

CAN IT BE THAT IT WOULD HAVE BEEN EVEN THOUGH IT MIGHT NOT HAVE BEEN?

Keith DeRose
Yale University

1. A Skeptical Puzzle Involving “Might” and “Would” Counterfactuals

The score was tied in the bottom of the ninth, I was on third base, and there was only one out when Bubba hit a towering fly ball to deep left-center. Although I’m no speed-demon, the ball was hammered so far that I easily could have scored the winning run if I had tagged up. But I didn’t. I got caught up in the excitement and stupidly played it half way, standing between third and home until I saw the center fielder make his spectacular catch, after which I had to return sheepishly to third. The next batter grounded out, and we lost the game in extra innings.

This thought haunts me:

(A) If I had tagged up, I would have scored the winning run.

Given the circumstances as I described them above, this is close to being as clear a case of a contingently true counterfactual conditional as one might hope to find.

But a skeptic might suggest a little caution here. She puts forward the following conditional, inviting me to agree:

(B) If I had tagged up, I *might* have tripped, fallen, and been thrown out.

Certainly, I agree, *if* I had tripped and fallen, I might—in fact, probably would—have been thrown out. But, I protest, I *wouldn’t* have tripped and fallen. (While not overly fast, neither am I very clumsy. I’m certainly not prone to tripping and falling—at least no more so than the next person.) The skeptic answers that she is not asking me to agree that I *would* have tripped and fallen had I tagged up, only that I *might* have. Am I prepared to deny even this? Will I insist that it’s *impossible* that I would have tripped if I had tagged up? This seems a bit much. Certainly (B) has its appeal.

But it also has its danger. For, the skeptic continues, given the truth of (B), shouldn't we conclude that

(C) If I had tagged up, I might not have scored the winning run?

It seems hard to deny (C) while agreeing to (B). Am I to insist that I *couldn't* have failed to score the winning run had I tagged up, even while admitting that I *might* have been thrown out in those circumstances?

But if I agree to (C), how can I continue to accept (A)? These two seem mutually inconsistent. It certainly is uncomfortable to conjoin them:

(C + A) If I had tagged up, I might not have scored the winning run; nevertheless, I would have scored the winning run if I had tagged up

is not a happy utterance, to say the least. It seems inconsistent. And it *is* inconsistent according to the standard account of the relation between "might" and "would" counterfactuals that we'll encounter below in section 2. Since (C) is true and is incompatible with (A), I should give up (A).

Or so the skeptic claims. How shall we resist her argument? As I've noted, none of the three possible ways out—denying (B), accepting (B) while denying (C), or accepting (C) while continuing to hold to (A)—is altogether comfortable. Still, any of these options seems preferable to giving up (A): If (A) isn't true, what counterfactual conditional *is* (contingently) true?¹

At this point, our skeptic offers to soften the blow. She proposes the "skeptical solution" of denying (A), but, in its place, accepting

(D) If I had tagged up, I *probably* would have scored the winning run.²

Simply denying (A) is hard to do, but perhaps some of the sting of this denial is removed if (D) can still be affirmed. And some will be able to work up some enthusiasm for the thought that there are no (contingent) facts about what would have been the case in various counterfactual situations, but only about what *probably* would have been the case.

Some, but not me. I cannot live with this skeptical solution to our problem. For, as I hope is clear, the skeptical argument just presented seems fairly easily adaptable to target just about any "would" counterfactual conditional—at least any such conditional which does not have a "probably" or other such hedge built into it—that seems contingently true. Are we to suppose that *no* such conditionals *are* true? Must we, if we are to avoid speaking falsely, *always* hedge our conditionals? I can't believe it's so. Still, once (D) is allowed, this skeptical solution doesn't initially seem that much worse than the three anti-skeptical solutions mentioned above.

What we have on our hands, then, is a puzzle, and it's clear that a big part of what's needed to solve the puzzle is an account of the relation between "would"

and “might” counterfactuals. In what follows, I present and argue for the “Epistemic Thesis” (ET) of this relation, largely on the grounds of its ability to solve our puzzle. But let’s first look at ET’s main rival.

2. Lewis’s Duality Thesis (DT) and Our Puzzle

Perhaps the most popular account of the relation between “might” and “would” counterfactuals is what we may call the “Duality Thesis” (DT) that David Lewis advanced in [8]. Letting “ $P \Box \rightarrow Q$ ” denote “would” counterfactual conditionals like (A), and letting “ $P \Diamond \rightarrow Q$ ” denote “might” counterfactuals like (B), the alleged relation is as follows:

$$(DT) (P \Diamond \rightarrow Q) \equiv \sim(P \Box \rightarrow \sim Q)^3$$

This account blocks one of the three possible escapes from the skeptic’s argument. On DT, $(C + A)$ really is inconsistent, and so, if the skeptic is to be stopped at all, she must be stopped before she reaches (C).

No wonder, then, that DT is a favorite weapon of counterfactual skeptics. In philosophy of religion, where many such skeptics lurk, a debate has been raging in recent years over whether God has “middle knowledge” of various counterfactuals involving what creatures would have done if they had been put in various situations in which they’d have had incompatibilist freedom. Since the parties involved agree that God is omniscient, and so would know such counterfactuals if they were true, the debate has turned on whether those counterfactuals *are* true—whether there are facts about what these creatures would have done in the relevant situations. The skeptics in the debate have marshaled DT, along with claims to the effect that the creatures, since they’d be free in the relevant situations, *might* have performed the acts in question and *might not* have, to reach conclusions to the effect that there’s no fact of the matter as to what they *would* have done. (Though it’s allowed that there are facts about what they *probably* would have done.)⁴

But accepting DT doesn’t necessarily make a counterfactual skeptic out of you. What it does do is face you with a lot of hard choices. When considered *individually*, (A) and (C) both seem true. (Or at least (C) seems to follow from something which does seem true—(B).) DT devotees will face similar dilemmas with respect to other counterfactuals. In each such case, adherents to DT must reject something with a good deal of initial plausibility. But, of course, as they’ll be quick to point out, *denying* DT is also counter-intuitive. When (A) and (C) are considered *together*, it seems they can’t both be true.

Enter ET, the hero of this story, according to which “would” counterfactuals are compatible with the relevant “might not” counterfactuals, but which *solves* our puzzle by *explaining* why conjunctions like $(C + A)$ can *seem* inconsistent. On ET, the metaphysical view that there are no facts about what would have been the case in relevant counterfactual situations, insofar as it’s based on arguments like the one in section 1, above, is generated by a misleading trick of language.

But before unveiling ET, we'll consider, in section 3 below, an analogue of our present puzzle which involves non-conditional statements regarding the past. An important advantage of ET is that it assimilates the puzzle of the section 1 to the analogue you're about to encounter.

3. A "Puzzle" Not Involving Conditionals

You're watching a baseball game on TV. The score is tied in the bottom of the ninth, and there's a speedy runner on third with only one out. We pick up the announcer's call as the batter makes solid contact: "It's a deep fly to left-center! Way back! It's going all the way to the warning track! Wimpy'll take it. He doesn't have much of an arm. Speedy's tagging up. There's the catch. Here comes Speedy. It's a weak throw. He'll score easily. . ." At that point, an electrical outage occurs. Far away, Speedy scores the winning run, but you don't get to see it, for your TV has gone dead.

You're satisfied that

(A') Speedy scored the winning run.

And the following argument wouldn't tempt you to conclude that Speedy didn't score the winning run:

(B') It's possible that Speedy tripped, fell, and was thrown out
 So, (C') It's possible that Speedy didn't score the winning run
 So, (not-A') It's not the case that Speedy scored the winning run.

This skeptical argument has a good deal of power through its second step, where the sub-conclusion (C') is reached. At that point, a skeptic may be ready to rationally motivate a disturbing conclusion regarding whether you *know* that (A') is true, or whether you're *in a position to flat-out assert* that it's true, rather than settling for the more modest

(D') Speedy *probably* scored the winning run.

I favor the following contextualist assessment of such assaults on knowability and assertability: Introducing the (mild) skeptical hypothesis that Speedy tripped, fell, and was thrown out,⁵ has the effect of raising the standards for knowledge to a level at which one doesn't count as knowing that Speedy scored the winning run.⁶ And since you represent yourself as knowing a fact if you flat-out assert it, you're in no position to flat-out assert that (A') is true in the skeptical context that's been created by the introduction of that skeptical hypothesis. This, I think, best explains the persuasiveness of such skeptical attacks on the knowability and assertability of (A'). But since, on this explanation, the skeptic only succeeds by raising the standards for knowledge (and there-

fore for proper assertability), the success of her attack does not jeopardize the propriety of our assertions of the likes of (A'), nor the truth of our claims to know the likes of (A'), made in other contexts, where the standards for knowledge are not so unusually inflated.

But whatever we say about such an attack on the knowability or assertability of (A'), the skeptic clearly oversteps when she attempts to conclude that (A') is *false*, for (not-A') clearly doesn't follow from (C'). Yet,

(C' + A') It's possible that Speedy didn't score the winning run, but he did,

does sound awful—about as bad as (C + A).

