it's both truth that, *were I to throw a party, John would come*, and that *were I to throw a party and were John to come, I'd have fun.* Then it must *already* be true that, *were I to throw a party, I'd have fun.* It would be weird if it's only true that I'd have fun were I to throw a party once we suppose *further* that John would come, since it's already true that John would come were I to throw a party. Would Conservatism underlies these intuitive judgments.

The problem is that Filtered Counterfactual Lockeanism invalidates Would Conservatism.[41] To see this, consider **Leafy** one last time, but suppose that $T = \frac{14}{15}$. Notice first that, by this threshold, $\neg shed(0)^{\Box} = \mathbf{1} \cup \mathbf{2} \cup \mathbf{3} \cup \mathbf{4}$.[42] Hence the following counterfactual is true:

(16) $\neg shed(0) \mathrel{\Box\!\!\rightarrow} shed(4)$

In words, *had Leafy not shed today, it would have shed within 4 days.*

Notice second that, upon strengthening the antecedent of (16) with $shed(4)$, it follows that we are not in **4**, since $(\neg shed(0) \& shed(4))^{\Box} = \mathbf{1} \cup \mathbf{2} \cup \mathbf{3}$. Hence the following counterfactual is also true:

(17) $\neg shed(0) \cap shed(4) \mathrel{\Box\!\!\rightarrow} \neg \mathbf{4}$

In words: *Had Leafy not shed today but within four days, it wouldn't have taken exactly four days to shed.*

Notice finally that, by Would Conservatism, it should follow that *Had Leafy not shed today, it wouldn't have taken exactly four days to shed*. But in fact

[41] As Goodman and Salow note with respect to the belief-revision analogue to Would Conservatism (Goodman and Salow fc, §7).

[42] For $\mathbf{1} \cup \mathbf{2} \cup \mathbf{3} \cup \mathbf{4}$ is the smallest union, containing any cell at least as likely as it contains, such that it's chance conditional on $\neg shed(0)$ at least $\frac{14}{15}$. Recall that since the actual world is in **0**, condition (iii) of Filtered Counterfactual Lockeanism is not activated.

this claim false on the current model, since as stated above **4** is a member of $\neg shed(0)^{\Box}$. That is, the following counterfactual is false:

(18) $\neg shed(0) \mathrel{\Box\!\!\rightarrow} \neg\mathbf{4}$.

Hence Filtered Counterfactual Lockeanism$_2$ invalidates Would Conservatism: it's both true that *had Leafy not shed today, it would have shed within four days* and that *had Leafy not shed today but still within four days, it wouldn't have taken exactly four days to shed*, but false that *had Leafy not shed today, it wouldn't have taken exactly four days to shed*.

I think this is a bad result in itself, but it moreover generates further problems. In particular, it means that we also also get counterexamples of the extremely appealing:[43]

Counterfactual Proof-By-Cases (CPBC)
If $(A\&B) \mathrel{\Box\!\!\rightarrow} C$ and $(A\&\neg B) \mathrel{\Box\!\!\rightarrow} C$ are both true, then $A \mathrel{\Box\!\!\rightarrow} C$ is true.

CPBC appears to underlie an impeccable piece of reasoning. Suppose I didn't throw a birthday party, but want to know what would have happened had I. I realize that, had I thrown a party and Felix came, I wouldn't have had fun. After all, Felix is always the center of attention at parties, and so, if he had come, he would take away attention from me on my special day. But I then realize further that, had I thrown a party and had Felix not come, I *also* wouldn't have had fun. After all, Felix is a good friend of mine, so him not turning up would have been disappointing. It seems I can conclude that I wouldn't have had fun had I thrown a birthday party. CPBC encodes this kind of reasoning.

But the above results give us the resources to offer a counterexample to CPBC, too. We just need to observe further that Filtered Counterfactual Lockeanism$_2$, along with any sensible theory of counterfactuals, will predict the truth of:

[43] Again, as anaologusly noted by Goodman and Salow for their theory of belief (Goodman and Salow fc, §7).

