On the Plurality of Counterfactuals*

Ben Holguín and Trevor Teitel

October 2026

Abstract

Counterfactuals are context-sensitive. However, we argue that various debates and doctrines in metaphysics and the philosophy of science are premised on ignoring the full extent of counterfactual context-sensitivity. Our focus is on the prominent "miracle" versus "no-miracle" debate about counterfactuals under the assumption that our laws of nature are deterministic. But we also discuss doctrines that employ counterfactuals in theories of rational decision, as well as doctrines that explain what it is to be a law of nature in terms of counterfactuals. We conclude by drawing some morals about the use of counterfactuals in philosophical theorizing more generally.

1 Introduction

Counterfactuals are context-sensitive. Take Quine's famous pair:

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

b. If Caesar had been in command in Korea, he would have used atomic bombs, not catapults.

It is easy to imagine contexts in which one would speak truly in asserting (1a). And it is easy to imagine contexts in which one would speak truly in asserting (1b). But it is difficult to imagine contexts in which one would speak truly in asserting both (1a) and (1b). So neither (1a) nor (1b) is true simpliciter; rather, which (if either) is true depends on the context in which it is used. In general, the question of the truth of a given counterfactual is notoriously slippery, and very often dependent on features of context.

We doubt there are many contemporary philosophers who would disagree with these claims in the abstract. But we will argue that much recent work in metaphysics and the

*Penultimate draft. Final version forthcoming in *Philosophical Review*. The authors contributed equally. For helpful comments and discussion, many thanks to David Albert, Bob Beddor, Harjit Bhogal, Kyle Blumberg, David Builes, Michael Caie, Rosa Cao, Sam Carter, Mark Crimmins, Cian Dorr, Samuel Elgin, Augie Faller, Simon Goldstein, Jeremy Goodman, Al Hájek, Thomas Icard, Zachary Irving, Nico Kirk-Giannini, Marc Lange, Matt Mandelkern, Trenton Merricks, Jack Mullaney, Antonia Peacocke, and Eran Tal, to audiences at Johns Hopkins, McGill, Princeton, Stanford, UT Austin, and the University of Virginia, and to this journal's editorial team and two anonymous reviewers.

philosophy of science fails to come to grips with just how context-sensitive counterfactuals really are. Indeed, our central contention in this paper is that certain prominent debates and doctrines in these areas are premised on ignoring the full extent to which counterfactuals are context-sensitive.

The central debate we have in our sights is between the so-called "miracle" and "no-miracle" views about the truth-conditions of counterfactuals under the assumption that our laws of nature are deterministic (henceforth 'determinism'). We'll argue that this debate makes sense only if the full extent of the context-sensitivity of counterfactuals is ignored. And we'll argue that once it is no longer ignored, it becomes unclear whether there's anything left to debate. We'll then suggest that similar morals can be drawn for other prominent applications of counterfactuals.

We'll discuss only one other application at length: attempts to ground facts about the laws of nature in facts about counterfactuals, defended chiefly by Marc Lange.[1] But in the conclusion we'll draw a general moral of our discussion that shows how our arguments generalize. This moral is that we should be especially wary of philosophical theories that tie facts about (relatively) fundamental features of reality to facts about counterfactuals. Ordinary counterfactual thought and talk is just too badly behaved to admit such connections. And though there may be *certain* specialized uses of the counterfactual for which such connections hold, properly delineating the relevant class will result in trivializing the putative connections. Or so we'll argue.

2 The Miracle Versus No-Miracle Debate

Our primary target is the miracle versus no-miracle debate. This debate stems from two supposed platitudes about counterfactuals that, as it turns out, can't both be true given determinism. The competing positions then diverge over which of the two supposed platitudes ought to be given up in light of the conflict.

Keeping things rather rough for now (we'll sharpen the discussion shortly), the first of these two platitudes is (LAWS), which holds that the worlds that witness the truth of a counterfactual preserve the actual laws of nature.[2] The second platitude is (PAST), which holds that the worlds that witness the truth of a counterfactual preserve the truths about the actual past—at least up until the earliest time relevant to the truth of the counterfactual's antecedent.

[1] We will also raise some general worries about the connections between counterfactuals and rational decision-making, though in a more tentative spirit than our criticisms of Lange's views.

[2] In this paragraph and elsewhere we talk in a way that presupposes a possible worlds semantics for the counterfactual, according to which '$a \mathrel{\square\!\!\rightarrow} c$' is true iff every member of a certain (possibly singleton) set of a-worlds is a c-world. In this setting, the "worlds that witness the truth of a counterfactual" are just these a-worlds. Our view is that this way of thinking and talking about counterfactuals makes the arguments of the paper more vivid and easier to express. That said, it's meant to be a purely notational convenience; we could, if necessary, work entirely with first-order statements about which counterfactuals are true or false. Thanks to an anonymous referee for encouraging us to clarify this aspect of our presentation.

To see why (LAWS) and (PAST) are taken to be platitudes, consider counterfactuals like (2) and (3):

(2) If I had conducted the double-slit experiment yesterday, I'd have observed an interference effect.

(3) If I had gone to the gym this morning, it would have been the first time in months.

It is hard to see why counterfactuals like (2) would so readily seem true if there were not some general presumption that, even if things were different from how they actually are, the laws of nature would nevertheless have remained more or less the same. Likewise, it is hard to see why counterfactuals like (3) would so readily seem true if there were not some general presumption that, even if things were different from how they actually are, the past would nevertheless have remained more or less the same, at least up until the earliest time relevant to the counterfactual's antecedent.

Here's the puzzle from determinism (likewise stated rather roughly for now). Given determinism, if the world were to differ in any respect (however minuscule) at *some* time, then the laws could remain true only if the world were also to differ in some respect at *every* other time, including the distant past and far future. Yet counterfactuals are about what would have been the case if things had been different.[3] So we seem faced with a choice: when evaluating counterfactuals under the assumption of determinism, we must either change some feature of the world at every time, including the distant past, or else we must suppose that the laws of nature would have been different. Whichever way we go, we can't cling to both (LAWS) and (PAST).

Note that the puzzle isn't just some hypothetical exercise: there is reason to believe that the actual laws of nature are deterministic. Indeed, just about every candidate physical theory considered to date has this feature. Quantum mechanics is no exception: determinism is preserved by the Everett/many-worlds interpretation, as well as hidden-variables approaches like Bohmian mechanics—arguably the two most popular current understandings of quantum theories. Moreover, even if future developments in physics change the state of play, it's still interesting to investigate this puzzle that arises for counterfactuals given the *epistemic possibility* of determinism. The fact that *for all we know* one of (LAWS) and (PAST) must be given up is not significantly less surprising than the categorical claim. So throughout we will simply assume that our actual laws of nature are deterministic. In some of our examples below we'll appeal to features of classical statistical mechanics, but in general our discussion does not rely on the details of any particular physical theory.

Still, (LAWS) and (PAST) are in need of refinement, since their first-pass forms are subject to obvious counterexamples. With respect to (LAWS), consider so-called "counternomics"—i.e., counterfactuals with nomically impossible antecedents:

[3] Here and throughout we ignore counterfactuals with true antecedents.

(4) If the cosmological constant were twice what it actually is, there wouldn't have been any galaxies.

In considering what would happen were the cosmological constant twice what it actually is, we are considering what would happen if the laws were different from how they are in actuality. So obviously the worlds that witness the truth of this counterfactual do not preserve the actual laws of nature.

Similarly, with respect to (PAST), some counterfactuals explicitly ask us to imagine the past being different at all times, and hence carry no presumption to the effect that the facts about the past are held fixed:

(5) Had every moment of the past been different in some way, things would have been different now too.

Indeed some counterfactual antecedents seem to explicitly contravene both platitudes:

(6) If the universe had consisted of only two qualitatively indiscernible angels singing for all eternity, there wouldn't have been any planets.

We don't take counterfactuals like (4)–(6) to show that there are no platitudes in the vicinity of (LAWS) and (PAST), or, by extension, that there is no puzzle about counterfactuals under determinism.[4] But we do take their existence to suggest that the platitudes we're calling (LAWS) and (PAST) were never meant to apply to *all* counterfactuals whatsoever, but instead only to a circumscribed range of counterfactuals. In particular, we will understand the platitudes as being restricted to counterfactuals whose antecedents express propositions that are: (i) nomically possible and (ii) consistent with some initial segment of the actual past.[5] We'll call a proposition *suitable* just in case it satisfies both (i) and (ii). And we'll call a counterfactual *suitable* just in case, in context, the proposition expressed by its antecedent is suitable.

We can now state (LAWS) and (PAST) more precisely. First, let L be the laws of the actual world. Second, where t is the earliest time relevant to the assessment of the antecedent at issue, let H_t be the conjunction of all true propositions that are entirely about matters that occur at or before t.[6] For example, if t is the present moment, then H_t includes propositions about past macroscopic states of affairs, like that there was a presidential election

[4] We do, however, take the existence of such counterfactuals to cast some doubt on how joint-carving principles in the vicinity of (LAWS) and (PAST) can really be. But our main objection makes for a less crude version of this argument, so we don't want to make too much of these present observations.

[5] The notion of consistency in this latter requirement can be spelled out more carefully in terms of metaphysical compossibility. But nothing in our argument turns on the more careful formulation, so we will work with the rough gloss in the main text.

[6] The notion of "the time relevant to the assessment of a counterfactual antecedent" is admittedly a bit opaque. However, we are content to follow in the tradition of the literatures we'll be engaging with and work with an intuitive, even if somewhat imprecise, understanding of the notion. The examples we'll employ that threaten (PAST) will often mention specific times in their antecedents, which will make it mostly unambiguous which times are earlier than those relevant to the assessment of the antecedent. Moreover, provided that the state of the universe in the *very distant* past—say, the time of the big bang—counts as being determinately earlier

in 2024, as well as propositions about the arrangement of particular microscopic particles moments after the big bang. Finally, let '$\boxrightarrow$' denote the counterfactual conditional. Then we have:[7]

(LAWS) For every context c, if 'a' expresses a false suitable proposition in c, then '$a \boxrightarrow L$' is true in c.

(PAST) For every context c, if 'a' expresses a false suitable proposition in c, then '$a \boxrightarrow H_t$' is true in c.

The puzzle about determinism goes through essentially as before.[8] Suppose that 'a' expresses a false but suitable proposition. And let 'p' be any sentence that expresses a proposition that implies that things are exactly as they actually are at some time prior to t. Given that the actual laws L are deterministic, if '$a \boxrightarrow L$' expresses a truth, then '$a \boxrightarrow p$' must express a falsehood. But if '$a \boxrightarrow p$' expresses a falsehood, then '$a \boxrightarrow H_t$' also expresses a falsehood—H_t is about *all* of the past up to t, after all. Therefore '$a \boxrightarrow L$' and '$a \boxrightarrow H_t$' can't both be true, and so at least one of (LAWS) and (PAST) must be false.[9]

The immediate question that arises at this juncture is which platitude ought to be given up. Both options have been explored. Those who reject (LAWS) and retain (PAST) defend what is typically called a "miracle" account of counterfactuals (we'll explain this label in a moment), whereas those who reject (PAST) and retain (LAWS) defend a "no-miracle" account—hence the "miracle versus no-miracle debate."

The canonical statement of the miracle view—indeed, the statement from which the debate's name is derived—is due to Lewis (1979).[10] On Lewis's version of the view, the falsity of (LAWS) is witnessed by "miracles": small, localized violations of the actual laws of nature. Lewis derives this result from his "four-fold" account of the closeness relation between possible worlds. The details of Lewis' theory needn't concern us here; what matters is that the account predicts that the truth-value of counterfactuals is standardly determined as follows. First, find a world whose past perfectly matches actuality's and that evolves in accordance with the actual deterministic laws of nature up until the time of the antecedent (this guarantees (PAST)). Then introduce a miracle: temporarily suspend the actual laws so as to allow a deviation from the actual course of history, one that brings about the truth of the counterfactual's antecedent (this guarantees a violation of (LAWS)).

than the time relevant to the assessment of mundane counterfactual antecedents concerning present events, the entire dialectic underlying the miracle versus no-miracle debate, and thus our paper, will be in good standing. We thank an anonymous referee for encouraging us to clarify these points.

[7] Here and throughout we suppose that 'L' context-invariantly denotes L and that 'H_t' context-invariantly denotes H_t.

[8] Here we assume that there is no relevant context-sensitivity in 'a', 'p', and so on. For more detailed discussions of the puzzle, see Dorr (2016a) and Gallow (2023).

[9] For reasons of space we will not be exploring views, like Nolan's (2017), that try to retain both principles by positing impossible worlds.

[10] For some other defenses of the miracle view, see Jackson (1977), Halpin (1991), Vihvelin (2000), Beebee (2003) and Glynn (2013).

Then, from that time onward, allow the world to go back to evolving perfectly in accordance with the deterministic laws of the actual world. If, in doing so, the truth of the counterfactual's consequent is brought about, then the counterfactual as a whole is true; otherwise it is false.

What then of the intuitions in favor of (LAWS)? Here the strategy is to replace it with a surrogate principle. Let QL be the disjunction of L and all laws that are just like L except in tolerating the occasional small, localized miracle. Then we have:

(QUASI-LAWS) For every context c, if 'a' expresses a false suitable proposition in c, then '$a \mathrel{\Box\!\!\rightarrow} QL$' is true in c.

(QUASI-LAWS) is just (LAWS) but with the weaker, miracle-tolerating QL in place of L. The principle allows the proponent of the miracle view to claim that, despite appearances, the actual deterministic laws are generally *not* held fixed when assessing the truth of suitable counterfactuals. What is instead held fixed is something sufficiently close to the actual laws: the proposition that the laws of nature are true give or take some small, localized exceptions. The hope is that the truth of (QUASI-LAWS) is enough to account for the intuitive pull of the alleged platitude (LAWS).

That's the miracle view. Proponents of the *no-miracle* view instead accept (LAWS) and give up (PAST), and likewise try to account for the intuitions in favor of (PAST) via a weaker surrogate principle.[11] The surrogate principle is motivated by noticing that (PAST) makes reference to *all* true propositions entirely about the past (up until the time of the antecedent). This leaves open the possibility of giving up (PAST) in its full generality, while still accepting that some restricted class of propositions entirely about the past is held fixed when assessing counterfactuals. In particular, where t is again the earliest time relevant to the assessment of the antecedent at issue, let MH_t be the conjunction of all true propositions that are entirely about *macroscopic* matters that occur at or before t—propositions entirely about tables, chairs, countries, and so on, and not about microscopic particles. One who gives up (PAST) might still accept:

(MACRO-PAST) For every context c, if 'a' expresses a false suitable proposition in c, then '$a \mathrel{\Box\!\!\rightarrow} MH_t$' is true in c.

The presence of (MACRO-PAST) allows the proponent of the no-miracle view to claim that, despite appearances, the totality of the past is *not* held fixed when assessing the truth of counterfactuals. What is instead held fixed is something sufficiently close: the totality of the *macroscopic* past. Once again, this is supposed to account for the intuitive pull of the alleged platitude (PAST).