Why? And, more generally, why do conjunctions of the form “P, but it's possible that not-P_{ind}” (the subscript “ind” indicates that the embedded P is in the indicative mood), like, to take another example, “It's raining outside, but it's possible that it isn't,” produce the feeling of inconsistency, though they are in fact perfectly consistent? Briefly, as I've argued elsewhere, it's because (a) in flat-out asserting that P, while one doesn't assert that one knows that P, one does represent it as being the case that one *knows* that P, and (b) the content of “It's possible that not-P_{ind}” is such that this second conjunct entails that the speaker *doesn't know* that P. Thus, what one says in asserting the second conjunct of “P, but it's possible that not-P_{ind},” while it's perfectly consistent with what one says in asserting the first conjunct, is inconsistent with something one represents as being the case in asserting the first conjunct. This supports our sense that *some* inconsistency is responsible for the clash involved in asserting the conjunction, while, at the same time, happily removing that inconsistency from the realm of what's asserted: The conjunction asserted is itself perfectly consistent, but in trying to assert it, one gets involved in a contradiction between one thing that one asserts, and another thing that one represents as being the case.⁷

4. The Epistemic Thesis (ET) and Our Puzzle from Section 1

Again letting “P □ → Q” denote “would” counterfactual conditionals like (A), and letting “P ◇ → Q” denote “might” counterfactuals like (B), the Epistemic Thesis is:

(ET) (P ◇ → Q) ≡ ⟨e⟩(P □ → Q),

where “⟨e⟩ϕ” indicates an epistemic possibility that ϕ, and where epistemic possibilities are understood to be possibilities of the kind that sentences of the form, “It's possible that P_{ind},” typically express. Such sentences express possibilities that are tightly connected to the concept of knowledge, and so are quite properly called “epistemic.” But I will leave open the *exact* nature of these possibilities' connection to knowledge until section 10, where a more precise understanding

will be useful. Until then, all we need is the idea that what a speaker asserts in saying, “It’s possible that P_{ind} ,” entails that the speaker doesn’t know that P is false. What else such an assertion might entail can be left open for now.

ET shares some important features with Robert Stalnaker’s account of the relation between “might” and “would” counterfactuals, as some might have noticed. In fact, it can be fairly characterized as a modification—hopefully, an improvement—of Stalnaker’s account. But there are crucial differences, which, I will argue, render ET superior. It proves most efficient to first develop ET the way I think it is best worked out, and then to explain the differences between it and Stalnaker’s account in sections 6–7, where I argue that these differences result in important advantages for ET over Stalnaker’s theory.

On ET, (C), from section 1 above, expresses the epistemic possibility that (A) is false.⁸ The puzzle from section 1, then, can be assimilated to that of section 3. (C + A) is perfectly consistent on this account, and the inference from (C) to the falsehood of (A) is invalid. But, on this approach, we have an *explanation* for why (C + A) *sounds* inconsistent: It’s a conjunction of an assertion together with an admission of the epistemic possibility that the assertion is false, and *any* such conjunction will sound inconsistent, for reasons we saw in section 3.

We face the same issues regarding knowability and assertability that confronted us in section 3. I suggest the same contextualist resolutions to these issues. But however the story turns out for the knowability and assertability of (A), ET provides a way of avoiding the *really* nasty conclusion—that (A) is false.

Question: If, as ET would have it, the skeptical reasoning of section 1 is so similar to that of section 3, why is this pattern of argument so much less persuasive when it’s aimed at the truth of (A’) than it is when its target is (A)? Answer: Because, prior to the argument, in the relevant situations, it’s far more obvious that either (A’) or (not-A’) is true than it is that either (A) or “If I had tagged up, I would not have scored the winning run” is true. Thus, it is easier for the argument to challenge the conviction that one of the latter pair is true.

5. ET vs. DT

First, a quick remark about symbolism is in order. Though I follow Lewis in my use of “ $\Box \rightarrow$ ” and “ $\Diamond \rightarrow$ ”, those symbols for “would” and “might” counterfactuals probably would not have been initially proposed by an advocate of a view like ET, and seem more appropriate if one is believer in DT, given the usual philosophical practice of using boxes to indicate necessity and diamonds for possibility. On ET, while “ $\Diamond \rightarrow$ ” seems a very apt symbol for “might” counterfactuals, “would” counterfactuals would then be more happily indicated by a simple arrow, “ \rightarrow ”. A related quick remark about initial appearances is also in order. To my ears, at least, while “might” counterfactuals clearly do include an element of possibility, there is no suggestion of a corresponding necessity in a “would” counterfactual. This is just to register my own sense that initial

appearances seem to favor a view like ET (or like Stalnaker's theory, which we'll encounter below), which keeps suggestions of anything like necessity out of "would" counterfactuals, over a view like DT. To initial appearances, or at least to how things initially appear *to me*, while DT might be an initially promising view about the relation between "might" counterfactuals, on the one hand, and something like "would necessarily", or perhaps "would certainly", counterfactuals on the other, it seems quite unpromising as an account of the relation between "might" counterfactuals and unadorned "would" counterfactuals like our (A).

Moving beyond initial appearances, let us now review the two views' handling of our puzzle from section 1 to see how they compare in that regard. A real solution to our puzzle must not only pick out which of the initially plausible statements constitutive of the puzzle to deny, but must explain (away) the plausibility of what it denies. ET solves our puzzle: It explains why the likes of (C + A) seem inconsistent while they're in fact consistent. DT does not solve our puzzle. Or at least it's not been shown how to solve our puzzle if we embrace DT. Rather, it forces us to face difficult choices, and provides little comfort as we swallow hard. Other things being equal, a solution that solves our puzzle is to be preferred to one that, for all we can see, doesn't.

Why accept DT? In [8], Lewis seeks to support DT by means of a single case, which, he claims, DT gets right, while it's botched by DT's rivals, including ET ([8], pp. 80–81). But, as I argue in [3], ET actually handles Lewis's case nicely, while, ironically, DT is refuted by a quite natural specification of Lewis's own (quite underdescribed) example.⁹

But Lewis's argument involving that single case can't explain DT's popularity (a popularity that has survived Lewis's own apparent abandonment of DT, as we'll see in section 9, below). I think DT's real attraction is that "might" counterfactuals do seem to be—as DT says they are—inconsistent with the corresponding "would not" counterfactuals, and "might not" counterfactuals seem inconsistent with the corresponding "would" counterfactuals. That is, despite the "initial appearance" that "would" counterfactuals contain no suggestion of necessity, still when such a counterfactual is conjoined with the corresponding "might not" counterfactual, the resulting conjunction—(C+A), above, for instance—certainly sounds inconsistent, as DT predicts. But as we've seen, that carries little to no weight against ET, because, given ET, we would expect the relevant conjunctions to seem inconsistent. Though we should perhaps prefer a theory that makes our intuitions come out true over one that explains them away, that's not the situation we face here. DT saves our intuitions concerning the contradictoriness of the likes of (C + A) only at the cost of violating other intuitions that it cannot explain away, while ET makes sense of all the intuitions involved by endorsing some and explaining away the plausibility of what it rejects. So not only does ET solve our puzzle, but it solves it in a way that undercuts the sole source of support for DT that remains after Lewis's argument from his single case has been debunked.

6. Stalnaker's Thesis (ST)

It's time to consider other rivals to ET, starting with Robert Stalnaker's thesis (ST) on the relation between "might" and "would" counterfactuals, articulated and defended in [12].¹⁰ Actually, as I've already noted, ET is close enough to ST that it can be fairly characterized as a modification of ST, rather than as a rival to it.¹¹ But ET is in that case a modification that improves on ST by solving its two most serious problems.

Stalnaker holds that " $P \diamond \rightarrow Q$ " expresses a possibility that " $P \Box \rightarrow Q$ " (p. 145); this is an aspect of ST that ET mimics. How close to Stalnaker's view ET comes depends upon what kind of possibility of " $P \Box \rightarrow Q$ " Stalnaker thinks " $P \diamond \rightarrow Q$ " expresses. His contention is that "*might*, when it occurs in conditional contexts, has the same range of senses as it has outside of conditional contexts" (p. 143). What is that range? Stalnaker thinks that, while "might" most commonly expresses epistemic possibility (p. 143), the range of possibilities "might" can express also includes non-epistemic and "quasi-epistemic" possibilities.¹² Here's what he says about the non-epistemic possibilities that "might" can express:

But *might* sometimes expresses some kind of nonepistemic possibility. *John might have come to the party* could be used to say that it was within John's power to come, or that it was not inevitable that he not come. (p. 143)

So, presumably, since the range of possibilities " $P \diamond \rightarrow Q$ " expresses is the same as those expressed by non-conditional *might*'s, sometimes " $P \diamond \rightarrow Q$ " expresses just such non-epistemic possibilities of " $P \Box \rightarrow Q$ " as those described above. At any rate, that's what I'll take to be Stalnaker's official view, and I'll label it "ST." So, while, according to ET, " $P \diamond \rightarrow Q$ " always expresses an epistemic possibility of " $P \Box \rightarrow Q$ ", on ST, it sometimes expresses a non-epistemic possibility of " $P \Box \rightarrow Q$ ".¹³ It's this that causes ST's second main problem.