(19) $(\neg shed(0) \cap \neg shed(4)) \mathrel{\Box\!\!\rightarrow} \neg \mathbf{4}$

In words: *Had leafy not shed today nor within four days, it wouldn't have taken exactly four days to shed*. Given (19) and (17) — *Had Leafy not shed today but within four days, it wouldn't have taken exactly four days to shed* — Counterfactual Proof By Cases would allow one to infer *Had Leafy not shed today, it wouldn't have shed in exactly four days*. But this counterfactual — (18) — we've already seen is false by Filtered Counterfactual Lockeanism$_2$. So Filtered Counterfactual Lockeanism$_2$ also invalidates Counterfactual Proof By Cases.

I think these problems are bad, and I am unsure how to fix them. One option is to bite the bullet, and accept that these principles do indeed fail. We've seen that Filtered Counterfactual Lockeanism$_2$ provides us with a variety of other theoretical advantages, so giving up on these principles may be worth the price.

Further, it might be possible to soften the blow of biting the bullet by demonstrating that counterexamples to Would Conservatism are somewhat limited. Notice, for instance, that in the above model we only get failures of Would Conservatism by strengthening the antecedent with a proposition that is particularly strong. Indeed, the antecedent of (17), $\neg shed(0) \cap shed(4)$ is equivalent to $(\neg shed(0))^{\Box}$, which is the strongest thing that would be true had Leafy not shed today. So the counterexample to Would Conservaitsm is being generated by strengthening the antecedent (17) to the *strongest possible fact* about what would be true. And if we weaken this strengthening even just a little bit — say we strengthen it with $shed(5)$ rather than $shed(4)$ — we no longer get a counterexample to Would Conservatism.[44]

So there is hope that, perhaps, Would Conservatism, even though strictly false, nevertheless encodes an inference that preserves truth *most of the time*. Moreover, those times when the inference is not truth preserving will be hard to identify. For it is presumably hard to distinguish the strongest proposition that would be true were A true from ever so slightly stronger

[44] For, by $T = \frac{14}{15}$ $(\neg shed(0) \cap shed(5))^{\Box} = \mathbf{1} \cup \mathbf{2} \cup \mathbf{3} \cup \mathbf{4}$ which is just equal to $\neg shed(0)^{\Box}$.

propositions that wouldn't be true were A true. So, by a Williamson-style (2000) margin-for-error principle, it's plausible that one cannot know of the strongest proposition that would be true were A true that it is in fact this strongest proposition. Indeed, it's quite intuitive that these propositions are hard to identify. Suppose that matter of what would have happened had Leafy not shed today arises in conversation. If you were asked "what, precisely, is the strongest proposition that would be true had leafy not shed today?" you'd struggle to come up with an answer, and you'd suspect that anyone who does offer an answer is merely guessing.[45]

In sum: there is hope that these failures of WOULD CONSERVATISM are both limited *and* hard to identify, giving us an explanation of this principle's appeal. Unfortunately, I don't have anything like a formal result demonstrating that all failures of WOULD CONSERVATISM will be constrained in the ways I've suggested. So this avenue of response is, for now, merely promissory.[46]

Another option is to abandon FILTERED COUNTERFACTUAL LOCKEANISM$_2$, and look for inspiration from alternative recent theories of belief to (Goodman and Salow fc) — like (Lin and Kelly 2012) and (Pearson ms) — the counterfactual analogues of which would use facts about $Ch(\cdot \mid A)$ to determine $A^{\Box}$ in a different way to FILTERED COUNTERFACTUAL LOCKEANISM$_2$. These approaches arguably fare better with respect to WOULD CONSERVATISM and CPBC, but each comes with a distinctive set of problems. Again, I'll have to leave a full exploration of these alternative options for further research.

To summarize, while I recognize that invalidating WOULD CONSERVATISM and CPBC is a serious drawback, I am hopeful that either (a) it is not so bad that it makes FILTERED COUNTERFACTUAL LOCKEANISM untenable, or (b) that there is at least some available theory, following the same general strategy

[45]Note that this observation doesn't entail any kind of skepticism about counterfactuals. Though it may be hard to know the *strongest* thing that would be true had Leafy not shed today, it is presumably quite easy to know other propositions that would be true, such that Leafy would have shed eventually.

[46]It's worth noting that FILTERED COUNTERFACTUAL LOCKEANISM$_2$—thanks to condition (iii)—does at least validate the weaker version of CPBC that, if $A \Box\!\!\rightarrow C$ and $\neg A \Box\!\!\rightarrow C$ are both true, then C is true.

from §2.1, that does better on this front.

