Most Counterfactuals are False

Alan Hájek

DRAFT

DEAR READER,

YOU MAY APPRECIATE SOME TIPS ON WHICH SECTIONS COULD BE SKIPPED ON A FIRST READING. IF YOU ARE LACKING IN TIME OR STAMINA, I RECOMMEND THAT YOU SKIP OR SKIM THROUGH SECTIONS 6 AND 7, SPARING YOU OVER 20 PAGES. I'VE PUT IN GRAY FONT THESE, AND OTHER BITS I SKIPPED IN MY APA TALK. FAST-FORWARDING THROUGH SECTION 5 WOULD SPARE YOU ABOUT ANOTHER 10 PAGES, WITHOUT LOSING THE GIST OF THE PAPER. [THERE ARE ALSO VERY OCCASIONAL NOTES TO MYSELF, USUALLY IN SQUARE PARENTHESES.]

A.H
.

1. Introduction

"Even the crows on the rooftops are cawing about the question of which conditionals are true". So said Callimachus over 2000 years ago.[1] And my answer to the question is: when it comes to counterfactuals, *relatively few* of them.

The crows have much to caw about. For counterfactuals are apparently implicated in much that we hold dear. They figure in influential analyses of causation, perception, knowledge, personal identity, laws of nature, rational decision, confirmation, dispositions, free action, explanation, and so on. Science freely traffics in counterfactuals, both explicitly (in drawing consequences of its theories[2]) and implicitly (by trafficking in concepts that are themselves tacitly counterfactual). And counterfactuals are also earning their keep in the social sciences—in psychology (e.g., in understanding emotions such as regret), in history (e.g., in studying the

[1] Attributed by Sextus Empiricus, Against the *Mathematicians*. Quoted in B. Mates, *Stoic Logic* (Berkeley & Los Angeles: University of California Press, 1961), 43.

[2] Consider this example from the classic physics textbook, Halliday and Resnick (19xx), p. 251: If we were to bore a hole through the earth, and drop a particle into the hole, it would move in simple harmonic motion.

incremental contribution by some commodity to economic growth), in the law (e.g., in apportioning responsibility), and so on.

Yet it has long been recognized that counterfactuals are strange beasts.[3] They involve a modality that makes empiricists uncomfortable; they resist truth functional analysis, yet the best-known possible worlds analyses of them make various philosophers uncomfortable; they putatively violate intuitive inference rules that material and strict conditionals obey, such as contraposition and transitivity and strengthening of the antecedent[4]; and so on. I will argue that they are even stranger than has been generally recognized: while we use them nonchalantly in daily conversation, and while they are staples of numerous philosophical analyses, most counterfactuals are false.

2. Counterfactuals under indeterminism: most ordinary counterfactuals are probably false

In what follows, by "counterfactuals" I will mean 'would'-counterfactuals of the form 'if X were the case, Y would be the case', denoted 'X → Y'. I will also speak of "'might' counterfactuals" (always so-qualified) of the form 'if X were the case, Y might be the case', denoted 'X ◊→ Y'. I will focus on two strategies for showing a counterfactual to be false: appealing to *indeterminism*—in particular, chanciness; and to *indeterminacy*—in particular, imprecision. Both are strategies for securing the truth of 'might not' counterfactuals that are, I will argue, incompatible with the corresponding 'would' counterfactuals. I will find it convenient to argue via the truth of the 'might-nots' to the falsehood of the 'woulds', but I think that the strategies can also be deployed directly: a counterfactual cannot second-guess

[3] A fairly reliable source tells me that in Sweden they tried to pass a law to abolish the subjunctive conditional.

[4] See Lewis 1973, §1.8. We will return to these rules in §7.

the outcome of a process that is chancy, or the resolution of an indeterminacy. The best way to understand what I mean is to see these strategies in action. I begin with indeterministic cases.

Suppose for now that coin-tossing is a chancy business: as a coin is tossed, it is a genuinely indeterministic matter whether it lands heads or tails; and for now assume for simplicity that these are the only possible outcomes. Here is a coin that in fact will never be tossed. Consider the counterfactual:

‘If the coin were tossed, it would land heads’,

symbolically,

‘Toss → Heads’.

I submit that it is *false*. I think you can see that directly: focus on the chanciness of the outcome of the toss. Or consider the ‘might’ counterfactual: ‘If the coin were tossed, it *might* land tails’, symbolically ‘Toss ◊→ Tails’. This is true—it is implied by our supposition of chanciness of the outcome. But I claim that the original ‘would’ counterfactual and the ‘might’ counterfactual are contraries: ‘Toss → Heads’ and ‘Toss ◊→ Tails’ cannot both be true. Since the ‘might’-counterfactual is true, it’s the ‘would’-counterfactual that must take the fall.

There are several ways to arrive at this conclusion.

The ‘would/might’ duality

The first way regards ‘would’ and ‘might’ counterfactuals as duals, à la Lewis (1973): X → Y is equivalent to ¬(X ◊→ ¬Y). Then ‘Toss → Heads’ and ‘Toss ◊→ ¬Heads’ are *contradictories*—they necessarily have opposite truth values. The duality of the ‘would’ and ‘might’ counterfactuals has been defended by a number of authors (e.g. Bigelow and Pargetter (1990, 103), and Bennett (2003, 192)), and assumed by others (e.g. Hawthorne 2005, and

Williams 2007). It has also appealed to some authors who have written on counterfactuals in connection with the debate over whether God has 'middle knowledge' of the truth of counterfactuals concerning free actions—see, e.g., Adams (1977) and van Inwagen.

But my argument does not need the full strength of the 'would/might' duality—it suffices that 'Toss → Heads' and 'Toss ◊→ ¬Heads' are *contraries.*

The elimination of possibilities

Just try saying out loud:

> "If the coin were tossed, it MIGHT land Tails, and it WOULD land Heads if it were tossed" (*)

There is surely something defective about (*). More generally, there is serious tension between assertions of 'might' and corresponding 'would not' counterfactuals, or between 'would' and corresponding 'might not' counterfactuals. DeRose (1999) calls this the phenomenon of "inescapable clashes". Here is an argument that the clash is semantic. The 'might' counterfactual recognizes Tails possibilities, while the 'would' counterfactual eliminates them. Of course, the 'would' counterfactual leaves open various possibilities, corresponding to various ways in which Heads could be realized—the coin landing Heads at noon, the coin landing Heads with a soft metallic tinkle, the coin landing Heads and then immediately afterwards vanishing, the coin landing Heads while John Howard impersonates a chicken, and so on. But they are all *Heads*-possibilities; Tails does not occur in any possibility left open by 'Toss → Heads'. On the other hand, the 'might'-counterfactual is committed to at least one Tails possibility remaining live. And rightly so: clearly, 'Toss ◊→ Tails' is true. Thus, 'Toss → Heads' is false.

Redundancy

Here is another way to see the point. I say: "if I were to toss the coin, it MIGHT not land heads". I then add: "And furthermore, it's not true that it WOULD land heads if I were to toss it." You ought to be puzzled. My utterance of "furthermore" primed you to expect more information, but none was forthcoming. Instead, what came next was redundant.

Disagreement and retraction

You claim that if the coin were tossed, it would land heads. I *disagree.* One way of stating my disagreement is to remind you that if the coin were tossed, it might not land heads. We can't both be right. (In fact, how else can I *merely* deny what you said, rather than committing myself to something stronger than the denial? That is another consideration in favor of the 'would/might' duality.)

Similarly, I may assert 'Toss → Heads' and later come to retract it, so that I have temporal stages of myself disagreeing with each other. One way that my later self could give grounds for his retraction would be to note that if the coin were tossed, it might land tails. (The earlier self may have had misleading evidence regarding the coin—e.g. being told by an unreliable source that it was two-headed.)

Evidence

Think of the *evidence* that I could muster in support of my claim that if I were to toss the coin, it might land tails—for example, the fact that it has landed tails in the past, or that another, similar coin that I just tossed landed tails. But to the same extent, these facts counter-support the 'would land heads' counterfactual.

I assumed for the sake of simplicity that ‘Heads’ and ‘Tails’ were the only possible outcomes. If the assumption is false, for example, because ‘edge’ is another possible outcome, no matter. In that case, ‘Toss → Heads’ and ‘Toss ◊→ Tails’ are clearly not contradictories, but they are still contraries. The truth of ‘Toss ◊→ Tails’ still implies the falsehood of ‘Toss → Heads’. Or we may bypass considerations of ‘Tails’ altogether: with some other outcome possible, *a fortiori* ‘Toss ◊→ ¬Heads’ is true (there being a further way that Heads could fail to happen), and it is contrary to ‘Toss → Heads’.

Our foot is in the door; now let’s kick it open. I deliberately did not specify the chance of Heads. It is fine if you assumed that the coin was fair, but I did not. The argument would go through equally well, whatever the chance of Tails, as long as it is a possible outcome (and our supposition of chanciness assures us that it is). Then ‘Toss ◊→ Tails’ is still true, and that is all I need to establish that ‘Toss → Heads’ is false. For example, let the chance of Tails be 0.000001. It remains true that the coin *might* land Tails if it were tossed, undermining the corresponding ‘would’ counterfactual concerning Heads. The point generalizes. Whenever we have a true ‘might’ counterfactual of the form X ◊→ Y, any corresponding ‘would’ counterfactual X → Z, where Z is a contrary of Y, is false. ‘If I were to play the lottery, I would lose’ is false, because I *might* win, no matter how many tickets there are in the lottery.

Indeed, we can drive the chance of the consequent all the way to 1 and *still* have a false counterfactual. Now suppose that the coin is to be tossed repeatedly infinitely many times. It *might* land Tails on every toss, even though the chance of this is 0. (Again, I deliberately did not specify the chance of Heads: as long as tails is a possible outcome of any toss, it is a possible outcome of *every* toss, assuming that the trials are independent.) Thus, I cannot truly say ‘if I were to toss this coin forever, it would land Heads eventually’. It might not.

This infinite coin toss experiment corresponds to an infinite, highly biased lottery: ticket #i wins iff the coin lands heads for the first time on the i^{th} toss. In this lottery, it is possible that no ticket will win (even though this can only happen in one way, and it can fail to happen in infinitely many ways—to which, moreover, the probability distribution is heavily biased if the coin is fair). So I cannot even truly say 'if I were to hold all the tickets in the lottery, I would win'. I might not.