[11] Defenders of the no-miracle view include Bennett (1984), Goggans (1992), Albert (2000), Kutach (2002), Loewer (2007), Maudlin (2007, ch.1, ch.5), Wilson (2014a), Albert (2015), Goodman (2015), and Dorr (2016a).

We thus have two competing views that supposedly carve out an interesting and open debate about how best to resolve the initial puzzle about counterfactuals under determinism. The proponent of the miracle view offers up (QUASI-LAWS) at the expense of (LAWS) but for the benefit of (PAST); the proponent of the no-miracle view offers up (MACRO-PAST) at the expense of (PAST) but for the benefit of (LAWS).

Whether (QUASI-LAWS) and (MACRO-PAST) can do the theoretical work that their proponents require of them is not an issue we'll spend much time on in this paper. The existing literature gets into these issues in some detail. But for our purposes, the dialectic between proponents and critics of miracles is beside the point. This is because all parties to this debate accept the *disjunction* of (LAWS) and (PAST). But we think that this is a mistake: both of these principles are false, and hence so is their disjunction. Indeed we even think that the weaker disjunction of (QUASI-LAWS) and (MACRO-PAST) is false. The issue is that all of these principles quantify universally over contexts, yet counterfactuals are just too context-sensitive to admit of generalizations of this sort. We'll see that it's easy to get into contexts where we run roughshod over the (quasi-)laws, as well as contexts where we run roughshod over the (macro-)past. So it is simply false that, in assessing suitable counterfactuals, we try to hold either the laws or the past fixed. And so we think the question at the heart of the miracle versus no-miracle debate simply has a false presupposition.

Thus, if the debate is to be preserved, its participants will have to restrict the range of contexts over which the various principles quantify. But we'll argue that there's no tenable way to carve out some restricted range of contexts so as to exclude all of our counterexamples—at least not without either trivially resolving the debate in favor of one view over the other, or rendering it too esoteric to be of broader theoretical interest. Counterfactuals are simply too context-sensitive for there to be an interesting debate here.[12]

We now turn to arguing for these claims (§§3–4). Afterwards we'll reply to various objections (§§5–6). We'll then consider another application of these ideas (§7), before drawing some general morals (§8).

One cautionary note before we begin. Once more: we think counterfactuals are context-sensitive. So we want to signal in advance that for each of the examples we discuss in the following sections, it probably won't be difficult to resolve the context-sensitivity of the counterfactual at issue in a manner other than how we intend to, thereby accessing

[12] We should note that Dorr (2016a), whom we've labeled as a proponent of the no-miracle view, is careful to acknowledge that the context-sensitivity of counterfactuals makes the status of universally quantified principles like (LAWS) and (PAST) somewhat precarious to begin with (see his pp.244-245). Indeed, Dorr prefers to work with an object-level statement of the puzzle (rather than one that involves meta-linguistic ascent, as ours does), with the proviso that "all counterfactuals [relevant to the puzzle] are henceforth to be interpreted according to a "standard" interpretation" (p.245). So although it's clear to us that Dorr rejects (PAST), it's not clear to us whether he would endorse (LAWS) as we've presented it. In any event, we'll have a lot more to say about the so-called "standard interpretation" of the counterfactual over the course of the paper. And our §6.2 will argue that the miracle versus no-miracle debate remains problematic even on Dorr's construal—i.e., when we anchor ourselves into an arbitrary context in which the puzzle of counterfactuals under determinism is gripping.

an alternate reading. Some of these alternate readings may even be driven by the sorts of judgments that motivate (LAWS) and (PAST). We embrace this. Again, (LAWS) and (PAST) are principles that quantify *universally* over contexts. The fact that there are some contexts in which we hold the laws or the past fixed falls far short of vindicating the idea that these are general platitudes governing counterfactual thought and talk. We welcome the malleability of judgments about counterfactuals. What we'll contest is that participants on either side of the miracle versus no-miracle debate are similarly free to do so.

3 Warm-Up: Ramps and Backtracking

This section will focus on some problems for (MACRO-PAST), and hence (PAST), which stem from issues about (i) "ramps" and (ii) "backtracking" interpretations of counterfactuals. The points here are familiar and common ground among all participants in the miracle versus no-miracle debate. Still, our view is that these problems are suggestive of deeper issues with the framing of the debate, issues that we'll bring out with more force in the next section.

(MACRO-PAST) says that, for every context c and sentence 'a' expressing a false but suitable proposition in c, the proposition expressed by '$a \mathrel{\Box\!\!\rightarrow} MH_t$' (i.e., that had a been the case, the macro-past up until t still would have been the same) is true in c. Though this claim may sound reasonable in the abstract, it doesn't take much reflection to notice that, taken literally, the idea that we hold the macro-past fixed all the way up until the time of the antecedent leads to absurdity. Even the most strident defenders of the intuitions behind (PAST) and (MACRO-PAST) must allow that, in assessing counterfactuals, we often make significant adjustments to the pre-antecedent macro-past in order to make counterfactuals *smoothly* true.

For example, suppose it's early July and we're in New York, and everything is essentially normal. Now consider the following two counterfactuals, uttered out of the blue:

(7) a. If we were at Wimbledon right now, we'd get to enjoy watching Alcaraz play tennis.

b. If we were at Wimbledon right now, we'd be completely bewildered, if not simply dead.

(7a) is perfectly sensible; (7b) is not. But if the past must be held completely fixed up until the time of the counterfactual's antecedent—namely the present moment—then since we were in New York one second ago, this would have to remain true when evaluating counterfactuals like (7a) and (7b). And this means that if we were at Wimbledon right now, we'd have to have gotten from New York to Wimbledon in under one second. And however this near-instantaneous transportation happens—whether by nomically possible thermodynamic fluke or nomically impossible (but isolated) miracle—the point is that we

would most certainly not be *expecting* it to happen. (We weren't expecting it a second ago, after all.) And so it looks like in just about every context (7b) rather than (7a) would have to be true, when exactly the opposite seems to be the case.

To get around this problem, some have suggested that counterfactuals routinely tolerate some amount of variation in the facts (including macroscopic facts) about the past—enough variation to get the antecedent to be true in a way that is at least somewhat normal (even at a miracle-world where the actual laws of nature are temporarily suspended).[13] This variation is referred to as a *ramp*. How much variation in the past is tolerated to make room for the ramp is a matter of significant indeterminacy. But we think the intuitive idea should be relatively clear. In the case of (7a) and (7b), for instance, the natural thought is that whatever makes the counterfactual possibility diverge from actuality happens early enough so that we have time to arrive at Wimbledon by conventional means of travel.

To account for ramps, then, the platitude about holding the past fixed must instead be understood as meaning something like "When assessing a counterfactual, hold the past fixed not up until the time of the antecedent, but up until the time of the ramp that smoothly brings about the antecedent." Hence, (PAST) and (MACRO-PAST), as stated above, are straightforwardly false. But it is easy to restate the principles to accommodate ramps. To do so, we just need to modify our definitions of H_t and MH_t so that t is tied to the *ramp* of the antecedent at issue, not simply the antecedent itself. Provided the relevant ramp does not extend all the way back to the initial conditions of the universe, the modified versions of (PAST) and (MACRO-PAST) will preserve the dialectic covered above. By itself, then, the need to accommodate ramps doesn't seem to pose any threat to the credentials of the miracle versus no-miracle debate.

Still, it is common ground in the debate that even the ramp-friendly versions of (PAST) and (MACRO-PAST) can't be quite right. The reason stems from what are called *backtracking* interpretations of counterfactuals. Consider the following example from Jackson (1977):

(8) a. If I had jumped out of this tenth-floor window, I would have died.

b. If I had jumped out of this tenth-floor window, I would have done so only because someone had put a safety net in place.

Assuming the tenth-floor window overlooks pavement, both (8a) and (8b) can be used to express true propositions. As far as (PAST) and (MACRO-PAST) are concerned, there's no special problem with (8a). Perhaps the near-past will have to be changed to get my jumping out of the window to come about in a "reasonable" way. But that's just a ramp. By contrast, (8b) seems to require substantial macroscopic changes to the actual past, changes that more straightforwardly problematize (PAST) and (MACRO-PAST). Not only do we

[13] See, e.g., Lewis (1979), Bennett (1984), Bennett (2003, §§79–81), and Dorr (2016a). For arguments in favor of *not* making significant adjustments to the pre-antecedent past, see Glynn (2013).

have to alter the past to make me willing to jump out of the window (perhaps the building I'm in must catch fire), we also have to alter it so that there's a net there to catch me. The usual thought is that there's a kind of "backtracking" involved here: on the relevant reading of the counterfactual, had things been such that I jumped out of the tenth-floor window, it would have been because the past was in some substantial way other than how it actually is.

The literature on the miracle versus no-miracle debate takes backtracking interpretations of counterfactuals to be reasonably non-standard, and easy enough to spot and quarantine. The usual reaction to the existence of such interpretations is simply to restrict the domain of contexts over which principles like (PAST) and (MACRO-PAST) quantify.[14] In particular, a distinction is drawn between so-called backtracking contexts and so-called *standard* contexts: in standard contexts (8a) is true and (8b) is false; vice versa for backtracking contexts. Principles like (PAST) and (MACRO-PAST) are then recast as universal generalizations exclusively over standard contexts.[15] Supposing backtracking contexts are in fact reasonably special—that standard contexts really are *standard*—it is not unreason-

[14] See, e.g., Lewis (1979), Goggans (1992), Albert (2000), Kutach (2002), Loewer (2007), Wilson (2014a), Dorr (2016a), and Gallow (2023). Khoo (2017) defends a view on which there is no special backtracking context *per se*, but also on which the true readings of examples like (8a) and (8b) involve different resolutions of a time parameter whose value is sensitive to context.

[15] An anonymous referee wonders whether the backtracking/non-backtracking interpretations are best understood in terms of a distinction between contexts. They put forward two alternative proposals in this regard.

> Proposal 1: the backtracking interpretation is tied to the *syntactic* features of the relevant counterfactual conditional—in particular, the presence of language like 'would have had to' or 'would have been because' in the counterfactual's consequent.
>
> Proposal 2: the backtracking interpretation arises whenever the counterfactual's consequent concerns some time before the times relevant to the assessment of the counterfactual's antecedent.

If either of these proposals succeeds, it would undermine our central claim in this section that principles like (PAST) and (MACRO-PAST) are tenable only if restricted to a certain class of contexts. But we are skeptical of both proposals. We will cite a few reasons why here.

We have two complaints about proposal 1. The first is that if the non-backtracking counterfactual 'Had it been that p, it would have been that q' and the backtracking counterfactual 'Had it been that p, it would have had to have been that $\neg q$' could both express true propositions within a single context, then we'd expect them to be felicitously conjoinable. But they appear not to be (cf. Dorr 2016a, p. 244):

(BC) ?? I know that if I had jumped out of this tenth-floor window, it would have resulted in my death. But I also know that if I had jumped out of this tenth-floor window, it would have been because there was a net there to save me.

This is surprising on any view that takes the difference between non-backtracking and backtracking counterfactuals to be lexicalized in the way suggested by proposal 1. Notice, by contrast, that there is nothing comparably infelicitous about conjoining an indicative conditional with its corresponding counterfactual contrary:

(K) ✓ I know that Kennedy was shot to death, so I know that if Oswald didn't shoot Kennedy, then someone else did. On the other hand, I don't know whether if Oswald hadn't shot Kennedy, someone else would have.

able to think that this restriction does little to diminish the interest of the miracle versus no-miracle debate, now understood to concern only which counterfactuals are true when we set aside backtracking readings.

On the other hand, one might be suspicious that the restriction of the debate to standard (i.e., non-backtracking) contexts diminishes its significance. If there are whole swathes of ordinary uses of the counterfactual that must be excluded to get a "puzzle" about (LAWS) and (PAST) going, then one might wonder how deep any such puzzle could really be.

To press this worry, first note that counterfactuals assessed in standard contexts still require ramps. A counterfactual like (7a, 'If we were at Wimbledon right now, we'd get to enjoy watching Alcaraz play tennis') is a paradigm case of a counterfactual whose natural interpretation is a non-backtracking one. But for the reasons discussed above, this interpretation involves a ramp: on the true interpretation of the counterfactual, the near past needs to be adjusted so that our arrival at Wimbledon occurs in a reasonably smooth manner. Plausibly, then, in contexts where (7a) expresses a truth, the following counterfactuals express truths too:

(9) a. If we were at Wimbledon right now, we would have gotten there by some conventional means of travel.

A second problem for proposal 1 is that the 'would have had to have been' and 'would have been because' language is neither necessary nor sufficient for getting a backtracking interpretation of the counterfactual. We'll see a number of counterexamples in both directions over the course of the paper. With respect to necessity, see (9a)–(9c), which we'll discuss shortly below, as well as (18), which we'll discuss in §4.2. With respect to sufficiency, the class of counterfactuals we will introduce in §4.1 (which we will call 'reverse-nomics') has the relevant morphology, but involves tinkering with the laws rather than the pre-antecedent past in the manner characteristic of backtracking interpretations. More directly though: to our ears a counterfactual like 'Had I jumped, I would have had to have died' seems to have a true, non-backtracking interpretation, despite containing the relevant language.

Turning to proposal 2, which ties the backtracking reading to the counterfactual's having a consequent that concerns times earlier than the times relevant to the antecedent, we think here too there are straightforward counterexamples. For one, some counterfactuals are most naturally heard in the backtracking way, despite the fact that their consequents exclusively concern times that are *subsequent* to the time of the antecedent:

(BT) I would never under any circumstances jump from a building unless I knew I was going to land safely in a net. So if I had jumped out of this tenth-floor window, I would have survived the subsequent fall.

For another, there are a number of counterfactuals that are naturally *not* interpreted in the backtracking way, despite having a consequent that concerns times antecedent to the times relevant to the antecedent. Consider, e.g.:

(NBT) If I had jumped, then the previous evening's dinner would have been my last.

(Some of the examples we discuss in §4.2 have this feature as well.)

The referee's alternative suggestions notwithstanding, we are inclined to preserve the assumption—endorsed by many in the literature (see the previous footnote)—that the distinction between backtracking and non-backtracking interpretations of counterfactuals is a matter of the resolution of their context-sensitivity.

b. If we were at Wimbledon right now, then at some point in the past we would have made plans to go.

c. If we were at Wimbledon right now, then we would have had different priorities than we actually ended up having.

These counterfactuals pose a challenge for those who would try to block examples like (8b) by distinguishing different "kinds" of contexts in which we interpret counterfactuals. For if the standard interpretations of a counterfactual like (7a) imply the truth of counterfactuals like (9a)–(9c), then the distinction between backtracking and non-backtracking contexts becomes a matter of degree rather than kind: the distinctive mark of a backtracking context is that it tolerates bigger ramps. And if that's all the distinction between the two "kinds" of contexts really amounts to, then the decision to restrict principles like (PAST) and (MACRO-PAST) to so-called non-backtracking or standard contexts starts to look pretty artificial.[16]

That said, going forward we will set aside issues concerning ramps and whether they threaten the putative distinction between non-backtracking and backtracking contexts. We will concede to both parties in the debate that there is a well-defined distinction between contexts in which we tolerate backtracking interpretations and contexts in which we do not, and that the latter kind of contexts are sufficiently common to preserve the interest of the miracle versus no-miracle debate. However, we will now argue that this falls far short of rescuing the interest of the debate—backtracking interpretations are just the beginning.