Stalnaker's first main problem is that he is working with an impoverished notion of epistemic possibility. This infects his account of non-conditional modal statements as well as his account of "might" counterfactual conditionals. For Stalnaker, an assertion that P is possible in the epistemic sense means that P "is compatible with what the speaker knows" (p. 143), while on ET, epistemic possibilities are those possibilities that sentences of the form "It's possible that P_{ind} ," typically express. As I said at the start of section 4, such sentences do typically entail that the speaker doesn't know that P is false, so Stalnaker's notion of epistemic possibility does have something in common with the notion of epistemic possibilities utilized by ET. But, on the account of what the relevant sentences mean that I present below in section 10, and argue for in [2] and [5], the epistemic possibilities expressed by the relevant sentences often go far beyond Stalnaker's meager notion of P's being compatible with what the speaker knows. On my more flexible analysis of epistemic possibilities, Stalnaker's notion is just one special case of what epistemic possibilities can amount to. There are *many* examples of

uses of modal statements which we *surely* should classify as epistemic, but which don't fit Stalnaker's narrow characterization of epistemic possibility,¹⁴ and Stalnaker's claim that "might" typically expresses epistemic possibility isn't true if epistemic possibilities are understood in his narrow way. But, as I've argued for the superiority of my account of epistemic possibilities over Stalnaker's elsewhere,¹⁵ and as this problem of Stalnaker's is not particular to his account of "might" counterfactuals, we should here move on to ST's second main problem.

7. ST and the Problem of the Inescapable Clashes

That second problem is that conjunctions of forms

- (i) $(P \Box \rightarrow Q) \& (P \Diamond \rightarrow \sim Q)$

and

- (ii) $(P \Box \rightarrow \sim Q) \& (P \Diamond \rightarrow Q)$

always clash or seem inconsistent, and that wouldn't be so on ST. To see why, look at what happens when *non*-conditional "might have" statements are put into some conjunctions.

Stalnaker writes that the conjunction

- (E) John might come to the party, although he won't

is "somewhat strange," and that this strangeness indicates that the epistemic interpretation of its first conjunct,

- (F) John might come to the party,

is "the dominant one for this example" (p. 143). For present purposes, what's important is not how to treat (E) and (F), but what by contrast happens to "might have" versions of them. Stalnaker reports that,

- (G) John might have come to the party, although he didn't

is "not so strange," which he takes to indicate that the epistemic interpretation of *its* first conjunct,

- (H) John might have come to the party,

is "less dominant" than is the epistemic interpretation of (F).

(G) is indeed "not so strange." This is partly because we can easily think of circumstances in which it's a perfectly fine thing to say. However, some may find

that their ears don't simply give (G) a clean bill of health because there are also ways of interpreting it such that it "clashes." What's going on is that (G)'s first conjunct, (H), is ambiguous. In the mouth of someone who can't remember whether John came to the party, (H) can express the epistemic possibility that John came to the party, and in a conversational context in which a topic of conversation is whether John did or did not come to the party, (H) will be interpreted as expressing just such a possibility. But from someone who knows that John wasn't at the party, but who is wondering what John would have done last night if he had been invited to the party, (H) expresses something quite different, and in conversational contexts where, for instance, a topic of conversation is what John would have done if he'd been invited, (H) will be interpreted as expressing this different thing, which seems compatible both with the second conjunct of (G) and with the what a speaker represents in asserting the second conjunct of (G).¹⁶

If we imagine (G) being asserted in the first type of conversational setting, where (H) will be interpreted as expressing the epistemic possibility that John came to the party, then (G) indeed does sound strange. (And this for the reasons given in sections 3–4.) But if (G) is asserted in the second type of context, where the other interpretation of (H) is called for, (G) sounds perfectly fine. If (G) is asserted in a context where it isn't clear which interpretation is called for, it may sound a bit strange, but I think we then naturally try to interpret (H) in the second way so that (G) makes sense. If (G) is considered out of context, as a philosophical example, then we get the results I've described above: It seems not so strange, because we realize it's often a perfectly fine thing to say, but the fact that it can also be interpreted in a way in which it would "clash" causes us to be a bit uneasy about it.

If ST were correct, and "might" counterfactuals sometimes expressed a non-epistemic possibility of their corresponding "would" counterfactuals, then, it seems, conjunctions of forms (i) and (ii) would behave like (G). The second conjunct of such a conjunction would have one (epistemic) reading which would cause a pragmatic clash with the first conjunct, since the second conjunct would deny the knowledge that the speaker would represent herself as having in asserting the first conjunct. But that second conjunct would also have another reading (see note 16, above) which would make it perfectly consistent both with the first conjunct and with what the speaker would represent as being the case in asserting that first conjunct. But type (i) and type (ii) conjunctions don't behave like (G). (C + A), which is of type (i), and

- (I) If he had been invited, John might have come to the party; nevertheless, he wouldn't have come if he had been invited,

which is a type (ii) conjunction, *invariably* sound awful.¹⁷

Stalnaker could perhaps respond that the reason (G) sounds acceptable is that the epistemic interpretation of its "might have" conjunct, (H), is not dominant. By contrast, he might claim, the epistemic interpretation of a "might have" counterfactual like

(J) If he had been invited, John might have come to the party

(the first conjunct of (I)) is dominant to some greater degree than is the epistemic interpretation of (H). This *might help a bit* in explaining the difference in how the conjunctions (G) and (I) strike us when we're considering them out of context. Perhaps we then naturally interpret their modal conjuncts in their most dominant form. But if we imagine (I) occurring in a context where a non-epistemic interpretation of (J) (if such an interpretation is possible) would be called for, (I) should there sound not so bad. Recall that when (G) is considered in a context in which its non-epistemic interpretation is called for, it sounds *perfectly fine*. That's perhaps a bit too much to ask for where the dominant interpretation is the epistemic one that produces a clash. Maybe with (I) we should expect a twinge of uneasiness, even in contexts conducive to a non-epistemic interpretation that should make (I) unproblematic. This twinge can be produced by the fact that, even though the epistemic interpretation isn't contextually relevant, it is the dominant reading, and may therefore intrude into our intuitions even in contexts where it's not the reading the context calls for.

Maybe. But "twinge of uneasiness" doesn't begin to cover it. You try it. If you think that (J) has a non-epistemic interpretation which makes (I) unproblematic, then devise a context which would most strongly call for such an interpretation. Then consider (I) in such a context. I think you'll find that (I)—and (C + A)—still sound absolutely terrible.¹⁸

Perhaps you, like me, and like Stalnaker, have a sense that, bad as they sound, type (i) and type (ii) conjunctions seem like they might nevertheless be true. Then you should *love* my treatment of them, according to which they can often be true (though you'll always at least represent something false as being the case if you try to assert them). But don't let your suspicion that they just might be true deafen you to the evident fact that they *sound* just awful.

What's left to be said in response? That "might" counterfactuals have an interpretation that makes type (i) and type (ii) conjunctions unproblematic, but that they never, in any context, manage to take on that interpretation? Or that they cannot take on that interpretation when they are conjoined with some other claim? Surely we don't want to resort to that!

Type (i) and type (ii) conjunctions invariably clash. That casts doubt on any theory of "might" counterfactuals that allows that they sometimes take on an interpretation that would make the conjunctions unproblematic. ST is one such theory. Most of ET's other main rivals share this problem with ST.

8. The Non-DT Version of Heller's Theory and the Problem of the Inescapable Clashes

In [7], Mark Heller puts forward what is potentially another non-epistemic reading of "might" counterfactuals. His main contention is that " $P \diamond \rightarrow Q$ " is true iff Q is true in at least one of the *close enough* P worlds, where a P world can at

least sometimes be close enough even though there are other P worlds that are closer to the actual world than it is. The cases Heller uses to argue for this contention do not present an obstacle to thinking ET always governs “might” counterfactuals. He trades in cases in which “might” counterfactuals are true, though they’d be false if “ $P \diamondrightarrow Q$ ” required for its truth that Q be true in some of the *closest* P-worlds. However, ET issues the intuitively correct verdict that the “might” counterfactuals of Heller’s cases are true, so these cases don’t count against ET.

Heller’s view isn’t clearly contrary to DT. In fact, Heller explicitly leaves open the possibility that “ $P \squarerightarrow Q$ ” is true iff Q is true in all the *close enough* P worlds, an account which, together with his above contention regarding the semantics for “ $P \diamondrightarrow Q$ ”, preserves the duality relation. However, he also leaves open the possibility that a less demanding semantics—one on which all that’s required for the truth of “ $P \squarerightarrow Q$ ” is that Q be true in all the *closest* P worlds—is correct for “would” counterfactuals. It’s only in the latter case that Heller’s account would be an alternative to DT.