In an indeterministic world such as ours appears to be, lotteries—in a broad sense—abound. I say 'appears to be', because we are not certain that the world we live in *is* indeterministic, but the evidence from quantum mechanics certainly seems to point that way.[5] And it isn't just the canonical quantum mechanical examples—radioactive decay, spin measurements on a particle in a Stern-Gerlach apparatus, and so on—that are indeterministic. The indeterminism reaches medium-sized dry goods (and even oversized wet ones), just less obviously so. Two billiard balls colliding may approximate a deterministic system, but even they are not immune from quantum mechanical indeterminism. One ball might spontaneously tunnel through the other, or to China, or to the North Star—incredibly unlikely, to be sure, but possible. Thus, I cannot truly say 'if the cue ball were to hit the 8 ball, the 8 ball would begin rolling'. Or again, whenever I jump in the air, there is a minute chance that I will not come down—I might vaporize instead, for the chance of that happening is non-zero. Thus, I cannot truly say 'if I were to jump, I would come down'. With indeterminism reaching so far, a surprisingly large array of ordinary counterfactuals really have chancy consequents. Thus, they are as 'bad' as the coin tossing counterfactual with which I began, as 'bad' as counterfactuals about lottery outcomes that are not guaranteed. Indeed, these ordinary counterfactuals are 'worse' than the earlier counterfactual about the infinite sequence of coin tosses, because

[5] To be sure, Bohmian mechanics is deterministic. In the next section I argue that determinism will not save most counterfactuals from being false.

unlike that counterfactual, the ordinary counterfactuals have consequents whose chances are less than 1. Ordinary the counterfactuals may be, but that doesn't save them from being false.

At this point you may want to protest.

> PROTEST: Granted, anomalous results such as a billiard ball quantum tunneling to China or a person vaporizing have positive chance of occurring. But in the *nearest* possible worlds in which the relevant antecedents are true, the results do not occur. A person's vaporizing mid-jump, for example, is such a *bizarre* event that a world in which it happens is rather remote from ours, and in particular more remote than worlds in which he falls normally. Thus, the counterfactuals come out true after all.

I believe that you have effectively denied that the 'might'-counterfactuals are true. For example, in claiming that all the nearest worlds in which I jump are worlds in which I fall normally, you have ruled out my vaporizing from being among them, thus apparently ruling out that I might vaporize were I to jump. But what entitles you to do that? (We will revisit this point later if you are not already convinced of it.)

Not this: the fact that the chance of my vaporizing is *small*. That would be reminiscent of a methodological precept—a bad one—adopted variously by Cournot, Kolmogorov, and Popper *to regard events with small probabilities as impossible*. You might think of it as a maxim to 'round off' small probabilities. I think of it, at least in Popper's case, as an attempt to fit the round peg of inductivism into the square hole of deductivism—to square the circle, if you will. The precept might make our lives easier, but it is bad for probability's philosophical foundations. Far from being impossible, events with small probabilities happen all the time—indeed, the perfectly normal jump that I just performed was one of them. Your reading these words, exactly as you are, is another one of them.

In any case, having a highly probable consequent is not sufficient for a counterfactual's being true, for reasons akin to the lottery paradox. Wherever you set the threshold for 'highly probable' (assuming that it is below 1), it will not be high enough. Suppose, for example, that you set the threshold at 0.999999—you claim, for example, that if a coin is biased at least 99.9999% towards Heads, then 'Toss → Heads' is true. But imagine a non-actual lottery with a 1,000,001 tickets. Consider the following list of a million and one counterfactuals, each of which has a consequent with probability above the threshold:

lottery is played → ticket #1 loses;

lottery is played → ticket #2 loses;

…

lottery is played → ticket #1000001 loses.

They can't all be true, because the counterfactuals are jointly inconsistent (there is no possible world in which the lottery is played and in which all of the tickets lose). So at least one is false —a counterexample to sufficiency of the 0.999999 threshold. And so it goes for any putative threshold (below 1); I will just imagine a lottery with enough tickets to thwart it. Notice that the analogue of a popular response to the lottery paradox—denying that belief is closed under conjunction—is not available here. For denying that *truth* is closed under conjunction would be a far more radical thesis than my thesis that most counterfactuals are false! Or you may instead insist that the threshold is context-dependent—unusually high in lottery contexts (so that all of the million and one counterfactuals above come out false), lower in various other contexts. It is unclear exactly what counts as a context, but there is a concern that on any reasonable understanding of it, the threshold account will deliver unintuitive results of the kind that it was intended to avoid. For example, counterfactuals like

lottery is played → the owner of the winning ticket is happy

and perhaps even

lottery is played → the sun rises the following day will come out false, since if the lottery is large enough, it will set up a context in which the threshold is so high that the probabilities of the consequents will not achieve it. And as I say, in a broad sense lotteries abound. I challenge the context-dependent-threshold theorist to come up with an account that correctly adjudicates as false the million and one counterfactuals above (and likewise for similar counterfactuals, however large the lottery), while adjudicating as true the ordinary counterfactuals about cup fallings and so on.[6]

Indeed, even setting the probability threshold *at 1* will not suffice. We have already seen a probability zero event that is possible: the coin landing Tails forever. Such events are forced upon us if we have uncountable probability spaces. For example, assuming that the radioactive decay laws involve continuous rather than step-functions, just as physics says they do, then probability zero events abound. So in this sense even zero probabilities cannot be 'rounded off'. They certainly are not equivalent to impossibilities—and given our assumptions, that's a genuine certainty, not merely the probability 1 kind!

I put the word "*bizarre*" in your mouth during your imagined Protest, and presumably that goes beyond merely "having small probability". After all, your hearing these words, exactly as you are, was antecedently improbable, but it is not (I hope!) a bizarre event. Fleshing out what "bizarre" really means will be no mean feat. Still, I agree that in some good, intuitive sense, there is something strange about the events I am imagining; I used the word "anomalous" myself. My ill-fated jump's positive probability does not save it from that epithet. But I am asking very little of that ill fate: all I need to be true is the modal claim that *I might vaporize*, were I to jump. If you want to take issue with that, you have not only me but also quantum mechanics to contend with.

[6] A detailed discussion of all the moves and countermoves that might be considered would derail this already lengthy paper. Many of them will parallel moves that Hawthorne (200x) considers in his discussion of how *knowledge* claims are incompatible with their chanciness. My skepticism about the truth of counterfactuals in the face of chanciness parallels his skepticism about knowledge claims in the face of chanciness.

So if nothing else, quantum mechanics is there to guarantee the truth of the ‘might’ counterfactuals that undermine the corresponding ‘would’ counterfactuals. But often less esoteric facts will equally do the job. Holding my cup tantalizingly over a hard floor, you say: “If I were to let go of the cup, it would fall. And if it were to fall and hit the floor, it would break.” Well, no, and no—it might not, and it might not. If I were to let go of the cup, a sudden gust of wind might lift it higher; and if it were to fall and hit the floor, another gust of wind might slow down its fall sufficiently to spare it a damaging impact. Or even less esoterically, I might catch the cup, sparing it an impact altogether. Quantum mechanics is just a handy, cover-all way for me to secure the truth of a huge raft of undermining ‘might’ counterfactuals in one fell swoop. But other anomalous happenings could do the job just as well on a case-by-case basis.

You may grant me that that there would be a *chance* of these anomalous happenings, but still deny that they *might* happen if the relevant antecedents were realized. You remind me that your protest did not question the chanciness of such bizarre outcomes, but rather their occurrence in the *nearest* worlds where the antecedents are true—their bizarreness relegates them to more distant worlds. Lewis (1986) would call them *quasi-miracles*—‘remarkable’ events that, while compatible with the laws in virtue of their chanciness, detract from overall similarity to the actual world. I confess to finding the notion of a ‘quasi-miracle’ rather nebulous, and one that courts anthropomorphism. There is a good sense in which quantum mechanical events are ‘remarkable’—we certainly find them so—yet there is no shortage of them in the actual world. Seen from the perspective of chancy laws, there should be no favoritism against the anomalous events that I am countenancing. They should not, as it were, be penalized twice—once for their low chance, and a second time for their bizarreness! Rather, the laws treat them even-handedly, much as we treat lottery outcomes—sensitive to their low chance, but no more. After all, ‘chances’ are written into the laws themselves, but ‘degrees of bizarreness’ are not.

But if we insist on pushing them out to more distant worlds—anthropomorphically, *I* insist—then note that we seem to get the result that *nothing* bizarre happens in our neighborhood of possible world space. For example, let’s agree that it is bizarre if a fair coin that is tossed a 100 times in a row lands all Heads—it is, in words that Lewis uses to characterize a quasi-miracle, “a remarkable coincidence” (60). Then we seem to have the result that even if the 100-toss experiment were run a googolplex times, in *none* of the experiments would we get all Heads; after all, for each experiment, *its* yielding the bizarre all-Heads outcome detracts from overall similarity to the actual world. More generally, it would seem that our neighborhood of worlds, despite its populousness, consists solely of *dull* worlds, ones free of any remarkable coincidences that are absent in our world. That in itself seems somewhat remarkable! (See Hawthorne 2005 and Williams 2007 to whom this paragraph is indebted.)

And if we insist on pushing out ‘bizarre’ possibilities to more distant worlds, then presumably we should *pull in* ‘familiar’ possibilities to *less* distant worlds, with implausible results. Bill Gates and a particular bunch of a million ordinary people have not entered a 1,000,001-ticket lottery, with a million dollar prize. What would have happened if they *had* entered it? Bad answer: Gates would have won. But there is nothing special about a world in which Gates wins a million dollars—that’s hardly a ripple in his financial fortunes—whereas we may suppose it to be a *major* change in the life of any of the other contestants. To be sure, none of these outcomes are bizarre in the way that a quantum tunneling is; but yet another million-dollar increase in Gates’s fortune is such a familiar occurrence that it is singularly *not* bizarre, while such an increase for an ordinary person is at least a little stranger. Here it is tempting to reply that all the worlds corresponding to the lottery outcomes are *exactly equally* distant, with nothing favoring the Gates-winning world for closeness. Thinking of the lottery as a purely probabilistic process, this may seem right, but then we are paying no attention that the differences in similarity that the *outcome* of the process can make. In that case we should think the same way about the cup 'lottery', one of whose outcomes is the cup's quantum tunneling. On the other hand, if the-cup-falls-normally worlds are regarded as closer than a quantum tunneling world because of their being more familiar (less bizarre), then the Gates-winning world should be accorded the same courtesy.

If you are not convinced by this example, consider instead: "If the lottery were to take place, it would be won by someone whose winning makes their world maximally similar to the actual world"—the appeal to Bill Gates was just meant to make this vivid. Or better still, consider the general recipe that the example is meant to exemplify: "If a chance process were to take place, it would yield an outcome that makes the resulting world maximally similar to the actual world." I hope that such second-guessing of the result of a hypothetical chance process strikes you as improper. And yet just such second-guessing underwrote your Protest.