4 The Plurality of Counterfactuals

It will be helpful to have a label for those contexts meant to fall under the purview of the miracle versus no-miracle debate. Let us now use the label *standard context* in this general sense (as opposed to defining it purely in terms of backtracking). So far, we have conceded for the sake of argument that standard contexts do not admit backtracking readings of counterfactuals. But we will now argue that the standard contexts must be restricted considerably further if principles like (LAWS), (PAST), (QUASI-LAWS) and (MACRO-PAST) are to have any hope of being true. Indeed, we'll argue that the notion of a standard context must be restricted to such a degree as to be "standard" in name only.

4.1 Reverse-Nomics

Generally speaking, if we are inclined to assert or assent to '$p \boxrightarrow q$', then we believe that the true (partial) answer to the question 'Were p the case, what else would be the case?' is expressed by 'Among other things, q would have been the case'. But questions of this form

[16] Here we take ourselves to be in agreement with Khoo (2017, p.872). See also Hedden (2023, p.747, fn. 29) for related discussion.

are extremely vague, and depending on the circumstances in which they are asked can be understood in a variety of different ways. One natural way to understand the question of whether q would have been the case had p been the case is as a question about what sort of possible *pasts* would have made p most plausible, give or take our knowledge of the actual facts. This way of understanding the question tends to induce backtracking readings.

But there are other, less past-centric ways to understand the question of whether q would have been the case had p been the case. A neglected fact is that this can sometimes be understood as a question about what sorts of *laws* would have made p most plausible, give or take our knowledge of the actual facts. This leads to what we will call the *reverse-nomic* interpretation of the counterfactual. Here are some counterfactuals where the reverse-nomic interpretation seems to us particularly natural:[17]

(10) Had Jesus performed most of the feats attributed to him in the Bible, he would have had to have had magical powers.

(11) Were Michelson and Morley to have measured a significant difference in the speed of light traveling in Earth's direction through the presumed luminiferous aether versus light traveling orthogonal to that direction, it would have been because there really was a luminiferous aether.

(12) Had Fizeau recorded the same results but with a slightly faster-spinning cog than in his actual experiment, the speed of light would have been slightly faster than it actually is.

We think these are perfectly reasonable things to assert or assent to, at least in some fairly ordinary contexts, and thus that each of (10)–(12) has true readings.[18,19] Crucially,

[17] See Hall (2011) for discussion of (10), and Lange (2009, pp.198-200) for discussion of examples like (11) and (12).

[18] We will defend this claim in more detail in §5.3. For now we'll note that we find it especially easy to recover the intended interpretations when one imagines uttering these counterfactuals in answer to questions. Consider the following variant on (12):

> Q: If Fizeau had recorded the same results but with a slightly faster-spinning cog than in his actual experiment, what would the speed of light have been?
>
> A: Had Fizeau recorded the same results but with a slightly faster-spinning cog than in his actual experiment, the speed of light would have been slightly faster than it actually is.

We deny that such dialogues are essential, however. We don't find it particularly difficult to hear the intended interpretations of (10)–(12) out of the blue.

[19] An anonymous referee worries that reverse-nomics like (10)–(12) are not sharply distinguished from backtracking counterfactuals, because reverse-nomics, like backtrackers, often involve the 'would have had to'/'would have been because' language. However, our sense is that this is something of a distraction. We employ this language because it helps make the intended readings salient, not because it is essential for getting the intended readings. Note, for example, that (12) does not involve any such language. (See also the question/answer pair in the previous footnote.) We also think these reverse-nomics are quite unlike backtracking counterfactuals in that their truth doesn't require that the pre-antecedent past be tinkered with in any meaningful way. Indeed, they strike us as counterexamples to the proposal that the 'would have had to'/'would have been because' language is the mark of a backtracking counterfactual, *pace* the suggestion

each of (10)–(12) is also suitable. That is to say: each has an antecedent that in its natural contexts expresses a false proposition that is nomically possible and consistent with some initial segment of the actual past. With respect to the not obvious claim about nomic possibility, the crucial observation is familiar from statistical mechanics: the dynamical laws themselves make an enormous range of propositions concerning non-modal, macroscopic states of affairs nomically possible.[20] You might think that propositions like the following are nomically impossible: that the egg in my frying pan re-collects back into its shell and lands back in its carton, or that this glass of lukewarm water gradually separates into a glass of hot water with some ice cubes floating in it. However, we know from statistical mechanics that these sorts of propositions are not nomically impossible, but merely extremely improbable (on the standard Boltzmann distribution), requiring highly fine-tuned initial conditions. Returning to our present examples, getting Jesus to walk on water is just a matter of fine-tuning the initial conditions of the universe to ensure that a series of extremely improbable (but nonetheless nomically possible) gusts of wind occur under his feet at the right times. And getting the experiments of Michelson and Morley and Fizeau to yield different results is just a matter of introducing certain kinds of human error—nothing as exotic as freakishly low-chance events is needed.

Because (10)–(12) are suitable, each has an interpretation that presents a straightforward counterexample to both (LAWS) and (QUASI-LAWS). Worlds in which Jesus has magical powers, light is a classical wave moving through a luminiferous aether, or the speed of light differs from its actual speed are worlds featuring systematic and widespread violations of the actual laws of nature, not merely the small isolated violations posited by the miracle view. It follows that unless the notion of a standard context is further refined, any principle along the lines of (QUASI-LAWS)—let alone (LAWS)—is simply untenable, and for a far more basic reason than that it is in tension with the combination of (PAST) and determinism.

Of course, we acknowledge that it's not difficult to also hear *false* interpretations of (10)–(12), especially when the nomic possibility of their antecedents is made salient. These interpretations can also be brought out with speeches like:

(10*) Had Jesus performed most of the feats attributed to him in the Bible, it would have been because of a series of extremely improbable but nonetheless nomically possible fluke events.

(11*) Were Michelson and Morley to have measured a significant difference in the speed of light traveling in Earth's direction through the presumed luminiferous aether versus light traveling orthogonal to that direction, it would have been because they made an experimental error somewhere.

of the referee discussed in footnote 15 above.

[20] For a more thorough defense of this claim, see Dorr (2016a, §4). For an accessible introduction to the relevant physics, see Albert (2000).

(12*) Had Fizeau recorded the same results but with a slightly faster-spinning cog than in his actual experiment, it would have been because some other part of the experimental apparatus had changed too.

But the mere fact that there are contexts in which (10)–(12) express falsehoods (namely the contexts in which counterfactuals like (10*)–(12*) express truths) is no reason to deny that there are also contexts in which (10)–(12) express truths. Again, counterfactuals are context-sensitive. Absent an independent explanation of how our judgments on these matters can sometimes be systematically mistaken (an issue we'll investigate in §5.3), we should take the fact that counterfactuals like (10)–(12) seem to have natural interpretations on which they express truths to show that they sometimes *do* express truths, even if the act of explaining what those truths amount to can, in certain settings, make those interpretations harder to access.

In light of examples like (10)–(12), the participants of the miracle versus no-miracle debate will have to understand the notion of a standard context so that it banishes the reverse-nomics along with the backtrackers. We remain suspicious of this approach to handling recalcitrant examples. But for the sake of argument, let us assume that this move is legitimate. Are the principles at the heart of the miracle versus no-miracle debate *now* in good enough standing to deserve serious investigation?

We think not. We'll now argue that, even setting aside backtracking and reverse-nomic interpretations, contexts that undermine the miracle versus no-miracle debate abound.

4.2 What-Ifs

Our next class of examples will draw on Randall Munroe's (2014) *What If? Serious Scientific Answers to Absurd Hypothetical Questions* (as well as some examples from his `https://what-if.xkcd.com/`). As is suggested by the title of Munroe's book, the counterfactuals in question tend to take the form of answers to 'What if?' questions—a general and pervasive part of ordinary counterfactual thinking.

Munroe's analysis of these questions typically puts the reader in a context that seems to favor (QUASI-LAWS) and hence the miracle view: he adduces interpretations that involve small, miracle-like violations of the actual laws, after which things run in accordance with something like actual-world physics (hence his 'serious scientific' qualifier). However, we'll see shortly that it's just as easy to get into contexts where Munroe's what-if questions induce interpretations of counterfactuals that are straightforward counterexamples to (QUASI-LAWS). So again, even given our more restricted conception of standard contexts, both the miracle and no-miracle views remain problematic.

4.2.1 Answering with Miracles

We'll start with the miracle-friendly interpretations Munroe tends to bring out in his analyses of these questions. Consider *Relativistic Baseball* (https://what-if.xkcd.com/1/).[21] This chapter opens with the question, 'What would happen if you tried to hit a baseball pitched at 90% the speed of light?'. We quote Munroe's answer at length (the italics are our emphasis; as usual c is the speed of light in a vacuum):

> Let's set aside the question of how we got the baseball moving that fast. We'll suppose it's a normal pitch, except in the instant the pitcher releases the ball, it *magically accelerates to* 0.9c. *From that point onward, everything proceeds according to normal physics.*

From there Munroe proceeds to describe how the baseball would travel so fast that the air molecules would fuse to it, thereby setting off a chain reaction that would result in a cloud of expanding plasma that:

> . . . hits the bat first, but then the batter, plate, and catcher are all scooped up and carried backward through the backstop as they disintegrate. The shell of x-rays and superheated plasma expands outward and upward, swallowing the backstop, both teams, the stands, and the surrounding neighborhood—all in the first microsecond.

Seconds later:

> Everything within roughly a mile of the park is leveled, and a firestorm engulfs the surrounding city. The baseball diamond is now a sizable crater, centered a few hundred feet behind the former location of the backstop.

Supposing Munroe's analysis is sound, it seems that the following counterfactual is true in the context of *Relativistic Baseball*:

(13) If a pitcher were to throw a baseball at 0.9c, it would cause a massive nuclear explosion.

Our claim is that, at a minimum, the natural miracle-friendly interpretations of counterfactuals like (13) are counterexamples to (LAWS), even given the more restricted understanding of standard contexts.

[21] We will work through only one example, but we encourage readers to consult https://what-if.xkcd.com/ to get a feel for Munroe's many others. Some of our favorites include "What if a huge mountain. . . had the bottom inch of its base disappear?" (https://what-if.xkcd.com/57/), "What would happen if everyone on earth stood as close to each other as they could and jumped, everyone landing on the ground at the same instant?" (https://what-if.xkcd.com/8/), and "How much would the sea level fall if every ship were removed all at once from the Earth's waters?" (https://what-if.xkcd.com/33/).

First, note that (13) does not involve a backtracking or reverse-nomic interpretation: we're not asking after what kinds of pasts or laws would make its antecedent most plausible. Note as well that (13) passes all of the tests for suitability. That the counterfactual's antecedent is compatible with some initial segment of the past is clear. And the antecedent is indeed nomically possible: if, via some extremely low-chance event involving fine-tuned initial conditions, the particles in the pitcher's hand were accelerated to 0.9c mid-throw (perhaps incinerating the pitcher's arm and/or most of the ball in the process), we'd have a nomically possible witness to (13)'s antecedent.

But the fact that there are far-fetched nomically possible worlds that make both 'A pitcher throws a baseball at 0.9c' and 'The pitch causes a massive nuclear explosion' true does *not* mean that proponents of (LAWS) have a compelling story about where the true interpretations of (13) come from. The reason is that (13)'s truth seems to be determined in the way Munroe describes: we assume that by magic the ball gets going to 0.9c, and then from that point onward there are no further exceptions to the laws of physics. That is to say, even when we're interested in offering physically realistic answers to what-if questions, we treat possibilities in which small, isolated violations of the laws of physics (that is, miracles) get the ball going to 0.9c as more relevant to the assessment of the counterfactual than possibilities in which the ball accelerates to 0.9c by nomically respectable means. We thus contend that in the sorts of contexts evoked by Munroe's discussion, counterfactuals like (13) present straightforward counterexamples to (LAWS).[22]

4.2.2 Answering with Systematic Violations

Still, they are not yet counterexamples to (QUASI-LAWS), for the exceptions to the laws generally introduced by Munroe's cases are paradigmatic miracles. But this victory for proponents of the miracle view will be short-lived: reliance on small, isolated exceptions to the laws of nature is not essential to the interpretations of what-if questions. Indeed, it is easy to turn the natural answers to these questions into intuitive counterexamples to (QUASI-LAWS) as well. To do so, one need only invoke a context where the laws of "folk physics"—or even a crude "Looney Tunes"-like picture of the workings of the universe—are more prominent than anything like actual-world physics, at least for certain phenomena.

For example, it's easy to imagine contexts in which it would be natural to answer Munroe's 'What would happen if you tried to hit a baseball pitched at 90% the speed of light?' with something like the following:

[22] One who works in the tradition of Stalnaker (1968) and Lewis (1973) could also put the point in terms of *closeness*. For on these sorts of accounts, (LAWS) is naturally paired with the following principle: for any suitable counterfactual '$p \mathrel{\square\!\!\rightarrow} \ldots$' and any two p-worlds w and v, if w is nomically possible but v is not, then w is closer to actuality than v is. And our contention is that in the context of Munroe's discussion, the natural interpretations of (13) suggest that this principle is false: we intuitively treat nomically impossible worlds in which the ball gets going to 0.9c by magic as closer to actuality than certain nomic possibilities with extraordinarily fine-tuned initial conditions.

(14) If a pitcher were to throw a baseball at 0.9c, the batter would hit it provided they timed the swing perfectly, which of course would require starting the swing long before they saw the pitch leave the pitcher's hand.

(14) is a suitable counterfactual, for the same reasons as (13). And many of the contexts in which (14) is true will bring about counterexamples to both (LAWS) and (QUASI-LAWS): as we just saw, setting aside the question of how the baseball gets to 0.9c, actual-world physics implies that the pitch would cause a nuclear explosion, destroying the ball, the batter, and any spectators. The truth of (14) would thus require a more thoroughgoing violation of the laws of physics than is tolerated by even the miracle view. So certain natural answers to Munroe's what-if questions pose a problem for both the miracle and no-miracle views.

Similar points are brought out by counterfactuals like:

(15) I can throw a baseball 40 mph. So if I were to throw a baseball from New York City to Los Angeles, it would take about 60 hours to get there.

(16) If I were to lift the Empire State Building, I would drop it somewhere I'd never see it.

These counterfactuals are both suitable. Getting their antecedents to be true in a nomically possible manner is just a matter of there being the right kind of extraordinarily fine-tuned initial conditions, such as those that give rise to fortuitous and sustained gusts of wind that keep the baseball airborne from New York City to Los Angeles at a constant velocity. Such possibilities are extraordinarily improbable, but they are nomic possibilities all the same. And as was the case with (14), each of these counterfactuals can be used to express true propositions in the right contexts.