Would such a non-DT version of Heller’s theory be viable? On any version of Heller’s view in which the set of P-worlds relevant for evaluating “ $P \diamondrightarrow Q$ ” are always the same as those relevant to “ $P \squarerightarrow Q$ ” (the first sentence requiring for its truth that some of those worlds also be Q-worlds, the second requiring that they all be Q-worlds), the duality relation will be retained (along with DT’s problems). This even if the range of relevant P-worlds varies with context, so long as the ranges of P-worlds relevant to the evaluation of the two types of conditionals vary together.¹⁹ So, to get a non-DT version of Heller’s view, there should be contexts in which the range of P-worlds relevant to the “might” counterfactuals is different from (no doubt broader than) the range of worlds relevant to the “would” counterfactuals. In such contexts, type (i) and type (ii) conjunctions should be unproblematic. But there are no such contexts; these conjunctions always clash. So any non-DT version of Heller’s view will succumb to the problem of inescapable clashes—the same problem that ST fell to.

9. Lewis’s Ambiguity Thesis (AT), His “Would-Be-(Non-Epistemically)-Possible” Reading of “Might” Counterfactuals, and the Challenge to ET

In [8], Lewis seems to hold that DT governs all uses of “might” counterfactuals. But in the later [9], Lewis explores the possibility that “might” counterfactuals are ambiguous between one “not-would-not” sense governed by DT, and another, “would-be-possible,” sense (see pp. 63–64). We’ll call this ambiguity thesis “AT”. The second sense there explored is, according to Lewis, not epistemic (see especially p. 64, note 8), since the possibility involved amounts to there being a non-zero chance—where the chance in question is an objective, single-case chance—of the occurrence of what’s being said to be possible (see pp. 62, 64). This non-epistemic type of possibility is one that avoids collapsing into actuality only in cases of indeterminism. According to Lewis, then, “ $P \diamondrightarrow Q$ ”

sometimes expresses that if P had been the case, then there would have been a non-zero objective chance that Q would have been the case or, what I take it amounts to at least roughly the same thing, that if P had been the case, then it would not have been causally determined that Q wouldn't be the case.

Ultimately AT falls to the same problem of “inescapable clashes” that plagues ST and the non-DT version of Heller’s theory, as I’ll argue below in section 13. But AT poses a challenge to ET, because there certainly are occasions on which “might” counterfactuals seem to have at least something close to the second sense alleged by AT, and these are problem cases for ET.

Suppose, for instance, that yesterday two physicists were about to roll a certain ball through the “zone of indeterminism” they had set up in their lab. When a ball hits this zone, it either veers left or veers right, but it is not causally determined which way it will veer until the ball is well into the zone. But suppose that just before they were able to roll the ball, a fire alarm sounded, and our physicists had to flee their lab. Today, they might well say such things as

(K) If we had rolled the ball into the zone, it might have veered left

and

(L) If we had rolled the ball into the zone, it might have veered right.

These are not straight-forward counter-examples to ET, because their assertions seem true, and, given that our characters don’t know which way the ball would have veered, ET yields that intuitively correct verdict. (In section 13, we’ll consider what one might say in such a case if one does think one does know which way the ball would have veered.) Still, these are problem cases for ET, for to many it will seem that in our physicists’ mouths, these “might” counterfactuals can mean something pretty non-epistemic—something at least along the Lewisian lines of “If we had rolled the ball into the zone, there would have been a non-zero, objective chance that it would have veered left (right)” or “If we had rolled the ball into the zone, it would not have been causally determined that it would not veer left (right).”

10. Flexing ET to Meet the Challenge: Non-Conditional Possibilities

ET posits a tie between “might” counterfactuals and sentences of the form “It’s possible that P_{ind} ”. It’s important to note that apparently non-epistemic uses of the latter are generated in contexts similar to those that generate seemingly non-epistemic uses of the former. Thus, our physicists can seem to mean the non-epistemic “not causally determined not to be” or “there’s a non-zero objective chance that” if, before rolling a ball into their zone, they say the non-conditional,

(M) It's possible that the ball will veer left

or

(N) The ball might veer left.

Since I'm here giving the name of "epistemic" possibility to those possibilities expressed by sentences of the form "It's possible that P_{ind} ," then perhaps there's hope that what such problem cases call into question is not the tie between "might" counterfactuals and epistemic possibilities, but rather the epistemic nature of "epistemic" possibilities.

But maybe not even that. Looking at uses of "It's possible that P_{ind} ," which seem pretty clearly to be epistemic, in [2] I argued for the "Flexible Hypothesis" that

(FH) S's assertion, "It is possible that P_{ind} ," is true if and only if (i) No member of the relevant community knows that P is false and (ii) There is no relevant way by which members of the relevant community can come to know that P is false,

where both the issue of who is and is not a member of the relevant community and of what is and is not a relevant way of coming to know are very flexible matters that vary according to the context of the utterance of the epistemic modal statement (see [2], esp. pp. 593–594). In [5], I argue that uses of "It's possible that P_{ind} ," like (M), and corresponding non-conditional "might" statements, like (N), even in the problem cases like that of our physicists, are governed by FH, and so the possibilities expressed really are epistemic in nature. These results can be extended to "might" counterfactuals, as I'll argue below in sections 12–13, so it's worth quickly reviewing the situation of non-conditional possibilities.

Key to my treatment in [5] of the problematic uses of "It's possible that P_{ind} " is the flexibility that's posited in the matter of what's to count as a relevant way of coming to know in the second clause of FH. (The flexibility of who's to count as members of the relevant community needn't concern us here, so long as we observe the side-constraint that the speaker is herself a member. We may suppose that as the scientists of our story converse, the relevant community relative to which their epistemic modal statements are to be understood includes just the two of them.) A tremendous amount of flexibility is exhibited even in clearly epistemic uses of "It's possible that P_{ind} ," as I observe in [2]. Sometimes the relevant sentences mean roughly that we (the members of the relevant community) don't know that not-P, and can't come to know that not-P in any *readily available* way. In other contexts, however, *not-so-readily-available* ways of coming to know are relevant. Sometimes, for instance, very particular ways of coming to know, which we have especially in mind, are relevant in context, while other, simpler ways of coming to know are irrelevant.

Given all this flexibility, we can understand even the problematic uses of “It’s possible that P_{ind} ” and corresponding non-conditional “might” sentences as being governed by FH, I argue in [5], if we allow some *very* inaccessible ways of coming to know to count as relevant. In contexts like that of our physicists, I argue, prominently included among the “relevant ways of coming to know” that the ball won’t veer left is: by deducing (if it’s deducible) what will happen from the all the relevant facts concerning the present state of affairs together with all the relevant laws of nature. On that suggestion, (M) and (N), in this context, mean “(i) We (the scientists) don’t know that the ball won’t veer to the left, and (ii) We can’t come to know that it won’t veer to the left by deducing (if it’s deducible) what will happen from the all the relevant facts concerning the present state of affairs together with all the relevant laws of nature,” where—and this is important—the force of the “can’t” here isn’t just that they can’t deduce what will happen because of their own present limitations, in the way that one can’t bisect an angle with a compass and ruler when one’s only compass is broken or when one hasn’t yet learned how to perform the task, (though in another context, their modal sentences might mean that they can’t so deduce because of such personal limitations), but that they can’t deduce what will happen in the way that one can’t trisect an angle with a compass and ruler—because it’s not deducible. (“Can” and “can’t”, of course, are paradigm cases of semantically flexible words.)

Admittedly, the “way of coming to know” here is so inaccessible to the members of the relevant community (the scientists) that it can only in a very thin sense be classified with respect to them as a “way of coming to know.” What the second truth-condition (condition (ii)) amounts to here is really just that it’s not causally determined that the ball won’t veer left. But that helps to explain why the problematic sentences, in context, can seem to mean “not determined not to be”: Since it’s commonly known by our scientists that neither of them knows which way the ball will veer, and it’s thus presupposed that the first clause of the above truth conditions is met, the point they’d mainly be trying to convey by asserting (M) or (N) is that the second condition is met—that it’s not causally determined that the ball won’t veer left.

Still, the thinness of this “way of coming to know” may make one question just how epistemic is the “epistemic” possibility that our scientists are expressing, if the above account of their meaning is correct. But the important issue here is not how proper it is to call the possibility they express “epistemic”. What’s important is that the forms of sentences in question (indicative possibilities and simple “might”s) are *not ambiguous*. There is a gradual continuum of how accessible are the “ways of coming to know” involved in uses of the sentences in question, no sharp division that could be used to mark a borderline between different “senses” of “might” and “possible”.

And, at any rate, even in the case of our scientists, there’s this good reason for calling the possibilities they express “epistemic”: The undeniably epistemic first truth condition—that they don’t *know* that the ball won’t veer left—is still a part of the meaning of their utterance, even though, being common knowledge to

them, it's not the point they'd be mainly interested in conveying in asserting (M) or (N). This is shown by the "clashes" that are produced when, even in contexts like that of our scientists, modal claims like (M) or (N) are conjoined with claims to know otherwise, or with flat-out assertions that things are otherwise, than what these sentences say is possible. If one of our scientists—we'll call him Newton—believes that he knows that the ball will veer right, though it's not now causally determined that it will (perhaps because God is reliably revealing to him that the ball will veer right), he might properly say,

(O) I know the ball won't veer to the left, though it's not now causally determined that it won't

and

(P) The ball won't veer to the left, though it's not now causally determined that it won't.

While saying these things may be based on some strange beliefs of the scientist, he's making no linguistic error in how he expresses his beliefs, and there doesn't seem to be any clash or contradiction involved in either of them. By contrast,

(Q) I know the ball won't veer to the left, but it might

(R) The ball won't veer to the left, but it might.