[Said slightly differently: Abnormality comes in degrees. The worry is that if SLIGHTLY more abnormal outcomes are SLIGHTLY less similar, then we will get the wrong results for various counterfactuals, because comparative similarity is (allegedly) all that matters to their truth conditions. Suppose that Bill Gates and a million regular people enter a lottery. It is SLIGHTLY more abnormal for a regular person to add millions to his fortune than it is for Bill Gates to do so. The worry is that 'if the lottery were played, Bill Gates would win' will come out true as long as there is ANY penalty, however small, for the abnormality of others winning.]

Yet I am still being too concessive to your Protest. You have insisted that my mid-jump vaporization takes us to *less similar* worlds than worlds where things go as expected. Let me grant that for the sake of the argument. It is still a further step for you to reach the conclusion that the counterfactual 'if I were to jump, I would come down' is *true.* "Ah, but that follows from the usual Stalnaker-Lewis style semantics for counterfactuals", you say. Agreed—but perhaps we should question such similarity-based semantics. In the thick of a disagreement about the orthodoxy concerning the truth-values of counterfactuals, to presuppose the orthodoxy of some similarity-based semantics for counterfactuals is to put the cart before the horse. So let us revisit that orthodoxy. I will now argue that the connection between similarity and the truth-conditions for counterfactuals is far less straightforward than has been widely assumed.

A chancy coin is hooked up to a Doomsday machine. If the coin is tossed and it lands Heads, nothing interesting happens: it's business as usual, status quo. If the coin is tossed and it lands Tails, something very interesting happens: the Doomsday machine obliterates the world and surrounding districts, resulting in vast, widespread changes to the status quo. In fact the coin is never tossed. But what would happen if it *were* tossed? Bad answer: it *would* land Heads. Bad answer, because the chanciness of the coin should preclude us from giving this verdict. I take the chanciness of the coin to be incompatible with the truth of 'Toss → Heads'. Nevertheless, various similarity accounts appear to be committed to the truth of this counterfactual. After all, intuitively the nearest 'Toss' worlds are 'Heads' worlds—business as usual is more similar to the actual world than Doomsday.

This example evokes Kit Fine's famous argument against Lewis: it appeared that Lewis was committed to denying that 'if Nixon had pressed the button, there would have been a nuclear holocaust'. Lewis, equally famously, replied by laying down an ordering of what matters in judgments of similarity of worlds: "(1) It is of first importance to avoid big, widespread, diverse violations of law. … (4) It is of little or no importance to secure approximate similarity of particular fact" (1986, 47-48). I elide over the details of Lewis's ordering. The upshot was that he argued that this ordering vindicated the intuitively correct verdict on the Nixon counterfactual.

My example differs from Fine's in its *chanciness.* The chanciness of the coin bars us from judging 'Toss → Heads' to be true—that would be tantamount to second-guessing an avowedly indeterministic process. Fine's example also turns partly on intuitions about closeness of miraculous reconvergence worlds, which Lewis's ordering was meant to undermine. My example requires no miracles; if it matters, add chanciness at every stage of the relevant causal chains. I merely require the seemingly unassailable judgment that a business-as-usual world is more similar to actuality than a Doomsday-world. Of course, Lewis can argue that the Heads world is not *really* business-as-usual. There will be various traces of the coin landing Heads that are absent in the actual world, so at best the Heads world is only approximately similar to the actual world regarding particular fact; and approximate similarity of particular fact is "of little or no importance". But then he had better harden this line: approximate similarity of particular fact had better count for *nothing.* Otherwise, I will insist that the approximately-matching Heads world is at least *a little* more similar to actuality than the Doomsday world—offhand it sure seems that way—and since comparative similarity is all that matters to the truth-conditions of counterfactuals, that's good enough to secure the truth of 'Toss → Heads'. In other words, to respect my insistence that we must not judge 'Toss → Heads' to be true, we must judge the approximately-matching Heads world to be no more similar *at all* to actuality than the radically-non-matching Doomsday world. But surely that does violence to our intuitions about similarity. And once we do that, there is a danger of losing the intuitively correct verdicts about ordinary counterfactuals in any case. Recall that your imagined protest took it for granted that the nearest worlds to ours are ones in which billiard ball collisions or human jumps transpire much as we expect them to. If approximate similarity of particular fact counts for *nothing,* then all bets are off.

In any case, we can cut through much of this intuition-mongering about similarity of worlds by using the following recipe for generating counterexamples to at least some similarity-based accounts of counterfactuals. Let the proponent of such an account go first: tell us your ordering for similarity (much as Lewis did). I will then attempt to fashion a coin case accordingly: the scenario resulting from Heads is concocted to be something judged more similar *by that ordering* than the scenario resulting from Tails. The business-as-usual/Doomsday scenarios merely made vivid the recipe for a plausible-looking similarity judgment. Now, you may be able to thwart me by building criteria into your similarity ordering that prevent me from hooking up Heads to a more similar scenario, Tails to a less similar scenario. But you had better be careful that in doing so, you do little violence to our intuitions about *similarity* of *worlds* (or else admit that it is a 'similarity' account in name only). After all, the intuition that 'business *almost* as usual' is more similar than 'Doomsday' takes a lot of explaining away. And more importantly, you had better be careful that you deliver the intuitively correct results on ordinary counterfactuals, assuming that a desire to do so underlay your protest in the first place.

For example, you might say that all that matters to similarity is the *past* relative to the antecedent, and the coin's landing one way or another cannot affect that. (See Jackson 19xx.) It would be less misleading to call this a similarity of *pasts* account, rather than of total worlds. Then future similarity—indeed, even perfect match—counts for *nothing* for you. Whether business is almost as usual, or totally unusual, after the coin toss makes no difference on this view, because that's all *future* business. But then you would seem to have abandoned all resources for adjudicating counterfactuals with consequents that lie in the future relative to their antecedents. There's no saying what the billiard balls, or the jumping human, would do.[7]

[7] Or perhaps your similarity account is antecedent-relative, building in a dependence of the similarity relation on the antecedent itself (cf. Kmint 200x, Schaffer 200x). Then it would be a similarity of worlds*-relative-to-antecedents* account, rather than a similarity of worlds account per se. Let's not quibble about names, and let's

Or perhaps you claim that 'similarity' is a purely technical notion, a relation that simply induces an ordering on worlds suitable for the Stalnaker-Lewis style evaluation of counterfactuals. That flies in the face of the vast bulk of literature regarding 'similarity' accounts, which has been driven by intuitions about *similarity* and not something else, and which has apparently found them to be of heuristic value. If you are breaking free of that literature, again you would do better to give your relation a less misleading name, one with no prior associations—say, 'the R-relation'—in order to forestall misunderstandings. Then presumably we are not supposed to have intuitions one way or another about it, except perhaps by working backwards from the counterfactuals that we think are true to what the R-relation would have to be to deliver those verdicts. Clearly it would be question-begging to appeal to this R-relation to support your protest, when the truth-values of such counterfactuals are the nub of our disagreement. And your protest was plausible only to the extent that an intuitive notion of similarity was assumed—you insisted, remember, that the *most similar worlds* were ones in which bizarre things did not happen, and you *inferred* from this that the ordinary counterfactuals about billiard balls and jumps must be *true.* But if 'similarity' is some purely technical notion, again all bets are off. In any case, whether the relation is recognizably a 'similarity' relation or not, if it permits me to use my recipe, all is lost for an account of counterfactuals based on it. As long as I can fashion a case in which Heads yields a closer world, Tails a less close world *according to the ordering induced by R,* the account will predict that 'Toss → Heads' is true—an unacceptable result.

Our foot is in the door; now let's kick it open. I deliberately did not specify the chance of Heads. It is fine if you assumed that the coin was fair, but I did not. The argument would go through equally well, whatever the chance of Heads, as long as it is a possible outcome (and

avoid a lengthy digression into the pros and cons of this alternative to the usual Stalnaker-Lewis style semantics for counterfactuals. I merely want to stress that a quick invocation of that semantics in support of the imagined protest is surely *too* quick.

our supposition of chanciness assures us that it is). Now make the coin highly biased to Tails. Still, I will tailor my example to your similarity ordering, so that ‘Heads’ results in a more similar scenario by your lights. Then despite the bias to Tails, you will be committed to affirming that if we were to toss the coin, it would land Heads. Indeed, we can drive the chance of one counterfactual’s consequent all the way to 1 and *still* have you affirm the *other* counterfactual. We set up an infinite sequence of tosses, such that if there is *ever* a Tail, the resulting scenario is more dissimilar to actuality than if there is *never* a Tail. Then you must affirm that if we were to toss the coin infinitely many times, it would land Heads every time. Similarity theorists of counterfactuals: beware![8] This completes my reply to your Protest.

And so I reach the interim conclusion that *most ordinary counterfactuals are probably false*. There are three weasel words here: ‘*most*’, ‘*ordinary*’, and ‘*probably*’, and soon I will strip away two of them—the title of this paper, after all, is “Most Counterfactuals are False” without further qualification. *“Most”* remains, because soon I will concede that *some* counterfactuals are true. But it will turn out that they are typically *extraordinary*—rarified, recondite, recherché counterfactuals that philosophers may occasionally traffic in, but not normal people. In the meantime, I can do away with the qualifier ‘ordinary’. Time for some stripping.

3. Most counterfactuals are probably false

I have spoken of “*ordinary*” counterfactuals, but I don’t want to fuss much about *defining* when a counterfactual is ordinary. (Just try, if you think it’s easy!) Let me generously count as ordinary pretty much any counterfactual that you hear uttered on the street, or indeed outside a philosophical discussion. I’m being generous, because I’m prepared to count as ordinary many an arcane counterfactual from science—some biochemist’s counterfactual about a reaction

[8] My homage here to Lewis’s dust-jacket tribute to David Stove’s *The Plato Cult* is intentional.

rate, some astrophysicist's counterfactual about a galaxy red shifting, and so on, If you like, understand the claim that 'most counterfactuals are ordinary' so that it is an analytic truth (replace 'ordinary' by 'typical' if that helps).

Indeed, we will see in §6.3 that even large classes of *extraordinary* counterfactuals that appear to be true may not be so, thus swelling the ranks of the false counterfactuals still further. I will not attempt a census of just what proportion of all counterfactuals the true ones constitute—an impossible task—but I think that it will be *clear* that they are in the minority, and a small minority at that.