But these true propositions are not plausibly the kind that are compatible with either (LAWS) or (QUASI-LAWS), involving only fine-tuned initial conditions or small miracles. Rather, they seem to involve more thoroughgoing violations of the laws. In the case of (15), its natural interpretations involve worlds in which projectiles are affected neither by gravity nor by air resistance. Likewise, the natural interpretations of (16) require us to ignore the fact that the enormous lifting force would have to be applied across the surface area of two human hands, resulting in an amount of pressure that would in normal conditions cause the foundations of the building to collapse. No matter: we simply imagine a person lifting the building like Superman.

More generally, the contexts in which counterfactuals like (14)–(16) are most natural are not those in which we're wondering what sorts of laws would have been most plausible given such and such circumstances (generating reverse-nomic interpretations). Nor are they contexts like those involved in Munroe's miracle-friendly analysis of (13): it's not like we imagine some Lewisian miracle bringing about the antecedent and then running the actual laws as normal from there. Rather, they are contexts in which we simply ignore

the facts about the actual laws—or at least some non-trivial portion of them—and instead allow varying amounts of folk and Looney-Tunes physics to dictate the evolution of the world from the time of the ramp onward. Again, there's simply no general presumption that the true counterfactuals adhere as far as possible to the laws or even to the quasi-laws.

4.2.3 Answering Indeterminately

We'll close this section by revisiting one of the counterfactuals with which we opened the paper:

(1a) If Caesar had been in command in Korea, he would have used catapults.

There are contexts in which (1a) expresses a truth. Such contexts are plausibly standard, even granting all of the restrictions we have seen thus far: (1a) looks nothing like a backtracking or reverse-nomic counterfactual. And in any reasonable context (1a) will be suitable (see below). But what is the standing of (MACRO-PAST) and (QUASI-LAWS) in such contexts?

We don't have a decisive argument for the view that the natural interpretations of (1a) *must* involve violations of (MACRO-PAST) or (QUASI-LAWS). But we think these interpretations make the motivations for these principles (as well as their original, stronger counterparts) look rather tenuous. The reason is that it seems to us altogether *indeterminate* how much of the past and/or the laws are held fixed on the natural readings of these sorts of counterfactuals.[23] But if either (MACRO-PAST) or (QUASI-LAWS) were true, we wouldn't expect such levels of apparent indeterminacy.

For example, it is uncontroversial that there are nomically possible ways in which Caesar could have ended up in command during the Korean War. He could have been abducted by aliens and preserved via cryogenic freezing, before being redeposited on Earth in the mid-20th century. Or he could have been brought back by a kind of thermodynamic resurrection of the flesh: very improbably, his scattered particles might have re-formed spontaneously somewhere in the United States sometime before the Korean War, in just the internal configuration he was in shortly before his death. Likewise, it is clear that there are historically respectable ways (at least with respect to *some* portion of the actual past) in which Caesar could have ended up in command: history could have evolved exactly as it actually did up until shortly before the Korean War, at which point Caesar was somehow spontaneously reassembled and placed in command. Or maybe history could have been different all the way back to the time of Caesar's assassination (but not before), with Caesar somehow subsisting for another two millennia.

[23] Hedden (2023, pp.734-5) makes a similar observation about this famous example. So does Bennett (2003, §113), though he focuses on slightly different examples, while also disparaging them as "a marginal and uninteresting sort that I cheerfully relinquish to any philosopher who wants to spend time on them" (p. 255).

The important point is that the natural interpretations of (1a) are not invariably interpretations on which these sorts of claims hold. In asserting counterfactuals like (1a), we tend not to really care how the world got to be the way it is. We seem to simply stipulate that Caesar is there, and ask what he would have done subsequently.[24] Perhaps he arrived by nomically possible means, but in a manner that changed the course of much of human history; or perhaps he arrived by magic, leaving most of history unchanged. Who knows. Questions about what the past or the laws of such a world would have had to have been are likely both to lack a determinate answer and to be beside the point. As a result, the case for principles like (LAWS), (PAST), (MACRO-PAST) and (QUASI-LAWS) looks to us rather weak in the face of counterfactuals like (1a): if these principles were true, we'd expect it to be clear that the worlds that witness the truth of (1a) are exactly like actuality with respect to either their laws or their history (at least up until Caesar's birth); but in fact this isn't at all obvious.

Here is another way of seeing the worry, at least in the case of (PAST) and (MACRO-PAST). The standard argument for these principles rests on the fact that we so readily interpret counterfactuals like (17) as true:

(17) Had I gone to the gym this morning, Alexander the Great still would have died in 323 BC.

That is to say: the case for (PAST) and (MACRO-PAST) is just the observation that in many contexts we *do* make an effort to leave the past alone. Our point is that we make much less of an effort in many contexts naturally associated with certain kinds of what-if questions, like 'What would Caesar have done if he were in command in Korea?'. For example, we feel no analogous pressure to accept counterfactuals like:

(18) Had Caesar been in command in Korea, Alexander the Great still would have died in 323 BC.

But were (PAST) and (MACRO-PAST) the sort of platitudes their proponents take them to be, it is unclear why we wouldn't take (18) to express the sort of obvious truth (17) seems to.

Does this mean (PAST) and (MACRO-PAST) must be false? Not necessarily: one could insist that (18) is true on all its natural interpretations, and chalk up its apparent indeterminacy to an error on the part of philosophically unreflective speakers. The point is just that it's not clear that this move is dialectically legitimate, since our intuitive judgments about (18) suggest we *don't* always try to hold the (macro-)past fixed, even in cases where there is no in-principle barrier to us doing so.

We believe that similar points apply to one of Dorr's (2016a) central positive arguments for (LAWS). The argument starts as follows (p. 265):

[24] Cf. Kripke (1980), Plantinga (1974, ch.6), and van Inwagen (1985) on the so-called problem of "transworld identification".

> Suppose that L is a simple, true, deterministic law and that Frank, a philosopher of physics, has devoted his career to defending the truth of L. He is having a public debate with Nancy, who maintains (wrongly) that there are isolated exceptions to certain generalizations that follow from L, so that L is false. If we keep [(PAST)]... we have to say that if the circumstances of the debate had been different in any way whatsoever—for example, if someone had put a glass of water on Frank's lectern, or rudely interrupted his talk—then Nancy would have been right and Frank wrong. Thus [(F1)] and [(F2)] are true:

(F1) If we had given Frank a glass of water, his whole career would have been devoted to a mistake.

(F2) If you had told Frank that his whole career was devoted to a mistake, you would have been right.

Dorr claims that it is difficult to imagine either (F1) or (F2) being used to express a truth. We agree. Supposing this judgment is correct, it strongly suggests that (PAST) is false, since the principle entails that both (F1) and (F2) express context-invariant truths.[25] If we were to assume that at least one of (PAST) and (LAWS) must be true, then this argument would also provide strong evidence in favor of (LAWS). But of course we are *not* assuming this. So the question is whether our intuitive judgments about examples like (F1) and (F2) confirm (LAWS) in a way that goes beyond merely refuting (PAST).

Here we remain skeptical. For one, the proponent of (LAWS) remains on the hook for explaining what's going on with all of the law-breaking interpretations of counterfactuals we have surveyed so far. For another, it's easy enough to run an analogous argument against (LAWS). Consider, for instance:

(F3) If upon entering this room Nancy had pointed a wand at Frank and proclaimed 'Now is the perfect time to reveal to you that I'm a witch with magical powers—*abracadabra*!', at which point Frank had gone flying about the room in accordance with the motion of Nancy's wand, then Frank's whole career would have been devoted to a mistake.

Unlike (F1) and (F2), (F3) seems to express a truth. But (F3)'s antecedent is suitable, and so (F3) is a straightforward counterexample to (LAWS) (and plausibly also (QUASI-LAWS)).

Similarly, though it is not obvious that (F4) expresses a truth, it is not obvious that it expresses a falsehood either:

(F4) If Caesar were a participant in this debate, then Frank's whole career would have been devoted to a mistake.

[25] Notice also that an appeal to (QUASI-LAWS) would be of no help to the proponent of (PAST) here: Frank's beliefs concern the actual laws L themselves. Hence the argument for (LAWS) at the expense of (PAST).

But again, were (LAWS) true, we would expect (F4) to have only false interpretations, when in fact its truth-value seems indeterminate.

As in the case of (18) and (PAST), the claim is not that our intuitive judgments about counterfactuals like (F4) are *inconsistent* with (LAWS). But it does suggest that abductive arguments for (LAWS) grounded in intuitive judgments about counterfactuals—arguments like Dorr's—are dialectically fraught in a way that has not been properly appreciated.

4.3 Taking Stock

How should those engaged in the miracle versus no-miracle debate react to all of this? One option is to continue pursuing the strategy we've already seen: lump the various kinds of what-if interpretations with both backtracking and reverse-nomic interpretations, so that the contexts in which these interpretations arise also count as non-standard, and thus not under the purview of the debate.

We will discuss this reaction further in §6, but for now we'll note that it seems to us somewhat desperate. The kinds of what-if uses brought out by Munroe's examples or Quine's Caesar example strike us as perfectly ordinary interpretations of counterfactuals. If the notion of a "standard" context is being used to exclude not just backtracking and reverse-nomic interpretations, but also these sorts of interpretations, then it seems to us standard in name only. And setting aside the issue of what deserves the label 'standard', the more important question is whether there is any theoretical interest in principles that generalize only over such a circumscribed range of contexts. It is not at all clear to us that this could be so.

As a result, we think the various considerations we've seen up until now make a compelling case that there are no interesting general principles in the vicinity of (LAWS), (PAST), or their weakenings, and hence that the miracle versus no-miracle debate is at best of marginal theoretical interest, and at worst misleading as to the actual workings of counterfactuals. We'll soon draw some more general morals from this discussion. But first we will consider some objections.

5 Objections and Replies

5.1 First Objection: Elided Material

As we explained when introducing (LAWS) and (PAST), proponents of these alleged platitudes do not take them to be motivated by the *full* pattern of our ordinary use of counterfactuals. Instead, they are meant to be motivated by reflection on our usage of counterfactuals whose antecedents, in context, express *suitable* propositions—namely, propositions that are nomically possible and consistent with some initial segment of the actual past. In light of this, one might object that the various putative counterexamples we have pre-

sented against (LAWS), (PAST), and their weakenings violate this condition. Perhaps in each case the context is such that the relevant counterfactual's antecedent in fact expresses a nomically *impossible* proposition, rendering the true interpretations of these counterfactuals trivially consistent with the principles under consideration.

As we are understanding this objection, it claims that the "real" proposition expressed by the antecedent of, for instance, (15, 'I can throw a baseball 40mph. So if I were to throw a baseball from New York City to Los Angeles, it would take about 60 hours to get there') is something along the lines of: *that I throw a baseball from New York City to Los Angeles in a way that ignores the effects of gravity, drag, and the like*—at least in the standard contexts. Since this proposition is non-suitable, this counterfactual falls outside the scope of principles like (LAWS) and (PAST) by fiat.[26]

Reply: It is crucial to distinguish the question of what proposition is expressed by a counterfactual's antecedent from the question of what world (or set of worlds) is relevant to the assessment of the counterfactual's truth. We agree that in many standard contexts, a counterfactual expresses a truth because the world that witnesses the truth of its antecedent is a world in which the actual laws of physics (and perhaps actual history) are violated. Indeed, it is crucial to the success of our argument that this is true. But this does not imply that the proposition expressed by the counterfactual's antecedent *entails* such violations, for we know there are nomically possible worlds where, due to factors akin to pervasive thermodynamic luck, people just so happen to throw baseballs across continents.

Once this point is acknowledged, however, it is hard to see what might motivate this objection. We see no evidence that the antecedents of the counterfactuals at issue contain elided material. The mere fact that such a view would allow one to champion principles like (LAWS) and (PAST) is not a compelling reason to accept it.[27]

[26] For defenses of this reply, see Bennett (1984, pp.84-85), Lange (1993) and Lange (2009, pp.197-98).

[27] Lange (2009) offers a more sophisticated argument for the elided material diagnosis, which we quote in full (pp.197-198; the boldface is our emphasis):

> Suppose we say "$p \mathrel{\square\!\!\rightarrow} q$ because r" (e.g., "Had the syringe been filled with arsenic, then it would not have been a law that such a large dose of arsenic is lethal, since the patient lived") and then ask "Is it true that $p \mathrel{\square\!\!\rightarrow} r$?" **If the first counterfactual's antecedent p tacitly included r, then this question has the effect of stripping r out of p in any of p's appearances as an antecedent in the near future.** (If r remains implicit in p, then the question "Is it true that $p \mathrel{\square\!\!\rightarrow} r$?" is trivial, and so to take p as implicitly including r is to give an uncharitable reading of the question.) **Therefore, if the answer to the question "Is it true that $p \mathrel{\square\!\!\rightarrow} r$?" is "No," then r must have been preserved under the original counterfactual's antecedent only by virtue of its having been implicit in that antecedent.**

Though we find this proposal intriguing, we think it faces some significant challenges. For starters, it's worth emphasizing that the second boldface claim—the one that actually licenses the elided material diagnosis—doesn't follow from the first. It could easily be that (i) whenever r is implicit in the antecedent of '$p \mathrel{\square\!\!\rightarrow} q$ because r', the answer to the question "Is it true that $p \mathrel{\square\!\!\rightarrow} r$?" is "No", without it being the case that (ii) whenever the answer to the question "Is it true that $p \mathrel{\square\!\!\rightarrow} r$?" is "No", r must be implicit in the antecedent of '$p \mathrel{\square\!\!\rightarrow} q$ because r'. And in fact we think it is not so difficult to find counterexamples to this second claim. Consider, e.g.:

Worse, the claim that the propositions expressed by the relevant antecedents entail violations of the laws quickly leads to some rather implausible consequences. In particular, if it were correct, it would seem to follow that in the contexts in which (15) is natural, the following counterfactual would express a proposition that could be known *a priori*:

(15*) If I were to throw a baseball from New York City to Los Angeles, and if it were to take about 60 hours to get there, gravity and air resistance would have no effect on the ball's trajectory.

But on all ways we can interpret this counterfactual, it is as *a posteriori* as can be.

5.2 Second Objection: Expanding the Notion of "Suitability"

We do not see a plausible path toward the conclusion that reverse-nomics and (the relevant kinds of) what-ifs require interpretations on which their antecedents are non-suitable, given our definition of 'suitability'. But perhaps the issue is simply with how we've defined 'suitability'.

Our reason for restricting attention to counterfactuals whose antecedents express propositions that are consistent with the laws and some segment of actual history was to make the case for (LAWS) and (PAST) not entirely hopeless. But maybe the restrictions need to be more substantial than that. For notice that the antecedents of many of the counterfactuals discussed above express propositions concerning states of affairs that are rather abnormal, to put it mildly: we're asking what would happen if Jesus were to walk on water, a pitcher were to throw a baseball at 90% the speed of light, and so on. We saw that these antecedents are all nomically possible, but one might object that they are nevertheless sufficiently exceptional that they ought to fall outside the scope of the miracle

(A) Had the assassin aimed a bit more to the right, his target would have been killed, since the bullet missed just barely to the left.