(S) I know the ball won't veer to the left, but it's possible that it will

and

(T) The ball won't veer to the left, but it's possible that it will,

do seem to clash. Now, given the context, if Newton *were* to utter one of (Q)-(T), we'd probably be able to figure out what it is he's *trying* to say—in each case he'd be trying to say something roughly like (O) or (P). But (Q)-(T) are wrong here. Why? Because the modal second conjunct of each is invariably governed by FH, and is therefore invariably at odds with the knowledge either claimed or represented in the (assertion of) the first conjunct, for reasons we saw in section 3.

If a simple *might* or a statement of the form "It's possible that P_{ind}" *could* mean anything like "not causally determined not to be," then (Q)-(T), we'd expect, would here succeed in expressing the consistent (and pragmatically consistent) thought voiced by (O) or (P), and would be acceptable here. The context strongly suggests such a reading, and the modal statements are in conjunctions where such a non-epistemic reading is needed for the conjunction to make sense. That they're wrong even here is strong support for the claim that such sentences are invariably governed by FH.²⁰

11. Flexing ET to Meet the Challenge: “Might” Counterfactuals

It’s now clear how the proponent of ET should meet the challenge issued in section 9, above. That challenge, recall, was to explain why, in certain contexts, like that of our scientists after they had to flee their lab before performing their experiment, the “might” counterfactual,

(K) If we had rolled the ball into the zone, it might have veered left,

can seem to mean that if they had rolled the ball into the zone, it would not have been causally determined that would not have veered left, and, more generally, to explain why in some contexts, “might” counterfactuals can seem to take on the second sense alleged by AT.

On ET, (K) expresses the epistemic possibility of

(U) If we had rolled the ball into the zone, it would have veered left.

So, on the best—FH—construal of epistemic possibilities, the content of (K) in our scientists’ mouths, is that (i) they don’t know that (U) is false, and (ii) there is no contextually relevant way by which they can come to know that (U) is false. But which ways of coming to know *are* contextually relevant here? Taking our cue from the treatment of non-conditional possibilities in section 10, above, we’d expect that prominently included among the “relevant ways of coming to know” that (U) is false is: by deducing (if it’s deducible) what would have happened from the all the relevant facts concerning the state of affairs that would have obtained when they rolled the ball into the zone together with all the relevant laws of nature. Pulling all this together, then, (K), *in context*, means:

- (i) We (the scientists) don’t know that the ball wouldn’t have veered to the left if we had rolled it into the zone, and (ii) We can’t come to know that it wouldn’t have veered to the left by deducing (if it’s deducible) that it would not have veered left from the all the relevant facts concerning the state of affairs that would have obtained if we had rolled the ball into the zone together with all the relevant laws of nature,

where—and this is again important—the force of the “can’t” in clause (ii) isn’t just that they can’t deduce what will happen because of their own present limitations, but rather because it’s not deducible. Given that understanding of “can’t”, what clause (ii) amounts to in this context is the second, “would-be-(non-epistemically)-possible” sense of “might” counterfactuals that AT alleges.

So we can explain why “might” counterfactuals like (K) can in some contexts (like that of our scientists) seem to take on that second sense, and we can thus meet the challenge registered in section 9 in a way that closely parallels our treatment of problematic non-conditional modal statements from section 10: Since

it's commonly known by our scientists that neither of them knows which way the ball would have veered, and it's thus presupposed that the first clause of the above truth conditions is met, the point they'd mainly be trying to convey by asserting (K) is that the second condition is met. Thus, since, as we saw, that second clause in context corresponds closely to the second sense alleged by AT, AT's second sense would be the main point be conveyed by our scientists' use of (K). It's no wonder, then, that (K) seems to take on that meaning in that context.

12. ET vs. AT: Preliminaries

So, from the perspective of ET, we can explain why “might” counterfactuals can seem to take on the second sense that AT alleges them to have. Indeed, it was to be expected that they would seem to take on such a sense, since, in similar contexts, expressions of the epistemic possibility of non-conditional propositions can seem to take on a related metaphysical sense involving indeterminism or non-inevitability, and, according to ET, “might” counterfactuals express the epistemic possibilities of the corresponding “would” counterfactuals. But even if we can now see how the problematic uses of “might” counterfactuals can be accommodated within ET, why *prefer* ET over AT? After all, of course, AT also accounts for why “might” counterfactuals can seem to take on the second sense that AT alleges them to have, and accounts for it in a more straightforward way: They seem to take on that sense because they do take on that sense; that sense is one of the things “might” counterfactuals can mean.

It can be said on ET's behalf that, unlike AT, it accounts for the problematic uses of “might” counterfactuals without postulating an ambiguity in “might” counterfactuals. Here it's handy to view AT as a descendant of DT—which, historically, it is. The “problematic” uses of “might” counterfactuals—the uses in which they seem to take on Lewis's “would-be-(non-epistemically)-possible” sense—are problematic for both ET and for DT. AT is a modification of DT which “handles” this problem by simply introducing a new sense for the “might” counterfactuals and claiming they are ambiguous. ET, by contrast, is able to handle the problem without making such a radical move. In fact, as noted in the above paragraph, given what happens with expressions of the epistemic possibilities of non-conditional propositions, such “problematic” uses of “might” counterfactuals were to be expected. It would have been a cause for concern for ET if such “problematic” uses did *not* occur.

It will be objected that while ET doesn't postulate an ambiguity to handle the problematic uses of “might” counterfactuals, it does make use of a *lot* of flexibility to account for them. The response is that this is a flexibility that we need anyway to best account for the operation of a variety of modal sentences which express epistemic possibility. Given the variety of cases that can be handled by (FH), especially by utilizing the flexibility it exhibits in clause (ii), it is safe to conclude that positing this flexibility is well-motivated long before we come to the case of “might” counterfactuals.

But this line of argument depends on the success of (FH) in other areas, and I shouldn't repeat the whole case for that success here. (See especially [2], but also [5].) Though the issues of how ET's handling of "might" counterfactuals fits into a broader approach to a wide variety of modal sentences are extremely important selling points for ET, for present purposes, some more direct evidence favoring ET over AT is called for. Again taking our cues from the case executed in [5] and described above in section 10 for the epistemic character of some non-conditional modal statements, that evidence is to be found in the problem of the inescapable clashes.

13. ET, Its Rivals, and the Problem of the Inescapable Clashes

So, again following the treatment of section 10, suppose now that Newton believes that, even though it would not have been causally determined which way the ball would have veered if it had been rolled into the zone, still, there is a fact of the matter as to which way it would have veered, and that God, being both existent and omniscient, knows what that fact is. What's more, suppose that Newton thinks that God is reliably revealing to him that the ball would have veered to the right, not the left, and that, therefore, he, Newton, now knows this to be so. Reporting all this, Newton might properly say,

- (V) If we had rolled the ball into the zone, it would not have veered to the left, though, if we had rolled it into the zone, it wouldn't have been causally determined that it wouldn't veer to the left

Saying this may be based on some strange beliefs of Newton's, but he's making no linguistic error in how he expresses his beliefs, and there doesn't seem to be any clash or contradiction involved in his utterance. But consider this troubled conjunction—which replaces the second conjunct of (V) with (K)—as uttered by Newton in such a context:

- (W) If we had rolled the ball into the zone, it would not have veered to the left, though it might have veered to the left if we had rolled it into the zone.

If "might" counterfactuals *could* take on AT's second, "would-be-(non-epistemically)-possible" sense, and could mean something like "it would not have been causally determined not to be", then the second conjunct of (W)—(K)—would take on that meaning here and (W) would succeed in expressing the thought voiced by (V), and would be acceptable here.²¹ The context strongly suggests such a reading, and such a reading is needed to avoid a "clash" with the first conjunct of (W). Now (to echo my comments in section 10 about (R) and (T)), if Newton *were* to utter (W) in such a context, we'd probably be able to figure out what he was *trying* to say—he'd be trying to say something like

(V). But, as I hope you can sense, even here, (W) is wrong; it clashes. In fact, it sounds just about as bad as does (C + A) from section 1. That (W)'s second conjunct—(K)—fails even here, where such a sense is so strongly called for, to take on the second sense that AT alleges it to have gives us strong grounds to conclude that “might” counterfactuals like (K) just don't have such a sense, and that AT is false.

By way of contrast, recall again the behavior of

(H) John might have come to the party,

which really is ambiguous. On one of its readings (the one on which it expresses the epistemic possibility of “John came to the party”) (H) is incompatible with the speaker's knowing that John didn't come to the party, but on its second reading (see again note 16), it is perfectly compatible with the speaker's knowing that John didn't come. The conjunction of (H) with the assertion that John didn't come,

(G) John might have come to the party, although he didn't,

when said in a “friendly” context—one which calls for the second sense that would make the conjunction even pragmatically consistent—sounds perfectly fine, though (G) sounds less fine if we imagine it being said in a context that calls for the first reading of (H). We are now considering (W) in what should be a “friendly” context if AT is true—a context which strongly calls for a reading of its second conjunct, (K), that should make (W) unproblematic like (V). If (K) really were really ambiguous in the way AT alleges, (W) in Newton's mouth should strike us much the way that (G) does in “friendly” contexts. But (W) is far from perfectly fine.