You may be tempted to say that there are uncountably many counterfactuals, of which uncountably many are true, in which case it makes no sense to speak of 'proportion', 'minority', and so on without some measure defined over sets of counterfactuals. When I quantify over counterfactuals, I don't mean all possible counterfactuals, the overwhelming majority of which could not even be asserted in a human lifetime because their antecedents and consequents are so complex. I mean instead the counterfactuals that one hears and reads in daily life. Imagine, if you like, a transcript of all counterfactuals ever uttered or written in the whole of human history, past, present and future. Needless to say, the set of all such counterfactuals is finite. And the vast majority of *them,* I am arguing, are probably false.

I say *"probably"* false, because so far I have assumed that the world that we live in is indeterministic. While I am not certain that it is, I think it is reasonable to be fairly confident that it is. My argument exploited possible ways in which things might turn out differently from what various 'would' counterfactuals claim, and so far I have appealed to indeterminism to guarantee that there are such possibilities. In a deterministic world, I risk losing that argument. But determinism is probably false.

No matter—determinism will restore truth to few of our counterfactuals in any case. Time for more stripping.

4. Counterfactuals under determinism: most counterfactuals are false

Even determinism will not save most counterfactuals from falsehood. A number of authors argue that determinism will not save the world from chanciness—"compatibilists" about chance such as Arntzenius (200x), Eagle (200x), Hoefer (forthcoming), Levi (19xx), and Loewer (2001). If they are right, it would seem that I do not lose my argument from chanciness to the falsehood of counterfactuals after all. But let's assume they are wrong, if only to make my job harder. I don't need that strategy for arguing for the falsehood of counterfactuals, because I have another one—the one that goes via considerations of *indeterminacy*.

For even if our world is deterministic, in the neighborhood of any trajectory of an object (of a billiard ball, or of a jumping human, or what have you) there will typically be some extraordinary trajectory in which things go awry. Here I appeal to statistical mechanics, whose underpinnings are deterministic. The point is familiar from the diffusion of gases, made vivid by Maxwell's demon. (Much as it is fair game for the epistemologist to remind us of the evil demon, it is fair game for me to remind us of Maxwell's demon.) For every set of initial conditions in which the air molecules in my office remain nicely and life-sustainingly spread throughout the room, there is a nearby initial condition in which they deterministically move to a tiny region in one corner—'nearby' as determined by a natural metric on the relevant phase space. The point generalizes to other deterministic systems. For every set of initial conditions in which the cue ball hits the 8 ball and each follows an expected trajectory, there is a nearby initial condition in which the balls behave anomalously. For every set of initial conditions in which I jump and land normally, there is a nearby initial condition in which I vaporize.

Now I exploit not indeterminism, as I did previously, but rather *indeterminacy*—in particular, the sort of indeterminacy that is due to *underspecification*. The antecedent of “if I were to jump, I would come down” is imprecise: I have not told you anything about the manner in which my hypothetical jump takes place, let alone given you a molecule-by-molecule specification of the jump. The antecedent, then, covers a huge range of initial conditions, each of which results in my jumping. Among them will be initial conditions that give rise to anomalous trajectories in which I vaporize, for the antecedent is too imprecise to rule them out. To be sure, the anomalous trajectories are sparse among all possible trajectories. But they exist all the same, compatible with the vaguely specified conditions given in the antecedent. So I If I were to jump, I might wind up on one of those anomalous trajectories. Thus, it is false to say that if I were to jump, I would come down. I might not.

Now you may want to lodge your second protest, confrontational as you are.

> SECOND PROTEST: Granted, these anomalous trajectories are compatible with the antecedent, but the *nearest* worlds in which the hypothetical jump takes place are ones in which you come down. So it is true after all that you *would* come down were you to jump.

This is similar to your previous protest, and my reply is similar; it’s just that now it is statistical mechanics rather than quantum mechanics that you are taking on. For you are denying that things *might* go anomalously in the ways I have imagined, while one of the most interesting features of statistical mechanics is to assert just that. (Again, we will shortly revisit this point if you are not already convinced of it.)

Or perhaps you are merely driving my argument in reverse, *tollensing* where I *ponensed.* You take as your starting point the apparent platitude that if I were to jump, I *would* come down. I am assuming you to agree with me that ‘would’ and ‘might not’ counterfactuals are contraries (we will drop this assumption in §§5.3 and 5.4). So you deny that if I were to jump, I *might not* come down. To be sure, they say that "one person's modus ponens is another person's modus tollens"; but all things being equal, I think it is sound counsel to side with whoever has physics on their side. And I claim that's me.

As before, often less esoteric facts will do the job of securing the truth of the 'might not' counterfactuals. If were to jump, a huge gust of wind might lift me higher. Statistical mechanics is just a handy, cover-all way for me to secure the truth of a huge raft of undermining 'might' counterfactuals in one fell swoop. But other anomalous happenings could do the job just as well on a case-by-case basis.

Yet I am still being too concessive to your Second Protest. As in your First Protest, you have insisted that my mid-jump vaporization takes us to *less similar* worlds than worlds where things go as expected. Again, let me grant that for the sake of the argument. Again, it is still a further step for you to reach the conclusion that the counterfactual 'if I were to jump, I would come down' is *true.*

For now consider the problems that imprecision or underspecification create for at least some similarity-based semantics for counterfactuals. Consider the counterfactual 'if I were at least 7 feet tall, I would be *precisely* 7 feet tall (precise to infinitely many decimal places)'.[9] I hope you agree with me that this is *false:* it seems absurd to affirm a counterfactual with such an imprecisely specified antecedent, and yet such a precisely specified consequent. Or perhaps you think that it is *indeterminate*, but you still agree with me that it is *not true.* Yet Lewis for one is apparently committed to it being *true.*

This example evokes Lewis's (1973) famous argument against Stalnaker's *limit* assumption[10]: that for any possible antecedent *X*, there is at least one nearest *X*-world. Lewis

[9] Igal Kvart tells me that he has similar examples in his (1986). [I must follow up on this.]

[10] And also an argument by Pollock (1976) against Lewis's denial of that assumption.

challenged this by considering counterfactuals of the form ‘if I were over 7 feet tall, then …’ What are the nearest worlds where the antecedent is realized? Try to pick such a world—say, one in which I am 7 feet 1 inch. Surely a world in which I am 7 feet ½ inch is closer to actuality—after all, in that world I am closer to my actual height. But a world in which I am 7 feet ¼ inch is closer still to actuality; and so on, ad infinitum.

Fair enough; but with the tiny tweak that I have given it, the example backfires on Lewis. Changing the antecedent from ‘I am over 7 feet tall’ to ‘I am *at least* 7 feet tall’ gives Lewis no wiggle-room—he is apparently committed to the most similar such worlds being those in which I am *exactly* 7 feet tall. After all, those are the ‘I am at least 7 feet tall’-worlds in which I am closest to my actual height.[11] He is *apparently* so committed—perhaps his example is only supposed to illustrate the possibility of a kind of structure that is problematic for Stalnaker’s theory, and we shouldn’t take this particular example too seriously. In that case it would be nice to see an example that we *should* take seriously, so that we are convinced that Lewis’s concern is not merely a theoretical possibility, that it actually arises for counterfactuals that we might assert or believe. I wager that I could rewrite my objection, mutatis mutandis, using any such example.

This suggests the following recipe for generating counterexamples to at least some similarity-based accounts of counterfactuals. Let the proponent of such an account go first: tell us your ordering for similarity. I will then attempt to fashion a counterexample accordingly: a counterfactual with a highly imprecise antecedent, and a comparatively precise consequent, such that all the closest worlds *by that ordering* that realize the antecedent realize the

[11] You may wonder whether atomism or genetics may cast some doubt on this. Maybe my exceeding 7 ft tall by sub-atomic distances does not detract from similarity; moreover, maybe facts about genetics could make my nomically possible heights ‘granular’, with some slight overshooting of 7 ft possible, but any smaller overshooting not. Of course, these would also be objections to Lewis’s original argument against Stalnaker. In any case, we may circumvent them with a minor revision to the example. Consider a line in an atomless world that is 6 ft long, and now entertain counterfactuals about what would be the case if the line were at least 7 ft long.

consequent. The danger is that you will then be committed to affirming the counterfactual, whereas we should baulk at its highly unspecific antecedent coupled with a comparatively specific consequent. Once again, similarity theorists of counterfactuals: beware! This completes my reply to your Second Protest.

So not even determinism suffices to save counterfactuals from falsehood. I said that determinism is probably false, but it turns out it doesn't matter either way. Whichever way our world goes, *most counterfactuals are false*.

*   *   *   *   *   *   *

Philosophy, like tight-rope walking, is a risky business.[12] And a philosopher, like a tight-rope walker, is well advised to have safety nets in place—especially when arguing for a position that is likely to be controversial, as I am. So in the next few sections I want to offer a series of fall-back positions. I have given you arguments for a radical position, but if you are unconvinced by them, you may still be convinced by close relatives of them. These relatives have weaker conclusions, but I think they are still radical enough to be interesting.

5. Fall-back positions

5.1 The counterfactuals are indeterminate rather than false?

You may say that in indeterministic cases such as I have discussed, and in deterministic cases with imprecise antecedents, there is no determinate fact of the matter of what *would* happen. You may insist, then, that these counterfactuals are not false, but *indeterminate*. For example, starting with the coin that will never be tossed, 'Toss → Heads' is neither true nor false, where I too quickly concluded that it was false. And when I took you down the slippery slope from there, through biased coins and lotteries to billiard balls, jumps and cups, at each point you may judge the counterfactuals to be neither true nor false. Or so you say.

[12] Compare the opening words of Lewis (1969).

I reply: This still yields a striking conclusion, namely, most counterfactuals are *not true*. Striking, because it still undermines much of what we say, apparently *taking* such counterfactuals to be true. Try telling people on the street that ‘If I were to jump, I would come down’ is *not true*, and see what looks you get!

In fact, Edgington is also a no truth value theorist about counterfactuals – she thinks that like indicatives, they are governed by a version of the Adams Thesis that assertability goes by the corresponding conditional probability. One wonders what the probability of a counterfactual *is,* if not its probability of truth. This ‘no truth value’ account is surely not true across the board. What about counterfactuals of the form p → p, and more generally, those whose antecedents entail their consequents? In section 6 I will concede that various kinds of counterfactuals are true; I wonder how Edgington and co could deny that.

There’s also the familiar Frege-Geach problem of how counterfactuals are meant to embed in various contexts, such as in Boolean combinations, or modal contexts, or how they iterate. Counterfactuals can appear in arguments that offhand seem to be valid or invalid, and I don’t just mean some kind of watered down ‘probabilistic validity’. And if Edgington and co want to analyze other things in terms of counterfactuals, as philosophers so often do, what are they going to say about those other things—causation, knowledge, perception, personal identity, laws of nature, dispositions, and what have you? That claims involving these things don’t have truth value either? I’ll return to this point at the very end.