It is clear that there could be situations in which (A) is true. But it is also clear that in those situations, the correct answer to "Had the assassin aimed a bit more to the right, would the bullet still have missed just barely to the left?" will be "No". By Lange's test, it follows that the proposition that the bullet missed just barely to the left is implicit in the antecedent of (A). But this predicts that (A) is semantically equivalent to the absurd:

(A*) Had the assassin aimed a bit more to the right and the bullet had missed just barely to the left, his target would have been killed.

We thus have doubts about the general correctness of Lange's diagnostic. However, even setting those doubts aside, we're also skeptical that the diagnostic properly handles the full range of counterfactuals we've considered. Perhaps it properly handles (15)—at least if we assume that the relevant value of r in (the merely implicit?) '…because r' clause is 'gravity and air resistance would be absent', or something to that effect. But which nomically impossible proposition is supposed to be implicit in the antecedent of (e.g.) (16, 'If I were to lift the Empire State Building, I would drop it somewhere I'd never see it')—is it that the speaker lifts the Empire State Building by magic? To our ears, the natural answer to the question "If you were to lift the Empire State Building, would it be by magic?" could well be "Yes!".

versus no-miracle debate. Indeed, the kinds of intuitive judgments that are taken to motivate principles like (LAWS) and (PAST) typically target counterfactuals whose antecedents concern relatively normal states of affairs that are describable in macroscopic terms, and whose truth isn't hostage to extremely fine-tuned initial conditions.

So perhaps the platitude underlying the disjunction of (LAWS) and (PAST) is something more like: whenever we're in a standard context and we're dealing with a counterfactual whose antecedent expresses a proposition that is (i) consistent with the actual laws, (ii) consistent with some amount of the actual past, and (iii) *sufficiently normal*, then either we hold fixed the laws or we hold fixed the past. Supposing our recalcitrant examples have antecedents that express insufficiently normal propositions, then perhaps the miracle versus no-miracle debate can be salvaged.

Reply: Without a reasonably precise statement of the conditions under which the truth of a proposition counts as 'sufficiently normal', it's hard to assess a proposal of this sort. But we'll raise some worries that we anticipate arising for any plausible implementation of it.

First, we're skeptical this style of response is capable of doing the work it needs to do. The issue is that backtracking counterfactuals like (8b, 'If I had jumped out of this tenth-floor window, I would have done so only because someone had put a safety net in place') and reverse-nomic counterfactuals like (12, 'Had Fizeau recorded the same results but with a slightly faster-spinning cog than in his actual experiment, the speed of light would have been slightly faster than it actually is') have antecedents that express propositions about familiar properties of macroscopic objects, and that are true at perfectly normal possible worlds, with no need for fine-tuned initial conditions. So even if all of the what-ifs could be ruled out (and it's not at all clear that this would follow), there's still the question of what to make of backtracking and reverse-nomic uses.

Perhaps the objector can divide and conquer. They could employ the distinction between standard and non-standard contexts to account for the problematic counterfactuals with relatively normal antecedents, and suitability to account for the rest. But this leads us to our second reply, which is that it becomes especially difficult to see the theoretical interest of principles like (LAWS) and (PAST) when the uses of counterfactuals they are meant to cover have been so aggressively circumscribed. Again, without an independently plausible account of what kinds of antecedents are meant to be excluded by the expanded notion of 'suitability' (and likewise 'standard context'—though more on this in §6), this strategy presents a serious risk of trivializing (LAWS) and/or (PAST), insisting that the principles hold provided we set aside cases where they fail.

Third, supposing proponents of the expanded notion of suitability venture to give an account of the relevant notion of 'sufficiently normal', they risk struggling to explain intuitive judgments about counterfactuals that their preferred principles would otherwise be well-suited to explain. For example, any account of normality that excludes fine-tuned

initial conditions will plausibly classify counterfactuals like (19) as non-suitable:

(19) If all of the air molecules in this room were to collect near the ceiling for the next ten minutes, I'd suffocate.

Likewise for (20):

(20) If yesterday I had flipped a fair coin heads 1,000 times in a row, it would have been the first time in human history that anyone had done so.

And although strange, both (19) and (20) seem to be instances of the patterns that principles like (LAWS) and (PAST) are meant to explain. But the proponent of the expanded notion of suitability will not be in a position to account for this fact. This is because, on their view, (LAWS) and (PAST) have simply nothing to say about our intuitive reactions to counterfactuals like (19) and (20), since their antecedents are deemed non-suitable.

We thus think that the participants of the miracle versus no-miracle debate would do best to leave the notion of 'suitability' in its original, austere form, and to try to answer our challenge in some other way.

5.3 Third Objection: Error-Theory

The next response we will consider concedes both that the counterfactuals at issue have the contents they seem to have—at least in the relevant contexts—and that *if* these counterfactuals are true, they are genuine counterexamples to one or both of (LAWS) and (PAST). What it rejects is the supposition that the counterfactuals in question are actually *true*, contrary to our initial judgments.

Perhaps there is simply no context in which we speak truly in saying (8b, 'If I had jumped out of this tenth-floor window, I would have done so only because someone had put a safety net in place'); after all, there is in fact no net, so had I jumped I would have died—end of story. Likewise, perhaps in no context do we speak truly in saying (11, 'Were Michelson and Morley to have measured a significant difference in the speed of light traveling in Earth's direction through the presumed luminiferous aether versus light traveling orthogonal to that direction, it would have been because there really was a luminiferous aether'); after all, there is in fact no luminiferous aether, so had they gotten those measurements, it would have been because of experimental error. And perhaps the same for (16, 'If I were to lift the Empire State Building, I would drop it somewhere I'd never see it'); after all, if I were to lift the Empire State Building, it would immediately collapse due to the enormous pressure exerted by my hands.

Proponents of this objection needn't accept all of these claims. Once again they could selectively divide and conquer, invoking standard contexts to handle certain problematic counterfactuals, and then error-theories for whatever else is left over. Perhaps backtracking counterfactuals and reverse-nomics are true in some contexts but always non-standard,

whereas the (relevant) interpretations of what-ifs occur in standard contexts but are always false. However the details are spelled out, the idea will again be to strike a balance between overburdening the notion of a standard context and predicting too much error in ordinary counterfactual judgments.

Reply: We are deeply skeptical of error-theoretic treatments of these counterfactual judgments. Even supposing the error-theory is applied only to certain what-if uses of counterfactuals—with backtracking and reverse-nomic uses being relegated to so-called non-standard contexts—we doubt that there are any general principles that can be cited in its favor. This is because we take it to be common ground between us and the proponents of the miracle versus no-miracle debate that counterfactuals are context-sensitive, and that a single counterfactual can be used to express both truths and falsehoods. It thus becomes mysterious why counterfactuals like

(15) I can throw a baseball 40mph. So if I were to throw a baseball from New York City to Los Angeles, it would take about 60 hours to get there.

(16) If I were to lift the Empire State Building, I would drop it somewhere I'd never see it.

must invariably express falsehoods, given that they *seem* to express truths, at least on some uses. Surely there could be an alien community that speaks a version of English exactly resembling ours, except where the meta-semantic facts relevant to the assessment of counterfactuals do not context-invariantly privilege worlds in which the laws and/or past are held fixed over those in which they are not. Given that we seem to speak exactly the way such a community would, why think our counterfactual isn't just theirs?

The error-theorist might reply that there *is* direct evidence that we are not like these imagined aliens. This is because there is an easy way to make counterfactuals like (15)–(16) seem false: namely, by making the workings of the actual laws of nature conversationally salient. For example, in reply to an assertion of (15), it would be wholly natural to point out that there are such things as gravity and air resistance, and thus that, whatever else would happen were you to throw the ball from New York City to Los Angeles, it almost certainly would not arrive in Los Angeles 60 hours later. Moreover, once this has been pointed out, it does not seem reasonable for the original speaker to stand their ground and reassert (15). At best the speaker will have to assert a new counterfactual that explicitly builds in a violation of the laws or a series of thermodynamic miracles into the antecedent itself to get the same point across.

One might take this to be evidence that the relevant what-if counterfactuals do not express truths, even in the most favorable contexts. For if they did express truths, then shouldn't competent speakers be able to stand their ground after asserting them? And wouldn't a response like 'The ball would stop moving long before it reached Los Angeles on account of gravity and air resistance' simply be misunderstanding the original asser-

tion? More plausibly, the objection goes, speakers who assert counterfactuals like (15) are sloppily failing to recognize how things would actually evolve in accordance with the laws of nature. This is what the error-theoretic response corrects.

We don't think the judgments concerning backtracking and reverse-nomic interpretations of counterfactuals pattern similarly, so one who motivates the error-theory in this way will probably appeal to standard contexts to set aside those uses. But the important point is that this version of the error-theory presents a genuine challenge to the relevant what-if interpretations outlined in §4.2.

Nevertheless, we take this more sophisticated error-theoretic diagnosis to fail. The mere fact that a certain interpretation of the counterfactual becomes difficult (or even impossible) to access after certain other propositions have been made salient is not good evidence that that interpretation was never there in the first place. This is because numerous uncontroversially context-sensitive expressions exhibit exactly this phenomenon. Take quantifier domain restriction. It is widely accepted that one can speak truly in uttering a sentence like 'Everyone is at the meeting' when one knows that all and only everyone *who is a member of the department* is at the meeting. An annoying person might respond to this utterance by pointing out that Obama isn't at the meeting, and thus that not everyone is at the meeting. And in the face of this "correction", it would not be felicitous to simply reassert 'Yes, but everyone is at the meeting'. One will have to make explicit that by 'everyone', one meant 'everyone who is a member of the department' or what have you. But this should *not* be taken to show that one did not speak both literally and truly in uttering 'Everyone is at the meeting'. Rather, it shows that it's possible to shift the context over the course of a conversation, and that certain kinds of shifts are harder to undo than others.[28]

The error-theorist might resist this analogy as follows.[29] Yes, by performing certain speech acts the annoying interlocutor can expand the domain of quantification to include Obama, thereby making subsequent assertions of 'Everyone is at the meeting' infelicitous. However, (i) one can still respond with an utterance like 'Hey, you know perfectly well what I meant!', and (ii) there is no temptation to say that one's original utterance ('Everyone is at the meeting') was somehow mistaken. Yet neither (i) nor (ii) seems to hold in cases of putative context-shifting with counterfactuals. If someone were to react to your utterance of (15) with a speech like:

> No, if you were to throw a baseball from New York City to Los Angeles, gravity and air resistance would stop it in its tracks long before it got there.

—it would seem unreasonable for you to reply 'Hey, you know perfectly well what I meant!', and there would seem to be real pressure for you to acknowledge that your original utterance was mistaken.

[28] Similar points arise for other context-sensitive expressions. For relevant discussion see Dorr (2014) and Schaffer & Szabó (2014).

[29] We thank an anonymous referee for encouraging us to discuss this further step in the dialectic.

We agree with the error-theorist's judgment that quantifiers and counterfactuals are dissimilar in this respect. But we suspect that the best explanation of this asymmetry is not that 'Everyone is at the meeting' has a true interpretation that can be made difficult to access, whereas (15) only has false interpretations. Instead, our view is that the best explanation of this asymmetry is simply that the contextual dynamics of counterfactuals are opaque to (otherwise) competent speakers in a way that the contextual dynamics of quantifiers are not.

Our reason for thinking this is broadly abductive: counterfactuals like (15) are not unique in failing the 'Hey, you know perfectly well what I meant!' or 'I guess I was wrong' tests; in fact, large swaths of "standard" counterfactuals fail them too. If failing these tests were enough to justify an error-theory, then we'd be threatened with the prospect of a very wide-ranging—and thus implausible—counterfactual skepticism.[30] So better just to reject the test.

Here is an example in illustration.[31] Suppose Sly Pete and Mr. Stone are playing poker. It's the final round of betting, and Mr. Stone has decided to go "all-in", wagering the remainder of his chips. The only hand Mr. Stone can lose to is two aces. As it happens, Sly Pete has a weak hand, which he folds. Sly Pete then declares:

(SP1) Had I had two aces, you would have lost all your money.

To our ears, (SP1) is a paradigm case of a true "standard" counterfactual—nothing as exotic as what's involved with backtracking counterfactuals, reverse-nomics, or what-ifs is present here. However, despite the apparent reasonability of Sly Pete's assertion of (SP1), we think Mr. Stone could easily reply:

(SP2) That's false. Had you had two aces, you would have played your hand in a way that would have led me to fold much earlier, and so I would have saved most of my money.

To our ears, (SP2) is a paradigm case of a true backtracking counterfactual: if Sly Pete had had two aces, he would have played the hand differently, and Stone would have folded long before going "all-in". So Mr. Stone's assertion likewise looks perfectly reasonable.

Here is the issue. Once (SP2) has been asserted, it becomes very difficult to hear the true readings of (SP1). Sly Pete cannot simply stand his ground and reassert it, or exclaim 'Hey, you know perfectly well what I meant!', or flat-footedly insist that he didn't say anything false. So is this evidence that we are in error in judging that (SP1) expresses a truth in its original context? We don't think so—otherwise we'd have reason to think that

[30] Of course, there are those like Hájek (2014, 2021) who are comfortable with wide-ranging counterfactual skepticism. We won't engage with such views directly here, though we'll note that the path to skepticism we're discussing is not the same as Hájek's, and indeed plausibly has an even broader scope.

[31] See also the discussion of "reverse Sobel sequences" in von Fintel (2001), Gillies (2007), Moss (2012), and Lewis (2018) for further examples of these dynamics.

huge swaths of ordinary counterfactuals fail to express truths, since it is not at all difficult to find situations that lend themselves to the (SP1)/(SP2) dynamics. Better, we believe, is to take all of this to be further evidence of the context-sensitivity of counterfactuals. The mere fact that the felicity of (SP2) seems to "trump" the felicity of (SP1)—in a way that seemingly blocks recourse to the 'Hey, you know what I meant!'-type defense—should not be taken to show that (SP1) must be context-invariantly false. Instead, it should be taken to show that the contextual dynamics of counterfactuals are often opaque, even to otherwise competent speakers.

6 Retreat to Standard Contexts

We now turn to a final reply we'll consider on behalf of the proponent of the miracle versus no-miracle debate. This reply insists that, contrary to what we claimed in §4, there *is* a way of understanding 'standard context' such that (i), the notion guarantees that the backtracking, reverse-nomic, and various what-if interpretations of the counterfactual surveyed above do not occur in standard contexts; and (ii), the resulting class of uses allows for a theoretically interesting miracle versus no-miracle debate.[32] We'll consider two ways of implementing this idea. On the first, 'standard context' is defined in terms of judgments about *rational decision-making*.[33] On the second, 'standard context' is defined by ostension to the puzzle of counterfactuals under determinism itself. We find neither view capable of delivering both (i) and (ii).

6.1 Standard by Decision

Under the influence of Stalnaker (1981) and Gibbard & Harper (1978), many philosophers have come to endorse some version of so-called "causal decision theory". Though there is significant variation in how exactly the view is spelled out, the core idea is that rational decision-making is a matter of maximizing expected *counterfactual* utility: one ought to choose the action that, in expectation, would make things go best were one to choose it.[34]

Setting aside certain recherché cases (e.g., Newcomb's problem) that distinguish causal decision theory from its main rival, evidential decision theory, the core idea is simple: very

[32] Note that (ii) rules out any definition of 'standard context' that settles the miracle versus no-miracle debate by fiat. So, for example, it won't do to characterize standard contexts as "scientific contexts"—contexts where we strictly adhere to the laws of nature in our counterfactual reasoning—since this would make the correctness of the no-miracle view a mere triviality.