The defender of AT may appeal here to the fact that, even where it's given an unfriendly reading, (G) is not a genuine contradiction, but only a pragmatic clash. By contrast, (W), according to AT, expresses a genuine contradiction on the unfriendly reading of its second conjunct. Perhaps, it may be suggested, if a conjunction has an ambiguous conjunct that has one reading on which the conjunction expresses an outright contradiction, that conjunction will strike us as fishy even in friendly contexts in which it is both semantically and pragmatically consistent, while if a conjunction only produces a pragmatic clash on an unfriendly reading, it will sound OK on its friendly reading. (G), on this not implausible suggestion, is not a good model for how (W) should be expected to behave given AT. Fair enough. But if we beef (G) up to

(G') John might have come to the party, but I know that he didn't,

then it expresses an outright contradiction on the unfriendly reading of its first conjunct—the reading on which that conjunct expresses the epistemic possibility

that John came to the party. Still, in a friendly context, where that first conjunct takes on its other sense, (G'), like (G), sounds perfectly fine. In fact, in such contexts, it's often just the thing to say. If (AT) were correct, (W) should behave like (G'). But it doesn't. Not by a long shot.

So, (W) clashes. What's more, closely in line with that observation, but perhaps not quite as intuitively clear, (W)'s second conjunct,

(K) If we had rolled the ball into the zone, it might have veered left,

seems not to be something Newton should assert when he takes himself to know that the first conjunct of (W) is true, i.e., when he takes himself to know that the ball wouldn't have veered left. But, on AT, (K) has a reading, in fact the reading that is strongly called for in Newton's context, on which it's true, and there's no apparent reason on AT why (K) wouldn't be just the thing for Newton to say.

But if "might" counterfactuals like (K) were instead governed by ET, we could explain why (W) clashes even in Newton's context in the same way that we explained, in section 4, the clash of (C + A) (though, recall, this explanation makes the clash a pragmatic one, and not a matter of genuine inconsistency), and we would have a ready explanation for why (K) isn't the thing for Newton to say.

To wrap up the results of our comparison of AT with ET, though in some contexts "might" counterfactuals can *seem* to take on the second sense that AT alleges, ET can account for this, and, in fact, given what happens with other modal sentences, predicts this seeming. What's more, even in the contexts where this "second sense" is most strongly called for, the unassertability of "might" counterfactuals like (K) in situations where AT predicts they'd be assertable, and their inability to happily enter conjunctions that AT predicts would be unproblematic, shows that AT doesn't really get the content of the "might" counterfactuals quite right. Though it may not always be part of the point a speaker is primarily intending to make, that he doesn't know that the corresponding "would" counterfactual is false *is* part of the content of his assertion of a "might" counterfactual.

In short, like most of ET's other rivals, AT falls to the problem of the inescapable clashes—the problem that conjunctions of forms

(i) $(P \Box \rightarrow Q) \ \& \ (P \Diamond \rightarrow \sim Q)$

and

(ii) $(P \Box \rightarrow \sim Q) \ \& \ (P \Diamond \rightarrow Q)$

invariably clash or seem inconsistent, though most of the theories we've looked at predict that such conjunctions should be acceptable in various contexts. The only theories that escape this problem are ET, on which the conjunctions only

seem to be inconsistent, and DT, on which they really are inconsistent.²² And we've already seen in section 5 why ET is to be preferred over DT.

14. ET, Indeterminism, and the Return of the Counterfactual Skeptic

But it's a bit too quick to simply say that on ET the conjunctions only seem inconsistent. The meaning of "might" counterfactuals is highly context-dependent, and it's quite possible that *sometimes* "might" counterfactuals *in context* express thoughts inconsistent with the corresponding "would not" counterfactuals (and that "might not" counterfactuals sometimes express thoughts inconsistent with the corresponding "would" counterfactuals). For instance, on our account of the content of our scientists' use of

(K) If we had rolled the ball into the zone, it might have veered left,

it means:

- (i) We (the scientists) don't know that the ball wouldn't have veered to the left if we had rolled it into the zone, and (ii) We can't come to know that it wouldn't have veered to the left by deducing (if it's deducible) that it would not have veered left from all the relevant facts concerning the state of affairs that would have obtained if we had rolled the ball into the zone together with all the relevant laws of nature,

where the "can't" is understood in such a way that clause (ii) amounts to the claim that it would not have been causally determined that the ball would not veer left. *Perhaps*, so understood, clause (ii) here really is inconsistent with

- (X) If we had rolled the ball into the zone, it would not have veered to the left.

The counterfactual skeptic may sense some hope here. Perhaps, she may claim, while "might" counterfactuals aren't *invariably* inconsistent with the corresponding "would not" counterfactuals, they *sometimes* are (and "might not" counterfactuals are likewise sometimes inconsistent with their corresponding "would" counterfactuals), and it was such a use of the "might" counterfactuals that she had in mind when she was arguing for the non-truth of the "would not" conditional.

ET is powerless to completely extinguish this hope. And that's as it should be. If the counterfactual skeptic can make the case that

- (Y) If we had rolled the ball into the zone, it would not have been causally determined that it wouldn't veer to the left,

which contains no “might”s, is inconsistent with (X), then no account of the relation between “might” and “would” counterfactuals should have the power to deprive her of this argument, so long as the argument doesn’t go through a “might” counterfactual, exploiting our sense that conjunctions of forms (i) and (ii) are inconsistent. Perhaps some have a direct intuition that (X) can’t be true if (Y) is true. Perhaps some will have the strong sense that if (Y) is true, then there can be nothing that can ground the truth of (X)—nothing that can make it be the case that (X) is true—and that (X) can’t be true without such a grounding.²³ Perhaps some will be moved by some other argument, involving no “might”s, from the truth of (Y) to the non-truth of (X). ET is silent on all of this.

But insofar as the counterfactual skeptic does utilize a “might” counterfactual like

(Z) If we had rolled the ball into the zone, it might have veered left

to mediate the passage from the truth of the likes of (Y) to the non-truth of the likes of (X), ET advises us to be suspicious of her maneuver. And, even where no explicit argument going through (Z) is in play, still, insofar as one’s sense or intuition that (X) can’t be true where (Y) is true is being implicitly influenced by one’s sense that (Z) is inconsistent with (X), one has cause to be suspicious of this “intuition”. For even in contexts in which *no* ways of coming to know are contextually relevant to the meaning of the relevant “might” counterfactuals, and so the second clause of FH completely “drops out”, the “might” counterfactuals will still produce a clash of apparent inconsistency when conjoined with the corresponding “would not” counterfactuals, for reasons we saw in sections 3–4—reasons based entirely on the first clause of the meaning of the “might” counterfactuals, a clause the holding of which is clearly consistent with the truth of the “would not” statements, so long as omniscient God is kept out of the relevant community. (And, of course, the same point holds for conjunctions of “might not” with “would” counterfactuals.) So (Z) would *appear* to be inconsistent with (X) even if it were perfectly consistent with it. And, in general, conjunctions of forms (i) and (ii) will seem inconsistent even in contexts in which they are perfectly consistent (a fact that helps explain the popularity that DT enjoyed). So, in other contexts, in which it is arguable that those conjunctions really are inconsistent, we have good reason to distrust any sense we might have that they are. Moral: To avoid being misled by a trick of language, when you try to discern whether the likes of (X) can be true where the likes of (Y) hold, keep the likes of (Z) completely out of your mind.

How much of a limitation does this put on the allure of counterfactual skepticism? Quite a bit, I think. For (X) certainly doesn’t *sound* inconsistent with (Y). As I’ve already noted,

- (V) If we had rolled the ball into the zone, it would not have veered to the left, though, if we had rolled it into the zone, it wouldn't have been causally determined that it wouldn't veer to the left,

which is the conjunction of (X) and (Y), doesn't "clash" in the way that

- (W) If we had rolled the ball into the zone, it would not have veered to the left, though it might have veered to the left if we had rolled it into the zone,

the conjunction of (X) and (Z), does. If (X) and (Y) are somehow inconsistent, it is a different kind of inconsistency, less analytical, more metaphysical, perhaps. But I haven't seen any even reasonably convincing arguments to the effect that the likes of (V) are metaphysically impossible, much less any arguments approaching the power that would be needed to support the havoc such a conclusion would wreak on our judgments regarding the truth values of "would", "would not", "might", and "might not" counterfactual conditionals.²⁴

Notes

1. Conditionals like, "If I had been a bachelor, then I would have been unmarried," where the antecedent entails the consequent, seem immune to this type of attack, for the skeptic will have a hard time generating much enthusiasm for the likes of "If I had been a bachelor, then I might not have been unmarried."
2. In employing this maneuver my imagined counterfactual skeptic follows the lead of the very actual Robert M. Adams; see especially [1], p. 111.
3. [8], pp. 2, 21–24. On p. 2, Lewis introduces this as a simulative definition of his connective, $\diamond\rightarrow$. His formulation: $(P \diamond\rightarrow Q) = \text{df } \sim(P \square\rightarrow \sim Q)$. But by pp. 80–81, if not before, the tie he posits between $(P \diamond\rightarrow Q)$ and $(P \square\rightarrow Q)$ is certainly being presented as a substantive account of the relation between "would" and "might" counterfactuals in ordinary English.