4 Conclusion

As I said, I'm an optimist. I think that many of the counterfactuals asserted in ordinary discourse are true. But I also want to preserve a strong link between counterfactual truth and chance. While the obvious way to have both optimism and a chance-truth connection, Counterfactual Lockeanism, faces serious problems, I have shown that attractive alternatives are available, like Filtered Counterfactual Lockeanism, that avoid those problems. I therefore take Filtered Counterfactual Lockeanism to provide a highly attractive theory of counterfactual truth.

Less ambitiously, I hope to have at least illustrated the fruitfulness of a particular methodological approach: that of drawing analogies between epistemology and the literature on counterfactuals in order to make progress on the latter. Drawing these analogies has been done before — see especially (Moss 2012) and (K. Lewis 2016) — but I think these analogies are still underexplored, and I take the above as a demonstration as to how drawing upon them can be fruitful. I invite the reader to investigate these analogies even further. Moreover, as I mentioned above, there is more work to be done with respect to the analogy I am exploring. The counterfactual analogue of Goodman and Salow's (fc) theory of the belief is not the only way to implement the general strategy outlined in §2. It may be that by taking as our point of inspiration the similar, but importantly different, theories of belief like those in (Lin and Kelly 2012), (Leitgeb 2014) or (Pearson ms) will deliver an overall more attractive view. I plan to explore these alternatives in further research.

In closing, I want to consider an alternative reading of this paper. In particular, the paradox I outlined in §1 was presented as a puzzle about counterfactual *truth*. But you can also present it as a puzzle about counterfactual *knowledge*. For notice that the following epistemic analogues of

Chance-Truth Link and Possibility Preservation are also extremely plausible:

Chance-Knowledge Link
If $Ch(C \mid A) \geqslant T$ then you cannot know $A \boxright C$.

Epistemic Possibility Preservation
If you know $A \boxright C$, but you don't know $A \boxright \neg B$, then you know $A\&B \boxright C$.

And if these principles are both correct, then in a similar style argument to the one in §1, from the claim that you can know *had Leafy not shed today, Leafy would have shed by spring*, and that for any day n, you don't know whether *had Leafy not shed after n days, Leafy would have shed in* $n + 1$ *days*, we can eventually derive the absurd conclusion that you can know *Had Leafy not shed by Spring, Leafy would have shed by Spring*.

So, even if you're already committed to some alternative optimistic theory to Filtered Counterfactual Lockeanism — say some theory that denies Chance-Truth Link — you'll still need to say something about about this *epistemic* version of the paradox from §1. For if you're an optimist about counterfactual truth, I imagine you're an optimist about counterfactual knowledge, too. (That is, you'll think that most ordinary counterfactuals can be known.)

At this point, you may wish to — and a few people have suggested this to me — follow something like the strategy suggested in §2 to construct, instead of a theory about counterfactual truth, a theory about counterfactual knowledge. If $A^{\Box}$ is interpreted, not as the strongest proposition that would be true were A true, but rather the strongest proposition one can *know* would be true were A true, we should be able to construct a plausible theory of counterfactual knowledge which validates Chance-Knowledge Link, yet invalidates Epistemic Possibility Preservation, thus solving the epistemic version of the puzzle.

Of course, the details of this option will need to be worked out in more detail; I won't fully spell out those details here. I do want to note, however, that taking this option in response to this epistemic version of the paradox in fact provides another, indirect argument for Filtered Counterfactual Lockeanism. For notice that in endorsing this strategy in response to the epistemic version of the paradox we now have two components: on the one hand, a theory of counterfactual truth, and on the other hand, a theory of counterfactual knowledge. Presumably, these two theories are not unrelated. And in particular, ideally the two will be related as follows: given the accepted theory of counterfactual truth, and given an independently plausible theory of knowledge, we ought to be able to derive the proposed theory of counterfactual knowledge.