[33] The second of Dorr's (2016a, §7) two major arguments against (PAST) and for (LAWS) is centered on such judgments. Much of what we say in §6.1 closely follows Dorr's discussion, except we focus only on the connections between counterfactuals and decision-making, whereas Dorr also discusses the connections between counterfactuals and emotions like *regret*. But the dialectic in both cases is quite similar, and we see no reason to think the concerns we're about to raise for a decision-theoretic gloss on 'standard context' would work any less well on a regret-theoretic gloss.

[34] The literature on causal decision theory is enormous. For a small sampling, see Lewis (1981), Skyrms (1982), Sobel (1994), Joyce (1999), Ahmed (2013, 2014), Bacon (2022), Elga (2022), Gallow (2023, 2024), Hedden (2023), and McNamara (2024).

often the best way to justify a decision is to explain what would have happened had one made a different decision. How did I know that it was a good idea not to jump out of the window? Because I knew that if I were to jump out the window, I'd die. How do I know that I should buy insurance? Because I know that it's sufficiently likely that if I were not to buy insurance, I'd be financially ruined. More generally: if I know I made a good decision in choosing to ϕ, then plausibly I know the counterfactual fact that things would have been worse had I chosen not to ϕ.

However, counterfactuals are context-sensitive. The counterfactual 'If I had chosen not to ϕ, things would have been worse' might in one context express a proposition I know to be true, but in another context express a proposition I know to be false. So if facts about rational decision-making are closely connected to facts about counterfactuals, then plausibly claims about rational decision-making inherit some of the context-sensitivity of the counterfactual.

What's less clear is whether they inherit the *full* extent of it. Suppose I'm looking out the window of my tenth-floor office, considering whether to jump. I want most of all not to die. But supposing survival is guaranteed, I much prefer the excitement of jumping safely out of a tall building into a net to the boredom of sitting in an office. Still, I see that there is no net to catch me, so I decide not to jump.

Given my preferences, it is obvious that I have made the correct decision. Indeed, there does not appear to be any sense in which I could have spoken truly in asserting:

(21) I should jump out of this tenth-floor window.

And this is despite the fact that it's relatively easy to interpret (22) as expressing a truth (as when in a backtracking context):

(22) I'm certain that if I were to jump out of this tenth-floor window, it would only be because someone had put a safety net in place.

But this is puzzling. Given my stated preferences, if both 'I'm certain that if I were to jump from this window, I'd fall into a net' and 'I'm certain that if I were not to jump from this window, I'd sit in my office' are true, then 'I'm certain I'd be better off jumping from this window' must be true too. And this would imply that in any context in which (22) is true, (21) is true. But there seem to be no true interpretations of (21), despite the fact that there do seem to be true interpretations of (22).

These considerations suggest that when questions of rational decision-making are salient, certain interpretations of the counterfactual are unavailable. And perhaps this is enough to reverse-engineer a workable notion of a "standard context": the standard contexts are just those contexts in which the natural interpretations of claims about rational decision-making are salient. Given this way of understanding 'standard context', it follows that if the only contexts in which a certain counterfactual has a true interpretation are contexts

that would make bizarre claims about decision-making express truths, then that counterfactual is false in any standard context.

We are happy to accept that this notion of a standard context is theoretically interesting. We are also happy to accept that the intuitive force of the puzzle of counterfactuals under determinism survives this restriction, since it is uncontroversial that our knowledge of the past and the laws plays a significant role in decision-making. The issue, however, is that we don't think this conception of a standard context does enough to rescue the miracle versus no-miracle debate, since both (LAWS) and (PAST) remain false, even when restricted in this way.

Dorr (2016a, §7) makes a compelling argument against (PAST) in this setting.[35] Recall Frank, the philosopher of physics who has devoted his career to defending the truth of L. Suppose Frank cares most of all that he has not devoted his career to defending the truth of a false physical theory. If (PAST) is true (and thus (LAWS) is false), then in any standard context, the inference from 'Frank ϕs' to 'Frank doesn't $\phi \mathrel{\square\!\!\rightarrow} \neg L$' is valid. And this in turn implies that for any decision Frank faces, he would be best off performing the action he is most confident he will perform, no matter how silly or otherwise inadvisable that action might initially seem.[36] After all, if Frank is extremely confident that (PAST) and L are both true and that tomorrow he'll impulsively spend all his money on lottery tickets, then he should be extremely confident that if tomorrow he were to do something other than spend all his money on lottery tickets, L would be false.[37] So in expectation Frank would be better off buying the lottery tickets than not, for the simple reason that this is what he antecedently expects himself to do. But this is absurd. So it must be that we sometimes hold the entirety of the laws fixed in reasoning about the counterfactual consequences of our possible decisions. Thus (PAST) is false, even when restricted to standard (i.e., decision-theoretic) contexts.

We find Dorr's argument against (PAST) persuasive. Once again, however, (PAST) being false is one thing, (LAWS) being true is another. And we think comparably compelling arguments can be run against (LAWS).

Suppose Lavinia is convinced of the quasi-Leibnizian view that the actual history of the world is the best of all possible histories. As a result, for any past time t, Lavinia cares more about the truth of H_t than any other relevant proposition. If (LAWS) is true, then the inference from 'Lavinia ϕs' to 'Lavinia doesn't $\phi \mathrel{\square\!\!\rightarrow} \neg H_t$' is valid. And this implies that Lavinia is in essentially the same predicament as Frank: for any decision she faces, she would be best off performing whichever action she is most confident she will take,

[35] See Ahmed (2013, 2014), Hedden (2023), Gallow (2023), and McNamara (2024) for related discussion.

[36] Here we are anchoring ourselves into an arbitrary standard context, so that for readability's sake we may use rather than mention the relevant counterfactuals.

[37] This argument assumes that if (PAST) is true, then we can be rationally certain that (PAST) is true. We don't see any strong reasons to be suspicious of this assumption, but also note that it's not really essential. So long as we can assign (PAST) high probability we can generate perverse verdicts about rational decision-making. (*Mutatis mutandis* for (LAWS) below.)

no matter how silly or inadvisable that action might initially seem. After all, if Lavinia is extremely confident that (LAWS) is true and that tomorrow she'll impulsively spend all her money on lottery tickets, then she should be extremely confident that if tomorrow she were to do something other than spend all her money on lottery tickets, then, for any past time t, H_t would be false. So in expectation Lavinia would be better off buying the lottery tickets than not, for the simple reason that this is what she antecedently expects herself to do. But this is absurd. So it must be that we sometimes hold the past fixed in reasoning about the counterfactual consequences of our possible decisions. Thus (LAWS) is false, even when restricted to decision-theoretic contexts.

Of course, it is possible to insist that if Lavinia really does care more about preserving the actual history of the world than anything else, then since causal decision theory is true and we live in a deterministic world, she really should make decisions on the basis of her antecedent expectations of how likely she is to make that very decision. But we do not think this move is dialectically appropriate, since pre-theoretically it is very hard to accept that one who has Lavinia's (strange, but seemingly not *irrational*) preferences ought to make decisions in this way (*mutatis mutandis* for Frank). As we see it, one who is not antecedently committed to (LAWS) would be better off either: (i), taking Lavinia's case to show that sometimes we hold the past fixed when reasoning about the counterfactual consequences of our possible decisions (and thus that (LAWS) is false); or (ii), taking Lavinia's case to show that judgments about ordinary counterfactuals are at best an approximate guide to the relevant decision-theoretic truths.[38]

We conclude that the notion of a standard context delivered by considerations of rational decision-making is not one that can generate a reasonable miracle versus no-miracle debate. Even when we restrict our attention to contexts in which questions of decision-making are salient, counterfactuals are too context-sensitive to obey principles like (LAWS) and (PAST).

6.2 Standard by Ostension to the Puzzle

We now consider one final proposal on behalf of the miracle versus no-miracle debate.[39] This proposal concedes a substantial portion of what we've been arguing for so far. In particular, it concedes that (LAWS), (PAST), (QUASI-LAWS), and (MACRO-PAST) are all false due to being universal generalizations about the truth-conditions of (suitable) counterfactuals across a wide range of contexts. What the proposal does not concede, however, is that this shows that the miracle versus no-miracle debate is defective. According to the proposal, what's puzzling about determinism is not that it implies that we cannot *always* (i.e., context-invariantly) hold both the laws and the past fixed when assessing counterfactuals. What's puzzling about it is that it implies that we can *never* (i.e., in no context) hold

[38] For more on this latter suggestion, see Gallow (2023).
[39] We thank an anonymous referee for raising the objection that led to this subsection.

both the laws and the past fixed when assessing counterfactuals. For it seems clear that there are at least some ordinary contexts in which we are antecedently committed both to the counterfactual robustness of the laws and to the counterfactual robustness of the past. And it is a consequence of determinism that, even in these contexts, we must abandon at least one of these two commitments.[40]

Building on this observation, the proponent of the miracle versus no-miracle debate could try to define 'standard context' by *ostension*. That is, they could simply define a 'standard context' as any context in which the puzzle of counterfactuals under determinism actually seems like a puzzle. This alone is plausibly enough to rule out many of the uses of the counterfactual we surveyed in §4, since there is little temptation to treat the laws and/or the past as counterfactually robust in contexts where backtracking, reverse-nomic, and certain what-if readings are salient. Indeed, it is plausibly a necessary condition on c being a standard context (so defined) that c be a context in which both the quasi-laws and the macroscopic past are held fixed, and in which *either* the actual laws (without isolated miracles) or the actual past (in full micro-physical detail) is held fixed. Combining these ideas, we get the following ostensive characterization of a standard context:

> c is a *standard context* iff: if 'a' expresses a false suitable proposition in c, then: (i) '$a \boxcircle\!\!\rightarrow (QL \wedge MH_t)$' is true in c; (ii) both '$a \boxcircle\!\!\rightarrow L$' and '$a \boxcircle\!\!\rightarrow H_t$' *seem* true in c; and (iii) exactly one of '$a \boxcircle\!\!\rightarrow L$' or '$a \boxcircle\!\!\rightarrow H_t$' *is* true in c.

This conception of 'standard context' guarantees the conjunction of (QUASI-LAWS) and (MACRO-PAST), as well as that, in every standard context, either the laws or the past is held fixed. It also avoids trivializing the miracle versus no-miracle debate, since neither view is such that it is guaranteed to be true by fiat.

So: do we now have a way of understanding (LAWS) and (PAST) whereby the question of which is true avoids having a false presupposition?

Not necessarily. This is because all that's guaranteed by this definition of 'standard context' is that for every standard context c, either c fixes the actual laws or c fixes the actual past. But it could well be that the standard contexts are non-uniform with respect to which of these two things is kept fixed. Perhaps c_L is a standard context that fixes the actual laws (ensuring that '$a \boxcircle\!\!\rightarrow L$' is true in c_L), while c_H is a standard context that fixes the actual past (ensuring that '$a \boxcircle\!\!\rightarrow H_t$' is true in c_H). And if that were the case, then the answer to the question:

> When one wants but cannot have both, which should be given up: the counterfactual robustness of the laws or the counterfactual robustness of the past?

would simply be: *it depends on the context*.

[40] Here, as in our initial presentation of the puzzle in §2, we gloss over some details required to generate a formal contradiction for ease of exposition. Doing so does not affect any of the points we wish to make.

An example will help make the point concrete. It is easy to imagine a context c such that, in c, we are inclined to accept both of the following counterfactuals:

(BP) If I had blinked once more a second ago, the distant past would have been exactly as it actually is.

(BL) If I had blinked once more a second ago, the laws of nature would have been exactly as they actually are.

We can assume, harmlessly, that c is a standard context (in the present, ostensive sense). So at least one of (BP) and (BL) is true in c. But determinism guarantees that they aren't both true in c. So one of them must be given up. Which should it be? Proponents of the miracle view say to keep (BP) and reject (BL), while proponents of the no-miracle view say the opposite. Suppose, without loss of generality, that the no-miracle view is right *as far as* c *goes*: in c, both (BP) and (BL) seem true, but only (BL) is actually true. Would this vindicate the no-miracle view? Not necessarily! For it could easily be that there is another standard context c', in which both (BP) and (BL) seem true, but only (BP) actually is true. c' would then be a context that vindicates the miracle view. And if that were the case—if there really were a pair of standard contexts like c and c'—then it seems that all we could say about the miracle versus no-miracle debate is that the miracle view is right as far as some standard contexts go (like c'), while the no-miracle view is right as far as other standard contexts go (like c). But then what would be left to debate?

Thus, if the ostensive conception of 'standard context' is going to deliver a workable miracle versus no-miracle debate, it needs to be that the standard contexts are *uniform* with respect to which of the laws or the past is held fixed in them—i.e.:

> Either (i) every standard context c is such that if 'a' expresses a false suitable proposition in c, then '$a \boxrightarrow L$' is true in c; or (ii) every standard context c is such that if 'a' expresses a false suitable proposition in c, then '$a \boxrightarrow H_t$' is true in c.

For again: if this "uniformity assumption" is false, then there is nothing more general to be said about the puzzle of counterfactuals under determinism other than that in some contexts the miracle view is right while in others the no-miracle view is.

The question, then, is whether proponents of the miracle versus no-miracle debate have grounds for endorsing the uniformity assumption. Do we have any reason to think that there is some general feature of the contexts in which the puzzle of counterfactuals under determinism gets its grip that explains why, in those contexts, it's the laws rather than the past that's held fixed (or vice versa)? Or are such contexts too variegated to admit simple generalizations of this form?

Our sense is that most participants in the miracle versus no-miracle debate have simply taken the uniformity assumption for granted. However, to the extent that we can recover

arguments for it, they mostly take the form of observations to the effect that there are many (ordinary) contexts in which we are inclined to treat counterfactuals like (BP) and (BL) as true, as well as arguments (like Dorr's) to the effect that if certain of these counterfactuals are false, then a number of surprising conclusions follow (in context). The issue is that these sorts of arguments tend to cut both ways (as we've argued in §4.2.3 and §6.1), and in any event are a far cry from establishing a conclusion as strong as the uniformity assumption.

Unsurprisingly, we believe that the uniformity assumption is false, and thus that the miracle versus no-miracle debate is defective even on its present, ostensive construal. We'll explain why with two analogies.

First, consider a standard Sorites argument:

(V1) A glass that is 100% full is pretty full.

(V2) If a glass that is 100% full is pretty full, then a glass that is 99% full is pretty full.

(V3) If a glass that is 99% full is pretty full, then a glass that is 98% full is pretty full.

...

(V101) If a glass that is 1% full is pretty full, then a glass that is 0% full is pretty full.

(VC) ∴ A glass that is 0% full is pretty full.

We believe each of this argument's 101 premises has considerable pre-theoretic appeal. But its conclusion is obviously false. The argument is valid, assuming modus ponens is.[41] So at least one of the premises must be rejected. Which is it?