Striking, but not striking enough if we subscribe to the would/might duality. By that duality, ‘Toss → Heads’ is equivalent to ‘¬(Toss ◊→ ¬Heads)’. If the former is indeterminate, then so is the latter; if the latter is indeterminate, then so is its negation, ‘Toss ◊→ ¬Heads’. But that might-counterfactual is not indeterminate, but unabashedly true. I subscribe to the would/might duality, and I am in good company, so I stand by my claim that most counterfactuals are *false.* But if I'm wrong about that, the fall-back position that most counterfactuals are indeterminate is disturbing enough.

5.2 The counterfactuals' truth values are context-dependent rather than false?

You may say that the counterfactuals that I have countenanced are not uniformly false, but rather have *context-dependent* truth-value. For example, in normal conversational contexts, it is true to say 'if I were to let go of the cup, it would break'. And if I press you regarding the undermining 'might' counterfactual 'if I were to let go of the cup, it might not break', you agree that it has opposite truth value (we will have you disagree with this soon enough): in normal conversational contexts it is false. If I then press you with quantum mechanical and other anomalous possibilities, you agree that the 'might' counterfactual becomes true, but only relative to a changed context. So the truth-value of 'if I were to let go of the cup, it would break' is context-sensitive. But since most contexts are normal ones, we can happily say that in most contexts it is true. Or so you may say.

In another version of this 'counterfactuals are context dependent' objection, you may allow both the 'would' and the 'might' counterfactuals to go *indeterminate* in some contexts, as well as true in others and false in still others. Still, you stick to your previous guns that in normal conversational contexts, 'if I were to let go of the cup, it might not break' is false—and so on.

I reply: The oft-heard slogan that ‘counterfactuals are context-dependent’ glides nicely off the tongue, the way that slogans often do. Moreover, it’s meant to have consequences for our various practices—for example, van Fraassen (1980) thinks that it has consequences for science. But what does the slogan mean exactly? Is it that *the truth values of all* counterfactuals are context-dependent? That’s clearly false: consider any counterfactual of the form p →p, and more generally any of the counterfactuals whose context-*independent* truth I will concede in §6. Is it that *some* counterfactuals have context-dependent truth value? That’s obvious: “if I were you, I’d lose that tie” depends on the contextually-sensitive terms ‘I’, ‘you’, and ‘that tie’.

But let’s set aside the context-dependence of counterfactuals that is parasitic on context-dependence of something else (indexicals, epistemic modals, knowledge ascriptions, or what have you). To be of interest, and of potential trouble for me, the thesis must be something like “*most* counterfactuals have context-dependent truth value”—perhaps, most counterfactuals on my imagined transcript of all counterfactuals uttered in human history are context-dependent? I haven’t seen a careful argument for that. I suppose the usual argument, to the extent that there is one, is roughly this: Quine’s famous ‘Caesar’ example is meant to be typical: “If Caesar had invaded Korea, he would have used catapults/nuclear weapons”—context-dependent! And this counterfactual is supposedly context-dependent because in one context we may focus on Caesar’s belligerent tendency to use the most powerful weapons at his disposal (which may include nukes), and in another context we may focus on the historical facts about his weapons (which exclude nukes).

I have several replies to this argument. Firstly, it is not clear that examples like the Caesar counterfactuals *are* typical. The even more famous “If Oswald had not shot Kennedy, then somebody else would have” is apparently straightforwardly *false*, independent of context.[13] Counterfactuals like this seem no less typical to me. Likewise, my considerably less famous ‘If the coin were tossed, it would land Heads’ surely has a context-independent truth-value (false, as I keep urging). Or again, ‘If I were at least 7 ft tall, I would be exactly 7 ft tall’ seems false, irrespective of context. So I’m not yet convinced that Quine’s example is so typical. To be sure, a single example suffices to show that a complete semantic analysis for counterfactuals will need a contextual parameter, and that example may as well be Quine’s; it’s just that the parameter may be idle for many counterfactuals. But then the context-dependence of counterfactuals does not yet pose a threat to my conclusion that most counterfactuals are false; there needs to be a further premise about how typically the parameter is activated.

[13] If you are a conspiracy theorist about the Kennedy assassination, append the words “at exactly the same instant”, and you should agree that the resulting counterfactual is false, independent of context.

Secondly, I agree that our similarity judgments about worlds may often be context-dependent, since what we hold fixed in those judgments may be context-dependent. But I have questioned similarity-based accounts of counterfactuals. Thirdly, it does not follow from the context-dependence of similarity judgments that the *truth value* of the 'Caesar' counterfactual is context-dependent—for it may well be that it comes out false *whichever* way we contextually resolve our similarity judgments. Indeed, that is exactly what we should expect, given my arguments. Focus on Caesar's belligerent tendency if you like; still, he *might not* have used nuclear weapons if he had invaded Korea (he might have used chemical weapons instead). Focus on the historical facts if you like; still, he *might not* have used catapults if he had invaded Korea (he might have used slingshots instead). So the corresponding 'would' counterfactuals come out false either way—false in *different* ways, but false nevertheless. So much for context-dependence!

More generally, even granting that our similarity judgments about worlds may be context-dependent, it does not follow that the truth-values of the everyday counterfactuals that we utter are context-dependent. They may well be stable across context shifts—and my position predicts that most of them are stably false. Compare: while our judgments of 'tallness' are context dependent, the truth value of 'all tall people have two heads' is stable across context shifts—stably false.

And what of the alleged context-dependence of the *truth value* of ‘if I were to let go of the cup, it would break’—true in ordinary contexts, you say, and false (only) in extraordinary contexts in which quantum mechanical or other anomalous possibilities are salient? Reflect on why you think it is false in the latter contexts. A good reason, I submit, is that the *chanciness* of the cup’s breaking is made salient in these contexts, and it undermines the claim that it *would* break, if it were let go. But the chanciness was there all along, before we made it salient or attended to it. Compare: Suppose a lover of the cup is poised all along to try to save it in case it is ever dropped; she will probably fail, but there is some chance that she will succeed. Then it is immaterial whether you happen to be attending to her or not—irrespective of whether this possibility is salient or not to you, and whether you realize it or not, it is simply false that if I were to let go of the cup, it would break. The chancy quantum mechanical and other anomalous possibilities are other ways the cup might be saved, whether you attend to them or not. But if you think that chanciness does *not* undermine the counterfactual, then why do you feel any inclination to retract the counterfactual when we move to a context in which the chanciness is salient? Either a given counterfactual is compatible with the chanciness of its consequent, or it is not. If it is, then you are wrong to retract it after attending to that chanciness. If it is not, as I claim it is not, then you are wrong to endorse it before attending to that chanciness.

But suppose for the sake of the argument that for a broad class of counterfactuals, their truth-values really are context-dependent. (It had better be broad to threaten my thesis.) We might follow Lewis’s account of knowledge (1996), for example, in contending that context determines which possibilities are and are not properly ignored. In a typical conversation about cup dropping, we may properly ignore bizarre quantum mechanical possibilities; in a typical conversation about jumping, we may properly ignore bizarre statistical mechanical possibilities. But once I draw your attention to them, you cannot ignore them (for a while, anyway), let alone properly ignore them. And as long as they are live, they will underwrite various outlandish 'might not' counterfactuals that oppose corresponding mundane 'would' counterfactuals. This has the unintuitive consequence that a good way to make your counterfactual assertions come out true is to make sure that you and your interlocutors are inattentive, ignorant, or unimaginative, and the more so, the better.[14]

The flip side of this is that counterfactuals are easily *made* false: as easy as context-shifting is. An extreme case of this, also reminiscent of Lewis on knowledge, is that philosophical contexts will be especially liable to make counterfactuals false, for in such contexts bizarre possibilities are fair game—bad news for those philosophers *in* such contexts who want to use counterfactuals to analyze other concepts. We will return to this point at the very end. And never mind conceptual analysis; we had better be careful, when asserting that counterfactuals are *entailed* by other things (e.g. the laws of nature), that we either embrace the context-dependence of these entailers, or else explain how their context-independence fails to transmit to their entailments.

Furthermore, a theory that portrays counterfactuals as being highly context-dependent risks not doing justice to the phenomenon of inescapable clashes. Even if you do not agree with me that the clash between 'would' and corresponding 'might not' counterfactuals is semantic, I hope you still agree that there is a clash. (The next section will explore another hypothesis about the nature of the clash.) There is not even a ghost of tension between my saying "It is now 4:07", and an hour later saying "It is not now 4:07"—"now" is obviously highly context-sensitive. If counterfactuals are too context-dependent, then there should not be even a ghost of tension between assertions of 'woulds' and corresponding 'might-nots' (say, uttered an

[14] Compare Elgin (1988).

hour later)—context could see to it that they both come out true. But there *is* tension, and more than a ghost of it.

[EASILY ANSWERED?: CONTEXT IS STABLE ENOUGH IN THE ONE UTTERANCE. (INSTEAD OF MAKING IT AN HOUR LATER, MAKE IT A SECOND LATER.) BUT ISN'T THE POINT: EVEN AN HOUR LATER, THE 'MIGHT NOT' CLASHES WITH THE 'WOULD' ASSERTED NOW. THAT'S PLENTY OF TIME FOR CONTEXT TO CHANGE, SO THAT THE CONTEXTUALIST CAN ALLOW BOTH TO BE FELICITOUS (WITHOUT THE RELEVANT UNDERLYING FACTS CHANGING).]

WORSE: THERE SHOULD NOT BE A GHOST OF TENSION BETWEEN THE UTTERANCE OF A 'WOULD' AT ONE TIME, AND ITS *NEGATION* AT ANOTHER.]

Suppose that the truth-maker for some claim we hold dear is not merely a counterfactual in isolation, but a complex pattern of counterfactuals. The contextualist had better hope that they *all* come out true in the relevant context. If they are too unstable, too sensitive to context, and the context changes sufficiently from one to another, then the dearly held claim may be undermined. For example, suppose that causation involves chains of counterfactual dependences (à la Lewis 1973), and consider a case in which the chain is long: *E* is counterfactually dependent on C_n, which is counterfactually dependent on C_{n-1}, which is …, which is counterfactually dependent on C_2, which is counterfactually dependent on a salient cause C_1. If the individual counterfactuals are too fickle, there may be no single context (let alone the relevant context) in which they all come out true. Then there is no single context (let alone the relevant context) in which the claim that C_1 causes E comes out true. This will have ramifications in turn for concepts that depend in turn on causation—say, moral responsibility. Indeed, the problem is writ large when a concept dear to us involves an entire network of causal chains—think of mental states on a functionalist analysis. The point generalizes beyond these illustrations: one should not be glib about the putative context-dependence of counterfactuals while being sanguine that complex patterns of counterfactuals, across which contexts may vary significantly, underwrite important claims of ours.