I take it that achieving this ideal is a highly non-trivial task. Observe, however, that this task is far more tractable if we accept Filtered Counterfactual Lockeanism as our theory of counterfactual truth. For example, combining Filtered Counterfactual Lockeanism with any account of knowledge on which knowledge is factive will immediately deliver Chance-Knowledge Link. For, by this combination, you can then only know true propositions, and the only true counterfactuals $A \mathrel{\Box\!\!\rightarrow} C$ are those where C has a high enough chance conditional on A. This is the indirect argument for Filtered Counterfactual Lockenaism.[47]

In a similar vein, notice that the paradox in §1 can also be reworked as a paradox concerning when counterfactuals can be *asserted*. Even counterfactual skeptics shouldn't (and don't — see (Hajek 2014)) deny that most counterfactuals are assertable. Yet, there are again assertion-based variants of Chance-Truth Link and Possibility Preservation which, when applied to **Leafy**, will lead to paradox. Just take Chance-Knowledge Link and Epis-

[47]I hope to explore alternative approaches in further research. From a preliminarily investigation, the most attractive alternative to just endorsing Filtered Counterfactual Lockenaism is to endorse the "randomization" variant of Stalnker-style selection function semantics (Stalnaker 1968), as discussed by (Boylan 2023), (Bacon 2015) and (Schulz 2017). Such an approach can, when combined with a bare-bones Hintikka (1962) semantics for knowledge, validate Possibility Preservation while invalidating Epistemic Possibility Preservation, thus providing a response to the epistemic version of the paradox. Despite this, I am not sure how to have such a view combine with a theory of knowledge to predict Chance-Knowledge Link — more work needs to be done.

TEMIC POSSIBILITY PRESERVATION, subbing in ”assert” for ”know”, and then interpret the principles as stating norms of assertion.

In response to *this* version of the paradox, one may desire to rework the theory in §2 as a theory of when counterfactuals are assertable, rather than when they are true or when they are known. But then a similar question arises. Our theory of counterfactual assertability ought not be unrelated from our theory of counterfactual truth. Indeed, our theory of counterfactual assertability will ideally follow follow from our theory of counterfactual truth plus our theory of the norms of assertion. And again, I take it this task is much easier if we accept FILTERED COUNTERFACTUAL LOCKEANISM as our theory of counterfactual truth: for combined with a theory of assertability in which (norm-abiding) assertions must be true, we should naturally predict the suggested theory of counterfactual assertability.

So, if you'd rather understand the theory in §2 as a theory of when counterfactuals are knowable or can be asserted, be my guest. But I contend that such a view makes most sense when one also, in addition, endorses FILTERED COUNTERFACTUAL LOCKEANISM.[48]

References

Bacon, Andrew (2015). "Stalnaker?s Thesis in Context". In: *Review of Symbolic Logic* 8.1, pp. 131–163. DOI: 10.1017/s1755020314000318.

Bennett, Jonathan (1974). "Counterfactuals and Possible Worlds". In: *Canadian Journal of Philosophy* 4.December, pp. 381–402. DOI: 10.1080/00455091.1974.10716947.

Boylan, David (2023). "Counterfactual Skepticism is (Just) Skepticism". In: *Philosophy and Phenomenological Research* 108.1, pp. 259–286. DOI: 10.1111/phpr.12957.

[48]This is still a work in progress — please do not hesitate to email me if you have comments. I would like to thank Kenneth Black, Kevin Dorst, Alan Hájek, Capsar Hare, Justin Khoo, and the attendees at MIT's work-in-progress seminar for their useful feedback so far.

Boylan, David and Ginger Schultheis (2021). "How Strong Is a Counterfactual?" In: *Journal of Philosophy* 118.7, pp. 373–404. DOI: 10.5840/jphil2021118728.

Dorr, Cian, Jeremy Goodman, and John Hawthorne (2014). "Knowing Against the Odds". In: *Philosophical Studies* 170.2, pp. 277–287. DOI: 10.1007/s11098-013-0212-9.

Dorst, Kevin and Matthew Mandelkern (2022). "Good Guesses". In: *Philosophy and Phenomenological Research*. DOI: 10.1111/phpr.12831.

Fintel, Kai von (Apr. 2001). "Counterfactuals in a Dynamic Context". In: *Ken Hale: A Life in Language*. The MIT Press. ISBN: 9780262316125. DOI: 10.7551/mitpress/4056.003.0006. eprint: https://direct.mit.edu/book/chapter-pdf/2308196/9780262316125_cac.pdf. URL: https://doi.org/10.7551/mitpress/4056.003.0006.