DT runs into some trouble with counterpossibles—counterfactuals with impossible antecedents. For, where P is impossible, then, as the standard semantics for counterfactuals is usually developed, both $(P \square\rightarrow Q)$ and $(P \square\rightarrow \sim Q)$ are (trivially) true. But then, DT has the result that both $(P \diamond\rightarrow Q)$ and $(P \diamond\rightarrow \sim Q)$ are false. But it is absurd to hold that $(P \square\rightarrow Q)$ is true while $(P \diamond\rightarrow Q)$ is false. This problem is easily fixed. The adherent to DT can follow Mark Heller (though Heller doesn't endorse DT) by stating that where P is impossible, then $(P \diamond\rightarrow Q)$ is true, and allowing such cases as exceptions to the duality thesis ([7], pp. 96–97), or, if one wants to continue to define "might" counterfactuals in terms of "would" counterfactuals, one can accept the following modification of DT, due to Edward Wierenga ([14], p. 94): $(P \diamond\rightarrow Q) \equiv \sim(P \square\rightarrow \sim Q) \vee (P \square\rightarrow Q)$.

4. See especially Adams's [1], pp. 110, 113, which initiated this line of argument. See also Hasker ([6], p. 28), Zagzebski ([15], pp. 134–135), and van Inwagen ([13], pp. 231–232), who follow Adams. Adams (see [1], p. 115) and Zagzebski (see [15], especially p. 141) allow—unwisely, I think—for some true counterfactuals regarding what creatures would have freely done, and thus aren't complete counterfactual skeptic-

tics, even with respect to counterfactual free actions. But they do use the pattern of argument I'm here discussing (involving DT) to attack the truth of other such "would" counterfactuals involving free actions.

I should point out that these skeptics have other grounds for their skepticism. In Adams's case, his *main* argument involves claims that there is no adequate ground or basis for the truth of the counterfactuals he's dealing with (involving free actions)—nothing that makes them true (see [1], esp. pp. 110–111). Though responding to (or even adequately explaining) this very different argument is beyond the scope of the present paper, let me here, however briefly and cryptically, point in the direction I'm inclined to take in response. The main difference between my views and Adams's, I believe, is that I think of "would" counterfactual conditionals as being "thinner" (easier to make true) than Adams does. Thus, I don't think it takes very much to make a "would" counterfactual true. What for Adams grounds the truth of a "would probably" counterfactual will often for me be adequate to ground a "would" conditional.

This "thinness" of "would" counterfactuals, however, makes them inadequate to be the objects "middle knowledge", if middle knowledge is to play the role it is typically assigned in Molinistic, no surprises, no dice-throwing accounts of God's exercise of providential control. I join Adams in denying that God has "middle knowledge", where middle knowledge is construed as knowledge of conditionals that would enable Him to exercise the providential control that he's assigned by Molinists. But I affirm, while Adams denies, God's middle knowledge, where this is construed as God's knowledge of "would" counterfactuals, like (A) of the present paper and the stock examples of "middle knowledge" in the literature—e.g., Adams's "If President Kennedy had not been shot, he would have bombed North Vietnam" ([1], p. 109) and Alvin Plantinga's "If Curley had been offered \$20,000, he would have accepted the bribe" ([10], p. 174). Since I take the first of the above two construals of "middle knowledge" to be the more important, I take myself to be on the basic side of Adams in the "middle knowledge" debate. But since the present paper is on "would" counterfactuals, I am here opposed to Adams.

(Some will wonder which conditionals I think God *would* have to know to exercise Molinistic providential control. Rather than past-directed counterfactuals, like "If President Kennedy had not been shot, he would have bombed North Vietnam" and "If Curley had been offered \$20,000, he would have accepted the bribe", I think God would need to know ahead of time (with Divine certainty) such future-directed *indicatives* as "If Kennedy is not shot, he will bomb North Vietnam" and "If Curley is offered \$20,000, he will accept the bribe." Indeed, it's a bit of a mystery to me why the whole middle knowledge debate has been barking up the wrong conditional tree, given that the past-directed counterfactuals seem relevant to providential control only to the extent that they are past-tense versions of the future-directed indicatives (that are true afterwards iff the indicatives were true beforehand), and given the dim prospects for any such a relation to hold between the two types of conditionals. While I have much I could say about this, let me here make do with the following speculation. I think this misdirection is at least partly due to the fact that (the 20th Century version of) the middle knowledge debate was kicked off at a time when counterfactuals were fairly well understood, and were a fairly hot topic, to boot, while indicative conditionals were relatively ill-understood and obscure. The debaters then were well-motivated to frame their discussion in terms of counterfactual conditionals.)

I should finally add that, though my position on the "thinness" of "would" counterfactuals renders them inadequate to be the objects of Molinistically providentially

useful “middle knowledge”, they still can play many of the *other* theoretical roles to which counterfactuals are often put, I think. Indeed, taking them to be as “thick” as Adams thinks they are may disqualify them from being put to some of these other uses.

5. This hypothesis may seem quite far-fetched in ordinary conversations, but it is mild in comparison with such philosophical creatures as the hypotheses that one is a bodiless brain in a vat or the victim of an all-powerful deceiving deity.
6. In [4] (see especially sections 11–12, pp. 33–38), I show how the raising of skeptical hypotheses raises the standards for knowledge via what I there call the “Rule of Sensitivity.”
7. See [2], pp. 596–601.
8. Assuming ET, (C) expresses *at least* the epistemic possibility that (A) is false. If the Principle of Conditional Excluded Middle is correct, then (C) expresses exactly that much. If the Principle is false, then (C) expresses the epistemic possibility of a clear contrary to—though not the contradictory of—(A).
9. Incidentally, in [3], p. 417, fn. 6, I refer to a paper of mine entitled, “‘Might’ and ‘Would’ Counterfactual Conditionals: The Epistemic Connection.” That paper has undergone a change in title and is the paper you have before you now.
10. The relevant portion of [12], on pp. 142–146, is, with the exception of some very minor alterations, identical to the corresponding material on pp. 98–101 of Stalnaker’s earlier [11].
11. Matters are complicated, however, by some lack of clarity in [12]: At one point, Stalnaker seems to embrace ET itself. See note 13, below.
12. [12], pp. 143–146. See note 13, below, for an explanation of Stalnaker’s “quasi-epistemic” possibilities.

I’m following Stalnaker, then, in construing “might” counterfactuals as expressing the possibilities of the corresponding “would” counterfactuals, and even in thinking of the possibility expressed as being at least usually epistemic (though I diverge from Stalnaker, at least on the official ST reading of him, in holding that the possibility is *always* epistemic). There’s another important way I’ve followed Stalnaker’s treatment. While he doesn’t seek to explain why conjunctions of the forms (i) and (ii) (see the beginning of section 7, below) seem inconsistent (why the duality relation seems to hold), and so doesn’t explain how his view can undercut the apparent support of DT or solve the puzzle of section 1, he does explain the problematic nature of conjunctions of the form

(iii) $\sim(P \Box \rightarrow Q) \ \& \ (P \Diamond \rightarrow Q)$

in roughly the same way I’ve followed in my explanations of (i) and (ii) (p. 144). Thus, the undercutting of DT’s support and the solving of our puzzle from section 1 are further applications of good Stalnakerian methodology.

While under a theory like ST or ET, the treatment of (iii) can be very similar to the treatment of (i) and (ii), things are not so similar from the perspective of DT. DT has no trouble with the problematic nature of (i) and (ii) type conjunctions—they seem inconsistent because they are inconsistent, according to DT. But, as Stalnaker argues, DT provides no good explanation for the troubles with conjunctions of type (iii). That’s yet another strike against DT. So ST or ET’s handling of (i) and (ii), which I’ve here provided, is needed for defensive purposes to match DT, but their handling of (iii) can be used, as Stalnaker uses it, for an offensive attack on DT.

13. But wait! After laying out ST, and after giving no indication that he's about to switch views, Stalnaker writes, "The main evidence that *might* conditionals are epistemic..." (p. 144), as if he instead means to defend something like ET, a view according to which " $P \diamond \rightarrow Q$ " always expresses an epistemic possibility of " $P \Box \rightarrow Q$ ". I'll reserve the title "ET" for the epistemic thesis as I think it's best developed. Let's then label as "ST2" the view that results from interpreting Stalnaker as holding that " $P \diamond \rightarrow Q$ " never expresses nonepistemic possibility.

Any plausible interpretation of Stalnaker must reflect the fact that he thinks that " $P \diamond \rightarrow Q$ " sometimes expresses a "quasi-epistemic" possibility of " $P \Box \rightarrow Q$ ". Stalnaker's gloss on what we'll call expressions of "straight-forward" epistemic possibilities (as opposed to "quasi-epistemic" possibilities) that P is that they express that P "is compatible with the speaker's knowledge" (p. 143). He introduces the notion of a "quasi-epistemic" possibility as follows:

Consider not what is, in fact, compatible with my knowledge, but what would be compatible with it if I knew all the relevant facts. This will yield a kind of quasi-epistemic possibility—possibility relative to an idealized state of knowledge. If there is some indeterminacy in the language, there will still remain some different possibilities, even after all the facts are in, and so this kind of possibility will not collapse into truth (p. 145).