Finally, even if we grant that most counterfactuals have context-dependent truth values, that it still quite compatible with my thesis that most counterfactuals are false. Recall my imagined transcript of all counterfactuals ever uttered in human history. It is quite compatible with contextualism that most of them were uttered *in contexts in which they were false*. Indeed, it is compatible with contextualism that *all* counterfactuals, as a matter of fact, are false in their context of utterance.

At this point the contextualist may want to invoke some version of a principle of charity—roughly, a constraint on interpreting one another's utterances so that most of them come out true, and where contextual factors make a difference, they are resolved with a presumption in favor of the truth of what is said. As a semantic thesis, I find the principle of charity too good to be true, and I don't know a charitable way of understanding it that will save it. It surely diminishes the achievement of *getting things right*. Think of how hard this can be in daily life, let alone in science. We should not be credited with too many true utterances, and we should get more credit for the true utterances that we do make. The principle also risks making disagreement harder to come by than it should be—indeed, it may be *uncharitable* to regard context shifts as ensuring that apparently disagreeing parties are each speaking truly, especially when they take their disagreement to be genuine. As a pragmatic strategy or heuristic for felicitous interactions with one another, something like the principle of charity has more going for it, although then it poses no threat to my semantic thesis about the counterfactuals' truth values. And even then the principle seems a little off the mark. More plausible is something more like the principle of *humanity*, which will make more room for false beliefs in the face of misleading or incomplete evidence, as is all too common.

In any case, I am not convinced that attributing falsehoods to each other need be uncharitable or inhumane. Falsehoods may serve useful purposes, and they may be assertable despite being false. Listen carefully to a typical conversation. You should be struck by how much of what is said is literally false, but understandably so because of other purposes that are served—think of exaggerations, jokes, short-cuts, approximations, paraphrasing, idioms, irony, metaphor, politeness, and so on. Arguably, even much of science is literally false, but close enough to true about enough of the propositions that we care about, to do the job we require of it. There should hardly be some special presumption in *favor* of the truth of counterfactuals that we utter, given so much falsehood in our utterances across the board! And we seem to be especially bad at other forms of modal reasoning—witness the literature started by Kahneman and Tversky on how bad people are at probabilistic reasoning. Why think that we're especially *good* at counterfactual reasoning? Still, I do think that counterfactuals fare especially poorly even compared to our generally low strike rate of truths in what we say, even the modally charged things we say. That should hardly come as a surprise, given my arguments. After all, the pronouncements of quantum mechanics and of statistical mechanics come as a shock. Quantum mechanical or entropy-decreasing possibilities fly in the face of folk wisdom, so we should hardly expect folk intuitions to be a reliable guide to the counterfactuals that they undermine. That said, in section 8 I will explain how our false utterances of counterfactuals can be vindicated by truths that approximate them. In the meantime, I suggest that if most of our counterfactual utterances are false, they are in surprisingly good company.

So even granting the context dependence of the truth-values of counterfactuals, I have as fall-back positions: *most counterfactuals are easily made false*, they are *especially easily made false in philosophical contexts*, *important claims of ours that are underwritten by complex patterns of counterfactuals risk being false*, and *most counterfactuals may be uttered in contexts in which they are false.* These positions are still somewhat unsettling.

Our judgments of counterfactuals are also relevant to our actions—so much so that some versions of causal decision theory explicitly write such judgments into their account of rational action. If the truth value of a given counterfactual is context-dependent, then presumably so is its probability of truth (as assigned by a rational agent). Then it would seem that by switching context, what it is rational to do can switch. Sometimes this should come as no surprise. But often it will be quite odd: a sudden change in the presuppositions of a conversation could change what you ought to *do.*

[It's unclear what a context even is. Suppose we suddenly take seriously the possibility of the cup quantum tunneling; allegedly, at that point the cup-dropping counterfactual suddenly goes false. Now suppose I say 'if I were to jump, I would come down'—is that false too? It's not obvious that in raising to salience the *cup* quantum tunneling possibility, I thereby raised to salience *quantum tunnelings of all objects.* Perhaps instead I raised to salience *weird things that could happen to the cup.* In that case, I may have created a context in which the cup suddenly being destroyed by an evil demon is a live possibility, but not the quantum tunneling of a jumping human. So the problem for the counterfactual contextualist is: exactly which new possibilities did the new context that I created admit? Absent a good answer to that question, there is no telling how the truth values of the counterfactuals vary with context.

People like to say that I prove too much: I would have most future contingent claims come out false. I think I can run this against contextualists: they'll think that future contingent claims are context-dependent. But they're not. The future doesn't care what you happen to presuppose, or what happens to be salient to you. Neither do counterfactuals.

According to one sort of contextualist, the following is a necessary truth, 'for all p: if p were the case, then only serious possibilities would obtain.' I think that's an unhappy result—if p were the case, some completely unforeseen possibility might obtain.]

5.3 The 'might not'/'would' clash is merely pragmatic?

My argument for the falsehood of most counterfactuals began in §2 with the collision between a 'would' and an opposed 'might not'

'if I were to toss the coin, it would land Heads' and 'if I were to toss the coin, it might land Tails'. I argued that they are contraries, so that the collision is logical. I asked you to try saying:

"If I were to toss the coin, it MIGHT land Tails, and it WOULD land Heads if I were to toss the coin" (*)

and to see whether you think you've said something consistent. You may agree that (*) sounds odd, but deny that the oddness is logical. It may instead be some kind of *pragmatic*, rather than semantic, inconsistency—the oddness that apparently underlies Moore paradoxical sentences such as 'It's raining and I don't believe it's raining'. In that case, both Toss ◊→ Tails and Toss → Heads could both be true. Then I lose my argument from the truth of the second conjunct to the falsehood of the first. And so it goes for the other counterfactuals that I went on to discuss. For example, 'Drop cup → Break' and 'Drop ◊→ ¬Break' are semantically consistent; they are merely pragmatically inconsistent. Or so you may say.

More specifically, de Rose (19xx), following Stalnaker (19xx), argues that 'if it were the case that p, it might be that q', should be analyzed as:

$\Diamond_e(p \rightarrow q)$

where '$\Diamond_e$' symbolizes 'it is epistemically possible that'. De Rose then appeals to the pragmatics of assertion…

First, notice that this renders the 'might not' counterfactual consistent with the corresponding 'would' counterfactual. Second, de Rose argues that "you represent yourself as knowing a fact if you flat-out assert it" (388). So if you flat-out assert '$p \rightarrow q$', then you are representing yourself as knowing $p \rightarrow q$. But by the above analysis, 'if it were the case that p, it might be that ¬q' means $\Diamond_e(p \rightarrow \neg q)$—"an admission of the epistemic possibility that the assertion [of '$p \rightarrow q$'] is false" (390). Conjoining the 'might not' and the 'would' counterfactuals (as in (*)), there is a genuine semantic inconsistency between what you actually say with the 'might not' counterfactuals, and how you represent yourself with respect to the 'would' counterfactuals. We thus have an ingenious explanation of inescapable clashes. They involve the same kind of pragmatic inconsistency as we find in assertions of the form 'X, and it might be that not-X', as in "It's raining, and it's possible that it is not raining"—semantically consistent, but representing oneself in a way that is inconsistent.

The Stalnaker/de Rose is by no means the only possible pragmatic account of the might-not/would clash, but it is particularly carefully worked out, and it will serve as something of a case study.

I reply:

First, some general worries about pragmatic approaches, and then some specific objections to Stalnaker/de Rose analysis.

Presumably, we would like to explain inescapable clashes even when they are not flat-out asserted—for example, when they are merely thought, or supposed, or mentioned, or are items in a data base that we are trying to assimilate. We would also like to explain them when the 'might not' counterfactual is explicitly *metaphysical,* rather than epistemic. In the cases I have imagined, that reading is available, and arguably even primary. The inferences from instances of indeterminism or of indeterminacy to metaphysical 'might nots' strike me as secure, and it is unclear to me what bearing pragmatic considerations could have on them. But if they are not secure, then let's cut out the middle-men, bypassing the 'might nots' altogether. After all, I maintain, indeterminism and indeterminacy undermine the relevant 'would' counterfactuals directly. I found it useful to appeal to the 'might nots', but I could have run my arguments without them. And however the pragmatic inconsistency of the clashes is cashed out, the fact that they are *clashes* that are pragmatically *inconsistent* tells us that there is still something seriously defective with asserting both the 'might not' counterfactuals and the corresponding 'woulds'. Yet I claim that there is nothing wrong with asserting the 'might nots'; so still, the 'woulds' must take the fall.

Turning now to the specifics of the de Rose/Stalnaker analysis: For starters, it does not do justice to the surface grammar of the 'might counterfactuals', in which the 'might' has narrow scope. To be sure, this counts for little; still, it's a point against, rather than a point for, the analysis. More importantly, it founders on our cases of indeterminism and indeterminacy. 'If the coin were tossed, it *might* land Tails' does not say the same thing as 'It is epistemically possible that: if the coin were tossed, it *would* land Tails', for the former is true, and the latter is surely false according to anyone who knows that the coin toss would be an indeterministic process. The 'would' claim is incompatible with their knowledge of the chanciness of the coin. It is the very nature of chancy processes that there is no fact of the matter of how they *would* eventuate.

Similarly, ‘If I were at least 7 ft tall, I *might* be 7 ft 1 inch tall’ can reasonably be regarded as true, while ‘It is epistemically possible that: If I were at least 7 ft tall, I *would* be 7 ft 1 inch tall’ can reasonably be regarded as false. Given the indeterminacy regarding my hypothetical height, I may know that 7 ft 1 is a live option; but I know that it is *not* the only live option, so I know that the ‘would’ counterfactual is too committal to be true.

But let’s suppose for the sake of the argument that deRose’s analysis of ‘might’ counterfactuals is correct. This does not yet show that the counterfactuals that we ordinarily take to be true are in fact so. For it does not follow from the fact that a given ‘might not’ counterfactual is true, and that the corresponding ‘would’ counterfactual is compatible with it (despite clashing with it pragmatically), that the ‘would’ counterfactual is true. (Compare: it is true that grass is green, and compatible with this that snow is black; it does not follow that it is true that snow is black.) We may grant that there is a possible world in which both the ‘might not’ and ‘would’ counterfactuals are true, without granting that the *actual* world is such a world. At best, deRose has shown that our pre-theoretical belief that various ‘would’ counterfactuals are true is *tenable*, not that it is *correct.*

What about deRose’s explanation of the pragmatic inconsistency? Arguably, one represents oneself merely as *believing* something when one asserts it, rather than knowing it. When one really does want to represent oneself as knowing something, explicitly saying so is not redundant: "The pub closes at midnight. Trust me: I *know* that it does." And it's not the difference of now representing oneself as knowing that one knows - that's too fancy for the folk, who often speak this way (even outside pubs).