Goodman, Jeremy and Bernhard Salow (June 2021). "Knowledge from Probability". In: *Electronic Proceedings in Theoretical Computer Science* 335, pp. 171–186. ISSN: 2075-2180. DOI: 10.4204/eptcs.335.15. URL: http://dx.doi.org/10.4204/EPTCS.335.15.

— (fc). "Belief Revision Normalized". In: *Journal of Philosophical Logic*.

Hajek, Alan (2014). *Most Counterfactuals Are False*.

— (ms). "A Chancy Account of Counterfactuals". Manuscript.

Hawthorne, John (2003). *Knowledge and Lotteries*. Oxford University Press.

— (2005). "Chance and Counterfactuals". In: *Philosophy and Phenomenological Research* 70.2, pp. 396–405. DOI: 10.1111/j.1933-1592.2005.tb00534.x.

Hawthorne, John, Daniel Rothschild, and Levi Spectre (2016). "Belief is Weak". In: *Philosophical Studies* 173.5, pp. 1393–1404. DOI: 10.1007/s11098-015-0553-7.

Hintikka, Jaakko (1962). *Knowledge and Belief*. Ithaca, N.Y., Cornell University Press.

Holguín, Ben (2022). "Thinking, Guessing, and Believing". In: *Philosophers' Imprint*, pp. 1–34.

Kratzer, Angelika (1986). "Conditionals". In: *Chicago Linguistics Society* 22.2, pp. 1–15.

Leitgeb, Hannes (2012). "A Probabilistic Semantics for Counterfactuals. Part A". In: *Review of Symbolic Logic* 5.1, pp. 26–84. DOI: 10.1017/s1755020311000153.

— (2013). "A Lottery Paradox for Counterfactuals Without Agglomeration". In: *Philosophy and Phenomenological Research* 89.3, pp. 605–636. DOI: 10.1111/phpr.12035.

— (2014). "The Stability Theory of Belief". In: *Philosophical Review* 123.2, pp. 131–171. DOI: 10.1215/00318108-2400575.

— (2017). *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford University Press.

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

— (1979). "Counterfactual Dependence and Time's Arrow". In: *Noûs* 13.4, pp. 455–476. DOI: 10.2307/2215339.

— (1996). "Elusive Knowledge". In: *Australasian Journal of Philosophy* 74.4, pp. 549–567. DOI: 10.1080/00048409612347521.

Lewis, Karen (2016). "Elusive Counterfactuals". In: *Noûs* 50.2, pp. 286–313. DOI: 10.1111/nous.12085.

Lin, Hanti and Kevin Kelly (2012). "Propositional Reasoning That Tracks Probabilistic Reasoning". In: *Journal of Philosophical Logic* 41.6, pp. 957–981. DOI: 10.1007/s10992-012-9237-3.

Makinson, David C. (1965). "The Paradox of the Preface". In: *Analysis* 25.6, pp. 205–207. DOI: 10.1093/analys/25.6.205.

Moss, Sarah (2012). "On the Pragmatics of Counterfactuals". In: *Noûs* 46.3, pp. 561–586. DOI: 10.1111/j.1468-0068.2010.00798.x.

Pearson, Joshua Edward (ms). "Belief Revision Revised". Manuscript.

Schulz, Moritz (2017). *Counterfactuals and Probability*. Oxford, United Kingdom: Oxford University Press.

Shear, Ted and Branden Fitelson (2018). "Two Approaches to Belief Revision". In: *Erkenntnis* 84.3, pp. 487–518. DOI: 10.1007/s10670-017-9968-1.

Stalnaker, Robert (1968). "A Theory of Conditionals". In: *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*. Ed. by Nicholas Rescher. Oxford: Blackwell, pp. 98–112.

Williams, Robert (2008). "Chances, Counterfactuals, and Similarity". In: *Philosophy and Phenomenological Research* 77.2, pp. 385–420. DOI: 10.1111/j.1933-1592.2008.00196.x.

Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press.

— (2007). "How Probable is an Infinite Sequence of Heads?" In: *Analysis* 67.3, pp. 173–180. DOI: 10.1093/analys/67.3.173.

— (2020). *Suppose and Tell: The Semantics and Heuristics of Conditionals*. Oxford, England: Oxford University Press.