We're not sure. We think that in just about any context in which all the premises seem true, it will be indeterminate which premise is in fact false, and, moreover, that these situations are likely to be non-uniform with respect to the culprit.[42] However, we also think that

[41] This is perhaps a tendentious assumption in the theory of vagueness, but innocuous in the present dialectical setting.

[42] Certain kinds of contextualists about vagueness (e.g., Raffman 1994) may be inclined to resist the claim that there is any single context (in the theoretical, *parameter of interpretation* sense of 'context') in which each of (V1)-(V101) seems true. According to the kind of contextualist we have in mind, what explains the impression that each of (V1)-(V101) seems true is that: (i) each premise in the argument can be associated with an interpretation (i.e., resolution of the context-sensitivity) of 'pretty full' on which that premise is clearly true; and (ii) moving through the argument (premise by premise) has the effect of continually shifting the context so as to make whichever premise happens to be under consideration clearly true relative to that particular context (despite the fact that that same premise may be clearly false relative to some of the other interpretations of 'pretty full' that are operative over the course of one's assessment of the argument).

For our purposes we need not take a stand on the plausibility of this approach to the Sorites. But we'll note that a similar strategy is available for the puzzle about counterfactuals under determinism. One could hold that there is no single context (in the parametric sense) in which both (BP) and (BL) seem true; rather, there are merely *situations*—i.e., spatiotemporally extended events—in which both (BP) and (BL) seem true, and these situations invariably involve a non-uniform resolution of the context-sensitivity of the counterfactual vocabulary in (BP) and (BL). This would be a view on which the puzzle of counterfactuals under

there is (at least) one premise that clearly ought *not* be rejected: (V1). And this is because this premise is determinately true. And the reason (V1) has this status is that it's a general feature of our use of the word 'pretty full'—one that manifests itself in every context—that the sentence 'A glass that is 100% full is pretty full' expresses a truth.

Now if one could find evidence that all but one of the premises of the Sorites argument had this status, then the question "Which premise in the argument from (V1)-(V101) to (VC) ought to be rejected?" would more clearly be worth trying to answer. For this would be tantamount to justifying the analogue of the uniformity assumption for 'pretty full': it would be a way of showing that either (i) for every context c, if all the premises in the Sorites seem true in c, then (V1) is false in c, or (ii) for every context c, if all the premises in the Sorites seem true in c, then (V2) is false in c ...or (ci) for every context c, if all the premises in the Sorites seem true in c, then (V101) is false in c. But we have strong reason to think that this assumption is false, since we have strong reason to think that, of the various ways of using the expression 'pretty full', some will make (e.g.) 'A glass is pretty full if and only if it is at least 70% full' true (though not necessarily determinately so), while others will make it false (though not necessarily determinately so). This is just a reflection of the fact that our patterns of usage of the expression 'pretty full' are variegated and highly sensitive to context. And this suggests that in some contexts (V32) will be the culprit, but in others it will be some other premise. This in turn makes the question 'Which of (V1)-(V101) is false?' not worth trying to answer—at least not in any sense of theoretical interest. The best we can hope for are partial answers like 'Not (V1)', or indirect answers like 'One of them, but it's indeterminate which' or 'It depends on the context'. Our general dispositions to use the expression 'pretty full' are simply too flexible for there to be anything deeper to say here.

We think the Sorites puzzle is suggestive in two ways. First, it shows how one could accept the claim "There exists an ordinary situation in which all of the premises of an argument seem true but the conclusion seems false" while rejecting the claim "There is a theoretically fruitful debate about which premise to reject"—as we do in the case of the puzzle about counterfactuals under determinism and the corresponding miracle versus no-miracle debate. And second, it conveys the reason why we think that facts about the generality of purported solutions to the puzzle—i.e., whether we *generally* speak in such a way that makes the relevant premises true or false—are germane to the plausibility of a solution as applied to a single (yet otherwise arbitrary) context. Our contention is that, given

determinism is essentially just a byproduct of unnoticed equivocation. Alternatively, one could adopt the weaker view that *some* instances of the puzzle are to be accounted for in terms of equivocation, while leaving open that there are ways of uniformly interpreting (BP) and (BL) such that they both seem true, despite the fact that at least one of them has to be false. We're neutral on whether either of these views ought to be adopted. But we'll note that neither is of any help to the proponent of the miracle versus no-miracle debate. The stronger view short-circuits the debate by showing that the two sides are talking past each other. The weaker view is less deflating, since it allows the proponents of the debate to frame it in terms of those situations in which (BP) and (BL) both seem true *and* happen to be interpreted uniformly. But then we're just back to the dialectic of the main text.

the arguments of §4, we should adopt a similar attitude to the uniformity assumption for counterfactuals.

Next consider a standard skeptical puzzle about knowledge:

(K1) I know that Trump is the president.

(K2) For all I know, Trump has suffered a fatal heart attack in the past few seconds.

(K3) If I know that Trump is the president, then I know that Trump is currently among the living.

(KC) ∴ I know that Trump is currently among the living, though for all I know he's suffered a fatal heart attack in the past few seconds.

We think it's easy to find contexts in which each of (K1)-(K3) seems true while (KC) seems false. So we have a puzzle. Unlike the puzzle about vagueness described above, there is a live debate among epistemologists about which of the skeptical puzzle's premises ought to be rejected. Proponents of the "skepticism" view say that (K1) is false (and that (K2) and (K3) are true). Proponents of the "no-skepticism" view say that (K2) is false (and that (K1) and (K3) are true). These two positions constitute the "skepticism versus no-skepticism" debate. But some epistemologists, the "contextualists," take a different attitude toward the puzzle. They believe that 'know' is context-sensitive, and that in no single context is each of the premises true (though often these contextualists maintain that in every context (K3) is true). The contextualists we have in mind think the best answer to the general question "Which is true: (K1) or (K2)?" is "It depends on the context".[43]

Suppose we had independent evidence—i.e., evidence that goes beyond the skeptical puzzle—that this sort of contextualism about 'know' is correct: that each of (K1) and (K2) is true in some contexts but false in others. Suppose, in particular, that we had independent evidence that there are "skeptical" contexts in which sentences like (K1) are taken to express clear falsehoods and sentences like (K2) clear truths, as well as "non-skeptical"

[43] Note that this is not a fully neutral characterization of contextualism about 'know'. To give one example: on the most straightforward way of interpreting Lewis's (1996) contextualist theory of 'know', any context in which (K2) seems true is likely to be a context in which (K1) seems false, since, in any such context, the possibility that Trump has had a fatal heart attack in the past few seconds is explicitly being attended to. Indeed, given Lewis's theory, it's not clear that it's even *possible* for there to be a context (in the theoretically relevant sense of the term) in which, for sufficiently attentive speakers, both (K1) and (K2) seem true. From this perspective, what we have been describing as a "context" in which both (K1) and (K2) seem true would better be described as a single *situation* in which both (K1) and (K2) seem true: in particular, a situation in which there is a (perhaps unnoticed) mid-discourse/mid-thought context-shift (cf. footnote 42). Thus, the kind of contextualism about 'know' we're imagining is one that is not so rigidly wedded to Lewis's attention-theoretic meta-semantics, and that allows that the situations in which both (K1) and (K2) seem true needn't always be situations in which speakers are (unwittingly) equivocating on the meaning of 'know'—see, e.g., Holguín (2021) for a defense of an approach of this sort. In any event, for reasons similar to those discussed in footnote 42, we'd be happy to endorse the view that at least some of the instances of the puzzle of counterfactuals under determinism are best explained as involving (unwitting) equivocation. Again, such a view is of no help to the proponent of the miracle versus no-miracle debate.

contexts in which sentences like (K1) are taken to express clear truths and (K2) clear falsehoods. (Much in the same way that we have independent evidence—i.e., evidence that goes beyond the puzzle of counterfactuals under determinism—that each of (BP) and (BL) is true in some contexts but false in others, since there are contexts in which it's clear that the (quasi-)laws are violated and contexts in which it's clear the (macro-)past is violated.) What attitude should we then take to the question "When in a context where both (K1) and (K2) seem true, which premise is the false one?" Should we have much confidence that there is a determinate answer to this question?

Our sense is that we shouldn't. Absent some reason to think that the contexts in which the puzzle is gripping are likely to be uniform with respect to how they resolve the context-sensitivity of 'know', it's hard to see how there could be anything general to say about which premise ought to be rejected beyond "It's indeterminate" or "It depends on the context." If there were some way of identifying a non-trivial subclass of contexts in which either (K1) held uniformly or (K2) held uniformly, and of showing subsequently that such a subclass is likely to be the kind we occupy when we find puzzles like these gripping, then perhaps the question would be worth debating. But we know of no such identification. Indeed, there seems to be no general feature of our use of the word 'know' that would suggest that always (or even very often), if one is in a context where both (K1) and (K2) seem true, it's (K1) that's true and (K2) that's false (or vice versa). It's thus hard to see the point of the imagined "skepticism versus no-skepticism" debate once in the contextualist's frame of mind.

The puzzles about vagueness and knowledge are real puzzles. But given a background commitment to the vagueness and/or context-sensitivity of their central terminology, it seems ill-advised to attempt to resolve these puzzles by singling out a particular premise for rejection. The contexts in which the puzzles arise are not uniform enough for this strategy to work.

We believe the situation with the puzzle of counterfactuals under determinism is analogous. The proponent of the miracle versus no-miracle debate cannot rest content with an ostensive definition of standard contexts as those in which the puzzle gets its grip. Yes, this definition has the effect of ruling out the putative counterexamples to (LAWS) and (PAST) we've presented over the course of the paper, while also not trivializing the question of which of the two principles is correct. But due to the general context-sensitivity of counterfactuals, it is hard to see how or why this class of contexts would be uniform with respect to which of the laws or past is held fixed. Again: we frequently trample over the laws and the facts of actual history when engaging in counterfactual reasoning. Some purposes call for an interpretation of the counterfactual on which the laws are held completely fixed; others call not for fixity, but for small, isolated exceptions; some tolerate but do not demand widespread exceptions; others exist to give such widespread exceptions serious consideration. On the question of the past, just about every counterfactual involves

some amount of tinkering with actual history, for just about every counterfactual involves a ramp. But some counterfactuals exist explicitly for the purpose of considering alternate histories. And many counterfactuals seem massively indeterminate with respect to both laws and history in the worlds that witness their truth. In light of these facts, we see little in our general use of counterfactuals that would explain how or why the uniformity assumption is true—let alone anything that would allow us to identify which side of the miracle versus no-miracle debate is correct.

Consequently, we believe that among the contexts where both (BP, 'If I had blinked once more a second ago, the distant past would have been exactly as it actually is') and (BL, 'If I had blinked once more a second ago, the laws of nature would have been exactly as they actually are') seem true, some are contexts in which (BP) is true and (BL) is false, and some are contexts in which (BL) is true and (BP) is false. We think in many such contexts it will simply be indeterminate which is which, much in the same way it seems to us indeterminate how much of the (quasi-)laws and (macro-)past are held fixed in contexts where counterfactuals like (1a, 'If Caesar had been in command in Korea, he would have used catapults') seem true.[44]

We thus conclude that there is no way of understanding the notion of a standard context so that the question 'Which of (LAWS) or (PAST) is true?' can reasonably be expected to have an answer that is both determinate and non-trivial. The most we can say about the miracle versus no-miracle debate is that each side is correct as far as particular uses of the counterfactual go. Beyond that, neither side has identified a correct generalization about the truth-conditions of counterfactuals—not even when the generalization is restricted to contexts that would seem most hospitable to it, like those in which we reason about decision-making, or in which we grapple with the puzzle of counterfactuals under determinism itself.

7 From Counterfactuals to Laws of Nature

Let us now apply some of the morals we have drawn from the miracle versus no-miracle debate to another debate in metaphysics and the philosophy of science. This application concerns doctrines that try to explain what it is to be a law of nature in terms of certain

[44] One might wonder which features of a given standard context determine which of (BP) or (BL) is true in it. We're not sure. In fact, we're skeptical that there is any general theory to be developed here. But our purposes do not require such a theory. One doesn't have to know which features of context determine the interpretation of a context-sensitive expression to know that that expression exhibits the relevant kind of variability. Indeed, our sense is that those who would respond to the puzzles about vagueness and knowledge in terms of indeterminacy and/or context-sensitivity are in essentially the same situation: they know enough to know that one of the key premises of the puzzle is false in context, but beyond that they shouldn't expect a general recipe for figuring out which it is—at least not if all they know about the context in which the puzzle occurs is that it is one in which each of the premises is appealing.

counterfactuals.[45]

As we noted in §2, principles connecting laws and counterfactuals, such as (LAWS), are widely endorsed.[46] But notice that (LAWS), as stated, presupposes a notion of lawhood. This is because we use the notion of lawhood to define *nomic possibility*, which in turn is used to define *suitability*, which must be included in the principle in order to insulate it from counternomics like (4, 'If the cosmological constant were twice what it actually is, there wouldn't have been any galaxies') and (6, 'If the universe had consisted of only two qualitatively indiscernible angels singing for all eternity, there wouldn't have been any planets'). As a result, (LAWS) illuminates but fails to explain what it is to be a law of nature—rather, the notion of lawhood is taken for granted.

By contrast, the metaphysical doctrines at issue in this section set themselves the more ambitious task of explaining what it is to be a law of nature in independent terms. One widespread construal of such theories is as providing the metaphysical *grounds* of facts detailing what the laws are, revealing such facts to obtain in virtue of facts not involving lawhood. However, even skeptics about grounding can recast these doctrines in their preferred framework for understanding this kind of metaphysical project.[47]

We shall focus on the most thoroughly developed doctrine of this sort: Marc Lange's influential counterfactual theory of lawhood.[48] Glossing over various subtleties that are orthogonal to our criticism, Lange's theory begins by defining a *counterfactually stable set of true propositions* as a set of true propositions every member of which would still have been true under all counterfactual suppositions consistent with every member of the set.[49]

[45] One additional application, which we leave to a footnote, concerns theories of counterfactuals stated in terms of the probabilistic apparatus of statistical physics, which some have taken to lend support to the no-miracle view. See, for example, Albert (2000), Kutach (2002), Loewer (2007), Albert (2015, chs.1-2), and Loewer (2020). We omit detailed discussion of this additional application due to considerations of space. But very roughly, all of our examples where the laws are not held fixed—including all reverse-nomic interpretations and various what-if interpretations—reveal that in many mundane contexts the probabilities of statistical physics are a straightforwardly poor guide to which counterfactuals are true. The reason is that these probabilities are all defined over a space of nomic possibilities, and so trivially any proposition that implies a violation of the laws will receive probability 0, irrespective of which other propositions we conditionalize on. As a result, there is no hope of constructing a general theory of counterfactuals in terms of statistical physics, which is what would be required in order to independently support (LAWS) and the no-miracle view. Of course, it may be that in certain specialized contexts—say, where the laws of nature are made salient—our judgments about which counterfactuals are true correlate with the verdicts of these statistical theories of counterfactuals. But this result would not provide any independent defense of the no-miracle view; rather, the result takes for granted that we are dealing only with contexts in which the laws are held fixed.