Finally, and most importantly, the de Rose proposal is still open to a kind of counterfactual skepticism, one that I contend is rather more puzzling than the one that I am defending. Whenever a ‘might not’ counterfactual is true, the corresponding ‘would’ counterfactual cannot be *known,* even if it is in fact true. (Cf. Eagle 2007.) But then the 'would' counterfactual cannot reasonably be asserted either, for in doing so one would *misrepresent* oneself as knowing it. So we quickly reach the startling conclusion that *most counterfactuals are unassertable.* But this flies in the face of our linguistic practices, which are a surer guide to assertability than they are to truth. My skepticism about the truth of counterfactuals seems tame by comparison! By my lights, deRose has things exactly back to front: while I regard most counterfactuals that we utter to be false but assertable, deRose must regard most of them as unassertable, even though he thinks that they are true. He is throwing the baby out with the bath water. [THIS NEEDS TO BE MORE NUANCED. DEROSE CAN GO CONTEXTUALIST ON ASSERTABILITY, AND INSIST THAT IT'S ONLY IN PECULIAR CONTEXTS THAT THE COUNTERFACTUALS ARE UNASSERTABLE.]

5.4 'Would' and 'might-not' counterfactuals are semantically and pragmatically consistent?

We come to my final fall-back position. You may insist that there is no inconsistency whatsoever between the 'would' counterfactuals and the corresponding 'might not' counterfactuals—neither semantic, nor pragmatic. You grant me the truth of the 'might not' counterfactuals, but deny me that falsehood of the corresponding 'would' counterfactuals follows. Rather, you say, they can happily co-exist with each other, not even in Moorean tension. For example, 'if I were to jump, I would come down' is true (as commonsense would have it), *and* 'if I were to jump, I might not come down' is also true (as physics would have it). Or so you say.

I reply: I hope that you will still agree that there is *some* tension between these sentences. If not, our linguistic intuitions are at such odds that I'm not sure how we could move beyond an impasse. And if we do agree at least that much, then I offer you the challenge of doing justice to the tension that even you hear in (*), and other examples of inescapable clashes. For instance, if you think that $p \rightarrow q$ involves the nearest p-worlds, whereas $p \Diamond\!\!\rightarrow \neg q$ involves casting our net further afield to include more distant p-worlds, then it is unclear why there should be any tension *at all* between them.

But if there is really is no tension at all between them, their standing alongside each other in perfect harmony, then we still reach an interesting result, my final safety net. In that case, we still have an argument against the Lewisian (1973) semantics for counterfactuals, according to which the ‘would’ and ‘might not’ counterfactuals are in such tension that they are contradictories. And we have an argument against any proposal for there being a pragmatic inconsistency between assertions of them. These are not my arguments, but they are still rather interesting.

[MAYBE INCORPORATE SOME OF THE FOLLOWING:] But perhaps I can try to pump my intuitions in successive stages. Start with an innocent enough sentence, say:

“The Godfather could break his promise, but he will not break his promise”.

That’s fine. The first conjunct attributes a modal property to the Godfather, a potentiality of his; the second conjunct is a prediction that he will not realize it. Next, consider:

“The Godfather might break his promise; but he will not break his promise.”

That’s less fine. The speaker indicates with the first conjunct that she regards the Godfather’s breaking his promise to be a serious possibility; but this undermines her right to her second-conjunct prediction that he will not. Next, consider:

“If the Godfather were to make you a promise, he might break it; and if he were to make you a promise, he would not break it.”

That’s even less fine. Now, the two conjuncts are making conflicting modal claims about him. ‘Tension’ surely puts it mildly.

This, in turn, implies that his analysis of them is mistaken:

$X \rightarrow Y$ is true at world w iff there is an $X \& Y$ world nearer to w than any $X \& \neg Y$ world.

$X \Diamond\!\!\rightarrow \neg Y$ is true at world w iff there is no $X \& Y$ world nearer to w than any $X \& \neg Y$ world.

Clearly the second analysans is the negation of the first, and thus we have Lewis's duality of the 'would' and 'might' counterfactuals.

An obvious way to make both $X \rightarrow Y$ and $X \Diamond\!\!\rightarrow \neg Y$ come out true, *contra* Lewis, is to use shifting standards for which worlds figure in their truth conditions. In particular, we may be allowed to cast our possible-worlds net further out when assessing 'mights' than when assessing 'woulds', and if it catches any $X \& \neg Y$ world, then $X \Diamond\!\!\rightarrow \neg Y$ is true. For example, casting our net indefinitely far out, we have:

(*Unrestricted 'might'*) $X \Diamond\!\!\rightarrow \neg Y$ is true at world *w* iff there is any $X \& \neg Y$ world (no matter how remote from w).

Since the analysans does not depend on *w*, this has the perhaps unwelcome consequence that 'might' counterfactuals are either prily true or necessarily false, and thus not contingent. It also reduces $X \Diamond\!\!\rightarrow \neg Y$ to $\Diamond(X \& \neg Y)$, which seems to do violence to the linguistic data. 'If the coin were tossed, it might not land heads' just doesn't seem to say thing as 'It's possible that the coin both is tossed and does not land heads'.

So perhaps we should not cast our net *indefinitely* far out for 'mights', but still further out than for 'woulds'. For example:

(*Suitably restricted 'might')* $X \Diamond\!\!\rightarrow \neg Y$ is true at world *w* iff there is any $X \& \neg Y$ world within some suitable distance of *w* (and in particular, greater than the distance appropriate for the assessment of $X \rightarrow Y$).

This restores contingency to the 'might' counterfactuals, but renders the analysans nebulous. And while it is at least principled to restrict one's attention to 'nearest X-worlds' (as Lewis does), or to impose no restriction at all (as *Unrestricted 'might'* does), one wonders how one could impose a principled intermediate restriction, looking at worlds further afield than for the assessment of 'would' counterfactuals, but not *too* far afield.

So to someone who thinks that both X → Y and X ◊→ ¬Y are true (for the various X and Y that have been my examples), I offer a challenge: Give a principled semantics of both kinds of counterfactuals that does justice to our intuitions about them, and that explains away the apparent tension between them. If the challenge can be met, then we have a reason to reject the Lewisian semantics for them. That's still an interesting conclusion. But I have given my reasons for thinking that it is not the right conclusion.

* * * * * *

That's almost as much of my long paper as I can present here. Let me just gesture at what happens next.

Much philosophy is an attempted demolition of commonsense followed by damage control. I think that it is an item of commonsense that various ordinary counterfactuals, such as 'if I were to jump, I would come down' are true. I have argued that commonsense is mistaken about them; indeed, they are not even indeterminate. So much for commonsense. Now it is time for some:

6. Damage control

There are limits to how startling my conclusion can legitimately be. I now concede that *some* counterfactuals are true; it remains to be seen how much of a concession this will amount to in the end.

Let us begin with counterfactuals that are not just true, but necessarily so, in virtue of logical, mathematical, analytic, metaphysical, or nomological truths.

6.1 Strict conditionals

6.1a. Necessary consequents

Counterfactuals with necessarily true consequents are trivially true (even if recognition of their truth may be non-trivial), where the 'necessity' at issue may be logical, analytic, mathematical, metaphysical, or nomological. Thus, I happily concede that the following counterfactuals are all true:

if the coin were to be tossed, it would be self-identical;

if the coin were to be tossed, all bachelors would be unmarried;

if the coin were to be tossed, nothing would be red and green all over;

if the coin were to be tossed, Fermat's last theorem would be true;

if the coin were to be tossed, water would be H_20;

if the coin were to be tossed, Hesperus would be Phosphorus;

if the coin were to be tossed, then nothing would travel faster than light;

and so on.

Counterfactuals whose consequents assert something true about the past may well go the same way:

if the coin were to be tossed (now), World War II would still have occurred,

and so on. They may well be just further instances of necessary consequents (given the actual historical facts): most philosophers agree that it is impossible to change the past.

6.1b. Impossible antecedents

Likewise, I am prepared to concede happily enough that counterfactuals with necessarily false antecedents are trivially true:

if the coin were not self-identical, then it would land heads;

if the coin were not self-identical, then it would not land heads;

and so on.

Now, I am not sure that I am forced to make this concession. Counterfactuals with necessarily false antecedents may be used non-trivially in *reductio ad absurdum* reasoning, and this surely requires that they are not all vacuously true. Furthermore, consider counter-metaphysicals: 'If there had been exactly 17 possible worlds, then Lewis's views on possible worlds would have been correct in every detail.' That hardly sounds true, since Lewis devoted much ink to arguing that there are infinitely many possible worlds.

But let us set these aside, for they will not affect my conclusion: if even such counterfactuals can be false, all the better for my case.

6.1c. Necessary connections between antecedents and consequents

Likewise, I happily concede that counterfactuals whose antecedents necessitate their consequents—either in virtue of logic, or analyticity, or metaphysical necessity, or mathematics, or nomological truth—are trivially true:

if I were in Australia, then I would be in Australia or Morocco;

if were not married, I would not have a wife;

if the moon were made of green cheese, the moon would not be red all over.

if I were 7 ft tall, I would be √49 ft tall;

if Renée were drinking water, then Renée would be drinking H20;

and so on.

Each such counterfactual is of the form $p \rightarrow q$, where $(p \supset q)$ is true—the 'box' being an appropriate form of necessity. In other words, what underwrites these counterfactuals are corresponding strict conditionals. And while counterfactuals are sometimes called 'strong' conditionals (see e.g. Lewis 1980), strict conditionals are stronger still.

Really, *all* necessarily true counterfactuals are species of those whose antecedents necessitate their consequents, so we may regarded this category as subsuming the others. For if q, then for any p, $(p \supset q)$. And if $\neg\Diamond(p)$, then for any q, $\neg\Diamond(p \,\&\, \neg q)$, which is equivalent to $\neg\Diamond\neg(\neg p \ v \ q)$, and thus to $(p \supset q)$. (This in turn provides an argument for regarding counterfactuals with impossible antecedents as true after all: $(p \supset q)$ is stronger than $p \rightarrow q$, and since the former is true when p is impossible, so is the latter. So I have now given arguments for both verdicts regarding such counterfactuals.) We can then characterize *all* of the true counterfactuals that we have so far identified as ones in which *the antecedent necessitates the consequent*. It remains to be seen whether there will be any other true counterfactuals to add to the list.