On ST2, then, " $P \diamond \rightarrow Q$ " expresses either a straight-forward epistemic possibility or a quasi-epistemic possibility of " $P \Box \rightarrow Q$ ". ST, then, adds another option; according to it, " $P \diamond \rightarrow Q$ " can express either of the two things ST2 allows, but it can also express a non-epistemic possibility of " $P \Box \rightarrow Q$ ". It's the addition of this third option which gets ST into trouble, as I'll argue below in section 7.

14. The modal statements in the following examples in [2] are just a few such examples: Case CTC-1B (pp. 584–585), Hacking's Salvage Ship Case (pp. 586–587), Case CTC-2B (p. 587), and the Revised Salvage Ship Case (p. 590). The modal statements in each case are of the form "It's possible that P_{ind} ," but the arguments involving them would be just as strong with analogues of the sentence which feature "might" rather than "possible".
15. See [5], especially section ix, pp. 79–81.
16. Stalnaker would say that this different thing that (H) then expresses is a non-epistemic possibility of John's having come to the party—a possibility of power or of non-inevitability, perhaps. My own suspicion about such non-conditional "might have" statements is that they're ambiguous between two different *epistemic* possibilities: Where (H) doesn't express the epistemic possibility of "John came to the party," it expresses the epistemic possibility of "John would have come to the party." This difference between Stalnaker and me shouldn't matter to the argument that follows. What does matter is what the two views share: That (H) is ambiguous between an expression of the epistemic possibility of "John came to the party" and *some* other reading on which it can unproblematically conjoin with the other conjunct of (G), because it's perfectly consistent both with John's not having come to the party and with the speaker's knowing that John didn't come to the party.

As Stewart Cohen called to my attention, in an episode of the old television series, *Taxi*, a joke was built around this ambiguity in "might have" sentences. Jim, at the end of an account of a wonderful fling he had had a long time ago, declares, "I might have married that girl." Elaine, thinking Jim surely knows whether he married

the girl, assumes that he is not expressing the epistemic possibility of “I married that girl,” and is instead using “might have” in its other sense, and replies, “Oh, Jim, why didn’t you?” The joke is that Jim has indeed forgotten whether he married that girl, and was indeed expressing the epistemic possibility of “I married that girl”; he answers, as best as I can remember, “I didn’t say I didn’t marry her. I said I might have.”

17. Some find the “clash” involved even more jarring if we utilize Stalnaker’s device of breaking up the conjunction in a dialogue, so that one speaker can conjoin a “yes” or “no” answer to a question regarding one conjunct with an assertion of the other conjunct. Stalnaker uses this device to show the trouble with what, in note 12 above, we called a type (iii) conjunction (p. 144); we’ll now apply it to the type (ii) (I):

A: Might John have come to the party if he had been invited?

B: Yes, but he would not have come if he’d been invited.

18. When checking conjunctions for “clashes”, one should employ the methodology of “flat-footed, all-in-one-breath” conjunctions, that I describe and motivate in [5], pp. 70–72—that is, one should say the conjunctions quickly and without any unusual stress, intonation, or emphasis. The reason for this is that the methodology of conjunctions utilizes the general presumption that the parameters along which the meanings of context-sensitive terms vary are held constant through conjunctions said “all in one breath”; however, emphasis is a device often used to defeat that presumption—to indicate that one is switching meanings in mid-sentence. Thus, to test for semantic ties and clashes between potentially context-sensitive terms, we should utilize conjunctions said “all in one breath”, but without emphasis. Thus, for instance, it’s cheating, in checking (I) for a “clash”, to emphasize the “might” in its first conjunct, while the second conjunct is said flat-footedly. Such emphasis can perform the function of indicating that the epistemic standards—the standards for what counts as knowledge—that are being applied to the first conjunct are higher than are those being applied to the second. This could unfairly reduce the sense of “clash” that the conjunction produces, for the first conjunct will be interpreted as saying, roughly, that the speaker does not know *according to very high standards* that it’s not the case that John would have come to the party if he had been invited, while, in asserting the second conjunct, the speaker would only be representing himself as knowing *according to lower standards* that John would have come if invited. The speaker would then be completely consistent, with respect to both what he asserts and what he represents as being the case, but none of that is to the point. What’s relevant for our semantic purposes is whether the sentence clashes *where the standards for knowledge are held constant*.
19. Thus, if the range of relevant P-worlds is context-sensitive, the duality relation is retained even if asserting a “might” counterfactual has a different effect on what that range is than would be the effect of asserting a “would” counterfactual. So long as there’s at any one time just one group of relevant worlds, however dynamic its membership, duality is maintained. If, on the other hand, there are potentially two groups of relevant worlds, one for “might” and the other for “would” counterfactuals, but there’s a defeatable presumption that these groups will contain the same worlds, duality can fail. The presumption may cause the appearance of duality, but there would be contexts where the presumption is defeated and duality fails. The problem with such a view, if I’m right that type (i) and type (ii) conjunctions always clash, is that there seem to be no such contexts. This should make us conclude either that there’s just one range of worlds relevant to both types of conditionals—which lands us in DT—or that

- there's some other explanation for the clash of type (i) and (ii) conjunctions—e.g., the explanation I've given in sections 3 and 4.
20. See [5] for a more thorough statement of this argument.
 21. We perhaps shouldn't expect on AT that (W) will be as unproblematic as is (V), since on AT, (W)'s second conjunct is ambiguous and has a reading on which (W) is inconsistent. But AT should lead us to expect is that (W), as uttered by our scientists, shouldn't out-and-out clash, because it's there being asserted in a "friendly" context—a context which calls for a reading of its "might" conjunct which would make the conjunction unproblematic. What we really should expect (W) to behave like is (G'), below in the text.
 22. Another *possible* theory that would avoid the problem is one on which "might" counterfactuals are ambiguous between one sense governed by DT and another governed by ET. But, of course, ET is a more efficient and lovely theory than is the imagined ambiguity theory, so, unless some phenomena can be adduced that this theory can handle better than does ET, or some other relative advantage can be adduced for the imagined theory, ET is to be preferred over it.
 23. Here again my counterfactual skeptic follows the lead of Adams in [1] (see note 4, above), though Adams was interested in cases of counterfactuals concerning the free actions people would have performed in counterfactual situations.
 24. Thanks to Robert M. Adams, Graeme Forbes, Mark Heller, and David Lewis for insightful and helpful comments on earlier versions of this paper.

References

- [1] Adams, Robert M., "Middle Knowledge and the Problem of Evil," *American Philosophical Quarterly* (B1.A4) **14** (1977): pp. 109–117.
- [2] DeRose, Keith, "Epistemic Possibilities," *Philosophical Review* (B1.P5) **100** (1991): pp. 581–605.
- [3] ———, "Lewis on 'Might' and 'Would' Counterfactual Conditionals," *Canadian Journal of Philosophy* (B1.C3) **24** (1994): pp. 413–418.
- [4] ———, "Solving the Skeptical Problem," *Philosophical Review* (B1.P5) **104** (1995): pp. 1–52.
- [5] ———, "Simple 'Might's, Indicative Possibilities, and the Open Future," *Philosophical Quarterly* (B1.P48) **48** (1998): pp. 67–82.
- [6] Hasker, William, *God, Time, and Knowledge* (Ithaca: Cornell UP; BT131.H18 1989).
- [7] Heller, Mark, "Might-counterfactuals and Gratuitous Differences," *Australasian Journal of Philosophy* (B1.A8) **73** (1995): pp. 91–101.
- [8] Lewis, David, *Counterfactuals* (Oxford: Basil Blackwell; BC199.C66 L48 1973b).
- [9] ———, "Postscript to 'Counterfactual Dependence and Time's Arrow,'" in D. Lewis, *Philosophical Papers*, Vol. II (Oxford: Oxford UP, 1986; B29.L49 1983), pp. 52–66.
- [10] Plantinga, Alvin, *The Nature of Necessity* (Oxford: Clarendon Press, 1974; BC199.M6 P55).
- [11] Stalnaker, Robert, "A Defense of Conditional Excluded Middle," in W.L. Harper, R. Stalnaker, G. Pearce, eds., *Ifs: Conditionals, Belief, Decision, Chance, and Time* (Dordrecht: D. Reidel, 1981; BC199.C56 I38), pp. 87–104.
- [12] ———, *Inquiry* (Cambridge, MA: MIT Press; BD161.S675 1984).
- [13] van Inwagen, Peter, "Against Middle Knowledge," *Midwest Studies in Philosophy* (B1.M5) **21** (1997): pp. 225–236.
- [14] Wierenga, Edward, "Theism and Counterpossibles," *Philosophical Studies* (B21.P53) **89** (1998): pp. 87–103.
- [15] Zagzebski, Linda, *The Dilemma of Freedom and Foreknowledge* (New York: Oxford UP; BT131.Z34 1991).