[46] See the citations in footnote 11 above. See also Lange's (2009, ch.1) extended discussion of "nomic preservation", which is our (LAWS) modulo some minor differences that don't affect the present dialectic.

[47] For discussions and surveys of ground, see Schaffer (2009), Rosen (2010), Fine (2012), and Bliss & Trogdon (2021). For some skepticism, see Wilson (2014b). For some other prominent frameworks for understanding fundamentality, see Sider (2011), Bennett (2017), Raven (2016), and Dorr (2016b).

[48] Lange has meticulously worked out his theory over numerous articles. Here we focus on his most developed book-length statement of the theory in his (2009). For a related doctrine see Kment (2014).

[49] Lange often employs an unanalyzed notion of consistency among propositions (rather than sentences). One natural means of explaining such a notion is in terms of metaphysical compossibility (compare footnote 5 above). In the main text we follow Lange and simply speak of consistency.

The theory then holds that to be a law of nature just is to be a proposition in any non-maximal counterfactually stable set of true propositions.[50] The theory is also meant to reverse the standard order of explanation among laws and counterfactuals. We're accustomed to thinking that candidate laws of nature are stable under a wide range of counterfactual antecedents precisely *because they are laws*. By contrast, Lange's theory proposes instead to begin with the facts about counterfactual stability and consistency, with no mention of the notion of lawhood. The theory then explains what makes certain propositions laws in terms of antecedently given facts about counterfactual stability.

A widespread reaction to Lange's theory, which (the reaction) we are sympathetic with, is that the theory gets the order of explanation wrong from the outset: counterfactuals, precisely due to their vagueness and context-sensitivity, are ill-suited to form the basis of a metaphysical theory of laws of nature. The challenge is in making this initial reaction more concrete. Indeed, Lange's main reply to it is that his "philosophic conscience" concerning which notions are suited to form the basis of metaphysical explanations differs from that of his critics, particularly in light of all of the virtues and applications he sees in his own theory.[51] Such issues frequently lead to dialectical stalemate, and for this reason critics have sought sharper and more concrete criticisms of the theory in the form of direct counterexamples.

In order to be extensionally adequate, Lange's theory requires the no-miracle view to be true. After all, the theory is premised on laws of nature generally being held fixed under counterfactual antecedents that are nomically possible—it just ingeniously manages to characterize this requirement independently of lawhood via the notion of a counterfactually stable set. For example, if Lewis' miracle view is correct—and so no law of nature would still have been true even given mundane counterfactual suppositions like my having blinked one more time in the past hour than I in fact did—then Lange's theory predicts that there are no laws of nature. Lange is well aware that his theory requires the no-miracle view; to give just one example, he writes: "Although context wields great influence over counterfactuals, there is a limit to its influence: in no context does an accident take priority over a law under the counterfactual supposition [that one or the other is false]" (2009, p.35).

Given this requirement of the theory, it is unsurprising that various putative counterexamples to the no-miracle view have been marshalled to argue against Lange's theory.[52] Such counterexamples present putative cases in which an accident does indeed "take priority over a law under the counterfactual supposition that one or the other is false." Notice

[50] Non-maximality is required because the set of all true propositions (assuming there is such a set) would be trivially counterfactually stable given standard centering assumptions in the logic of counterfactuals. Yet not all true propositions are laws of nature.

[51] See Lange (2009, pp.188-89). The description of one's views concerning which notions stand in need of explanation as one's "philosophic conscience" of course traces back to Goodman (1955, ch.2.1).

[52] See, for example, Demarest (2012), Hall (2011), and the critiques by Woodward, Loewer, and Carroll in the symposium Woodward *et al.* (2011).

that a number of our counterfactuals above are cases of exactly this form, such as all of our reverse-nomics and many of our what-ifs (14)–(16). And of course Lange cannot insist that his account applies only in contexts in which the laws are held fixed, and hence such counterexamples do not arise, without rendering his theory blatantly circular.

Lange's preferred response to counterexamples of this form appeals to elided material in the counterfactuals' antecedents.[53] But we argued in §5.1 that this reply fails. Most importantly, reading the extant critiques of Lange, as well as his replies, conveys the impression that the problems for his theory simply stem from a few exotic and isolated counterexamples, which can then be dealt with on a case-by-case basis. But our arguments above bring out widespread and systematic problems for the very intuitions that are meant to motivate (LAWS), and hence widespread and systematic problems for the foundation on which Lange's theory rests.

8 Conclusion

A general moral of our discussion, alluded to at the outset, is that we should be especially wary of philosophical theories that tether some relatively fundamental subject matter to counterfactuals. We have seen how this moral applies concerning laws of nature, but the form of our argument generalizes. On the one hand, given some account of counterfactuals in terms of some independent subject matter, the immense context-sensitivity of counterfactuals makes it likely that the account at best accurately describes which counterfactuals are true in some circumscribed range of contexts. On the other hand, accounts of some subject matter in terms of counterfactuals require first delineating the relevant range of contexts in which the counterfactuals are interpreted, and a challenge arises as to whether this can be done without employing the independent subject matter itself, thereby rendering the theory circular (as in the case of Lange's counterfactual theory of laws of nature). This dialectic does not arise for subject matters that can plausibly be held to be as context-sensitive as counterfactuals themselves. A counterfactual theory of some such subject matter might then maintain that the subject matter's context-sensitivity perfectly marches in step with counterfactual context-sensitivity. But precisely due to their context-sensitivity, such subject matters are poor candidates to be relatively fundamental, just as counterfactuals themselves are.

Let us close where we began: counterfactuals are context-sensitive. We hope to have given this truism its due.

[53] See, for example, Lange (1993) and Lange (2009, pp.197-8). Lange (2009, pp.198-205) also considers some alternative replies, including various error-theories of the sort we argued against in §5.3.

References

Ahmed, Arif. 2013. Causal Decision Theory: A Counterexample. *Philosophical Review*, **122**(2), 289–306.

Ahmed, Arif. 2014. Causal Decision Theory and the Fixity of the Past. *British Journal for the Philosophy of Science*, **65**(4), 665–685.

Albert, David Z. 2000. *Time and Chance*. Cambridge, Mass: Harvard University Press.

Albert, David Z. 2015. *After Physics*. Cambridge, Massachusetts: Harvard University Press.

Bacon, Andrew. 2022. Actual Value in Decision Theory. *Analysis*, **82**(4), 617–629.

Beebee, Helen. 2003. Local Miracle Compatibilism. *Noûs*, **37**(2), 258–277.

Bennett, Jonathan. 1984. Counterfactuals and Temporal Direction. *Philosophical Review*, **93**(1), 57–91.

Bennett, Jonathan. 2003. *A Philosophical Guide to Conditionals*. Oxford University Press.

Bennett, Karen. 2017. *Making Things Up*. Oxford: Oxford University Press.

Bliss, Ricki, & Trogdon, Kelly. 2021. Metaphysical Grounding. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*, winter 2021 edn.

Demarest, Heather. 2012. Do Counterfactuals Ground the Laws of Nature? A Critique of Lange. *Philosophy of Science*, **79**(3), 333–344.

Dorr, Cian. 2014. Transparency and the Context-Sensitivity of Attitude Reports. *Pages 25–66 of:* García-Carpintero, Manuel, & Martí, Genoveva (eds), *Empty Representations: Reference and Non-existence*. Oxford University Press.

Dorr, Cian. 2016a. Against Counterfactual Miracles. *Philosophical Review*, **125**(2), 241–286.

Dorr, Cian. 2016b. To Be F Is To Be G. *Philosophical Perspectives*, **30**, 39–134.

Elga, Adam. 2022. Confession of a Causal Decision Theorist. *Analysis*, **82**(2), 203–213.

Fine, Kit. 2012. Guide to Ground. *Pages 37–80 of:* Correia, Fabrice, & Schnieder, Benjamin (eds), *Metaphysical Grounding*. Cambridge University Press.

Gallow, J. Dmitri. 2023. Causal Counterfactuals Without Miracles or Backtracking. *Philosophy and Phenomenological Research*, **107**(2), 439–469.

Gallow, J. Dmitri. 2024. Counterfactual Decision Theory Is Causal Decision Theory. *Pacific Philosophical Quarterly*, **105**(1), 115–156.

Gibbard, Allan, & Harper, William. 1978. Counterfactuals and Two Kinds of Expected Utility. *Pages 125–162 of:* Hooker, C. A., Leach J. J. & McClennen E.F. (ed), *Foundations and Applications of Decision Theory*, vol. 1. Dordrecht: D. Reidel.

Gillies, Anthony S. 2007. Counterfactual Scorekeeping. *Linguistics and Philosophy*, **30**(3), 329–360.

Glynn, Luke. 2013. Of Miracles and Interventions. *Erkenntnis*, **78**(S1), 43–64.

Goggans, Phillip. 1992. Do the Closest Counterfactual Worlds Contain Miracles? *Pacific Philosophical Quarterly*, **73**(2), 137–149.

Goodman, Jeremy. 2015. Knowledge, Counterfactuals, and Determinism. *Philosophical Studies*, **172**(9), 2275–2278.

Goodman, Nelson. 1955. *Fact, Fiction, and Forecast*. Cambridge, Mass.: Harvard University Press.

Hájek, Alan. 2014. Most Counterfactuals Are False. Unpublished manuscript.

Hájek, Alan. 2021. Contra Counterfactism. *Synthese*, **199**(1-2), 181–210.

Hall, Ned. 2011. Review of *Laws & Lawmakers: Science, Metaphysics, and the Laws of Nature*. Notre Dame Philosophical Reviews.

Halpin, John F. 1991. The Miraculous Conception of Counterfactuals. *Philosophical Studies*, **63**(3), 271–290.

Hedden, Brian. 2023. Counterfactual Decision Theory. *Mind*, **132**(527), 730–761.

Holguín, Ben. 2021. Knowledge by Constraint. *Philosophical Perspectives*, **35**(1), 275–302.

Jackson, Frank. 1977. A Causal Theory of Counterfactuals. *Australasian Journal of Philosophy*, **55**(1), 3–21.

Joyce, James M. 1999. *The Foundations of Causal Decision Theory*. Cambridge University Press.

Khoo, Justin. 2017. Backtracking Counterfactuals Revisited. *Mind*, **126**(503), 841–910.

Kment, Boris. 2014. *Modality and Explanatory Reasoning*. Oxford: Oxford University Press.
Kripke, Saul A. 1980. *Naming and Necessity: Lectures Given to the Princeton University Philosophy Colloquium*. Cambridge, MA: Harvard University Press.
Kutach, Douglas N. 2002. The Entropy Theory of Counterfactuals. *Philosophy of Science*, **69**(1), 82–104.
Lange, Marc. 1993. When Would Natural Laws Have Been Broken? *Analysis*, **53**(4), 262–269.
Lange, Marc. 2009. *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*. Oxford University Press.
Lewis, David. 1973. *Counterfactuals*. Blackwell.
Lewis, David. 1979. Counterfactual Dependence and Time's Arrow. *Noûs*, **13**(4), 455–476.
Lewis, David. 1981. Causal Decision Theory. *Australasian Journal of Philosophy*, **59**(1), 5–30.
Lewis, David. 1996. Elusive Knowledge. *Australasian Journal of Philosophy*, **74**(4), 549–567.
Lewis, Karen S. 2018. Counterfactual Discourse in Context. *Noûs*, **52**(3), 481–507.
Loewer, Barry. 2007. Counterfactuals and the Second Law. *Pages 293–326 of:* Price, Huw, & Corry, Richard (eds), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford University Press.
Loewer, Barry. 2020. The Mentaculus Vision. *Pages 3–29 of:* Allori, Valia (ed), *Statistical Mechanics and Scientific Explanation*. World Scientific.
Maudlin, Tim. 2007. *The Metaphysics Within Physics*. Oxford; New York: Oxford University Press.
McNamara, Calum. 2024. Causal Decision Theory, Context, and Determinism. *Philosophy and Phenomenological Research*, **109**(1), 226–260.
Moss, Sarah. 2012. On the Pragmatics of Counterfactuals. *Noûs*, **46**(3), 561–586.
Munroe, Randall. 2014. *What If? Serious Scientific Answers to Absurd Hypothetical Questions*. Houghton Mifflin Harcourt.
Nolan, Daniel. 2017. Causal Counterfactuals and Impossible Worlds. *Pages 14–32 of:* Beebee, Helen, Hitchcock, Christopher, & Price, Huw (eds), *Making a Difference*. Oxford University Press.
Plantinga, Alvin. 1974. *The Nature of Necessity*. Oxford University Press.
Raffman, Diana. 1994. Vagueness Without Paradox. *Philosophical Review*, **103**(1), 41–74.
Raven, Michael J. 2016. Fundamentality without Foundations. *Philosophy and Phenomenological Research*, **93**(3), 607–626.
Rosen, Gideon. 2010. Metaphysical Dependence: Grounding and Reduction. *Pages 109–136 of:* Hale, Bob, & Hoffmann, Aviv (eds), *Modality: Metaphysics, Logic, and Epistemology*. Oxford University Press.
Schaffer, Jonathan. 2009. On What Grounds What. *Pages 347–383 of:* Manley, David, Chalmers, David J., & Wasserman, Ryan (eds), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.
Schaffer, Jonathan, & Szabó, Zoltán Gendler. 2014. Epistemic Comparativism: A Contextualist Semantics for Knowledge Ascriptions. *Philosophical Studies*, **168**(2), 491–543.
Sider, Theodore. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.
Skyrms, Brian. 1982. Causal Decision Theory. *Journal of Philosophy*, **79**(11), 695–711.
Sobel, Jordan Howard. 1994. *Taking Chances: Essays on Rational Choice*. Cambridge, England: Cambridge University Press.
Stalnaker, Robert. 1968. A Theory of Conditionals. *Pages 98–112 of:* Rescher, Nicholas (ed), *Studies in Logical Theory*, vol. 2. Oxford: Blackwell.
Stalnaker, Robert. 1981. Letter to David Lewis. *Pages 151–152 of:* William Harper, Glenn Pearce, & Stalnaker, Robert (eds), *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Dordrecht: Reidel.
van Inwagen, Peter. 1985. Plantinga on Trans-World Identity. *Pages 101–120 of:* Tomberlin, James E., & van Inwagen, Peter (eds), *Alvin Plantinga*. Dordrecht: D. Reidel.
Vihvelin, Kadri. 2000. Libertarian Compatibilism. *Philosophical Perspectives*, **14**(14), 139–166.
von Fintel, Kai. 2001. Counterfactuals in a Dynamic Context. *Pages 123–152 of:* Kenstowicz,

Michael (ed), *Ken Hale: A Life in Language*. The MIT Press.
Wilson, Jessica. 2014a. Hume's Dictum and the Asymmetry of Counterfactual Dependence. *Pages 258–279 of:* Wilson, Alastair (ed), *Chance and Temporal Asymmetry*. Oxford University Press.
Wilson, Jessica. 2014b. No Work for a Theory of Grounding. *Inquiry*, **57**(5-6), 535–579.
Woodward, Jim, Loewer, Barry, Carroll, John W., & Lange, Marc. 2011. Counterfactuals All the Way Down? *Metascience*, **20**(1), 27–52.