It should come as no surprise that all these trivially true counterfactuals resist my argument. For clearly the might-counterfactuals that are contrary to them are false. 'Might' counterfactuals with impossible consequents are trivially false:

if I were to toss the coin, it might not be self-identical,

and so on.

It is less obvious that 'might' counterfactuals with impossible antecedents are trivially false:

if the coin were not self-identical, then it might land heads,

and so on. For it is a little odd that the 'would' counterfactual with the same antecedent and consequent is true, while the seemingly weaker 'might' counterfactual is false. Recall that I was unsure about my concession to the truth (let alone necessary truth) of ‘would’ counterfactuals with impossible antecedents. So if ‘might’-counterfactuals like this one come out to be true, all the better for me.

Again obviously, might-counterfactuals involving the failure of a necessary connection are trivially false:

if I were in Australia, then I might not be in Australia or Morocco,

and so on. In general, if (p ⊃q) is true, then ◊(p & ¬q) is false, so p ◊→ ¬q is false. If there is no world in which p and ¬q are both true, then still less is there a ‘nearby’ world in which they are both true, however you construe ‘nearby’.

Sometimes, the truth of a necessarily true counterfactual is secured by a contingent truth about the world. For example: in fact, the coin landed heads. I say, truly: “If I had bet on heads, with the coin landing as it in fact did, then I would have won.” But what underwrites this counterfactual is one in which the necessitation is laid bare: “If I had bet on heads, with the coin landing heads, then I would have won.” And I will happily concede its truth, too.

Still, my concessions are happy, because I have claimed that most ordinary counterfactuals are false, and these counterfactuals are not ordinary. They are not counterfactuals that you hear uttered on the street, or indeed outside a philosophical discussion (not that you would hear them much *inside* a philosophical discussion, either).

6.2 Counterfactuals under determinism with sufficiently precise antecedents

Let us explore further an interesting subclass of counterfactuals whose truth is secured by a nomological connection between antecedent and consequent.

I argued earlier that not even determinism suffices to save counterfactuals from falsehood: the problem was that vaguely specified antecedents encompass initial conditions that yield anomalous results. What is needed is determinism *plus* sufficient precision in the antecedents regarding the initial conditions, relative to the consequents. That way, the antecedents necessitate the corresponding consequents, and we have more instances of type i).

A counterfactual about my jump will be true provided that the antecedent together with the (deterministic) laws of nature imply the consequent. Thus, a counterfactual whose antecedent fully specifies my jump, molecule-by-molecule, will be true provided that all possible trajectories consistent with that specification result in the truth of the consequent—as it might be, 'I fall'. A counterfactual that specifies my jump more imprecisely may still be true, but only if there is sufficient imprecision in the consequent to tolerate it. As we saw above, even the seemingly imprecise consequent 'I fall' was too precise to be implied by the antecedent 'I jump', so the corresponding counterfactual was false. To make it true, either we have to precisify the antecedent so as to rule out all anomalous initial conditions that result in my not falling, or we have to *vagueify* the consequent so as to be compatible with all of them. Precisifying the antecedent sufficiently will be quite a task; while it may not require a molecule-by-molecule specification of the jump, it will require a lot of fancy footwork nonetheless, presumably intractably so. Vagueifying the consequent sufficiently is easier but risks rendering the counterfactual pointless—e.g., "if I were to jump, something would happen to me". Of course, we could secure the truth of the counterfactual by building in a logical connection between antecedent and consequent—e.g., "if I were to jump according to initial conditions that result in my falling, then I would fall". But now the counterfactual is trivial.

So our options for producing true counterfactuals—precisifiying the antecedent, vagueifying the consequent, or building in a necessary connection between antecedent and consequent—yield counterfactuals that are respectively intractably complicated, pointlessly vague, or trivial. In any case, they are hardly ordinary.

6.3 Counterfactuals with probabilistic consequents

There are many other counterfactuals that *may* be non-trivially true, namely those that explicitly state the appropriate probability in the consequent itself. ‘If the coin were tossed, it would land heads with chance 1/2’ may well be true for all I’ve said. There may well be a tiny real number, $\varepsilon > 0$, such that ‘if I were to jump in the air, I would come down with chance $1 - \varepsilon$’ is true. And so on.

There is no contradiction between a counterfactual with a probabilistic consequent and the opposite ‘might’ counterfactuals, even when the probability is high, just as there is no contradiction between ‘$P(X) = x$’ and ‘not X’, even when the probability is high. I said earlier that there are precious few valid inference rules taking us from modal claims to probabilistic claims; there are even fewer—namely, none—taking us from probabilistic claims to claims about how things actually turn out. The only thing that can contradict a probability statement is a contradiction, or another probability statement (that attributes a different probability), or something that entails such a statement. This is true even if $x = 1$ or 0, as we saw in the example of the infinite sequence of coin tosses, all of which might land tails. So we have to countenance the possibility that various counterfactuals with probabilistic consequents are true.

I suspect that when such a counterfactual is true, it is a special case of a counterfactual with a nomological connection between antecedent and consequent—that is, once again a case of i). The tossing of a coin is not lawfully connected to its outcome (under indeterminism); but it may well be lawfully connected to the *chance* of such an outcome.

So as before, perhaps I cannot appeal to a might-counterfactual that involves the failure of the lawful connection. Perhaps it is not true to say: ‘if the coin were tossed, it might not land

Heads with chance 1/2', since the laws may determine the chance of Heads to be 1/2. In that case, my argument from undermining might-conditionals will not go through. Similarly for all of the counterfactuals whose consequents are chancy, discussed in §2, as long as the correct chance is stated in the consequent. So counterfactuals with chancy consequents may be non-trivially true.

Then again, they too may be false, and for five different reasons.

i. The chance value claimed may be incorrect

The most obvious reason is that there is only one *correct* value for a given chance, and uncountably many *incorrect* values. The truth of 'If the coin were tossed, it would land heads with chance 1/2' implies the falsehood of 'If the coin were tossed, it would land heads with chance x' for all x ≠ 1/2, since the chance function is a *function*.

What we really mean is "If the coin were tossed, it would land heads with chance *roughly* half". Given that there is some chance that a real coin lands on its edge, we can be confident that the chance of heads is *not* ½. It would be remarkable if we got the chance *exactly* right, to infinitely many decimal places of accuracy. Indeed, if the true chance is transcendental, then with very few exceptions (involving π , e, and their kin) we cannot even express it with our current linguistic resources. So we cannot express any corresponding true counterfactual involving it, either.

ii. The antecedent may not be sufficiently precise

Secondly, the antecedent may specify the conditions too vaguely to yield a unique chance. Indeed, this is surely true of the coin-tossing example (so an unqualified concession three paragraphs ago to its truth would have been premature). If the coin were tossed very feebly an inch above the ground, it might not land heads with probability 1/2; if the mathemagician Persi Diaconis were to toss the coin, it might not land heads with probability 1/2… Chances must be relativized to a chance set-up; but if the set-up is incompletely described, there may be considerable leeway in the resulting chances.

This is closely related to the argument I gave in §4. There I considered counterfactuals under determinism, and contended that an insufficiently precise antecedent could be compatible with anomalous results inconsistent with the consequent of a given counterfactual. Reverting to counterfactuals with probabilistic consequents is a way of recovering determinism at one step removed, as it were—for while there is indeterminism governing the consequent, the connection *between* the antecedent and a *probabilistic* consequent is supposed to be perfectly deterministic. And much as vagueness afforded a way of undermining the truth of counterfactuals under determinism, so it affords a way of undermining the truth of counterfactuals with probabilistic consequents.

iii. The chances may be chancy

Thirdly, it could be that the chancy consequents are themselves chancy—that there are *higher-order* chances attaching to the propositions in question. Chances attach to propositions; a typical chance statement has the form

$$ch(X) = x.$$

(Relativize this to a time, if you like.) However, that chance statement itself picks out a proposition—call it Y. For all we know, Y is itself in the domain of the chance function:

$$ch[Y] = y,$$

that is,

$$ch[ch(X) = x] = y.$$

There is an issue here of just how profligate chances are—just how big that domain is. But supposing that there is no restriction on the set of propositions that can be the bearers of chance, then we have no obstacle to such self-referential statements of chance. Moreover, such higher-order chances may be non-trivial, intermediate in value. This would seem to be a live possibility on a 'Humean supervenient' account of chance—see Lewis (1994) and Hoefer (20xx)

Perhaps, then, a counterfactual with a second-order chance statement as its consequent is true—say, 'if I were to toss the coin, then with chance 0.99993 it would land heads with chance 1/2'? But the problem, if it is a problem, may arise again one level up. The 0.99993 figure may itself be chancy. And so on. The staircase of non-trivial higher-order chances may never end. Moreover, counterfactuals with probabilistic consequents are hardly ordinary; with each added step up the staircase, they become even less so. So even if there is hope that somewhere up the staircase we reach true counterfactuals, they will hardly swell significantly the ranks of those true counterfactuals that are ever uttered or written.

iv. There may be chance gaps

Fourthly, unless chances are maximally profligate, attaching to *all* propositions, there are *chance gaps*—propositions that don't receive any chance at all. We've already seen some candidates for such propositions: those picked out by chance statements. Elsewhere (2003) I have argued that there are other candidates. I focused especially on *free acts* and *non-measurable sets.*

It may well be that *free acts,* such as my raising my foot now, simply don't receive any chance value at all. After all, while chances should not be identified with (limiting) relative frequencies (see e.g. Hájek 1996), such frequencies are typically good evidence for the values of chances, and they typically nearly coincide. Yet we can easily drive the (limiting) relative frequencies of our free acts to whatever value we want (and in the limit, to no value at all), *in virtue of their very freedom*.

Similarly, certain symmetric experiments over uncountable spaces give rise, at least in principle, to *non-measurable sets:* sets that cannot be assigned any probability at all, consistent with certain natural looking assumptions. Randomly throwing an infinitely fine-tipped dart at a representation of the [0, 1] interval is arguably such an experiment. Such sets would be chance gaps. See Hájek 2003 for more explanation and defence.

The upshot is that various propositions may simply fail to receive a chance value at all, under a given counterfactual assumption. In that case, any counterfactual with that assumption as its antecedent, and *any* particular chance attribution to such a proposition as its consequent, is false. It claims that were that antecedent to be the case, the chance of the consequent would be such-and-such, when in fact that chance is undefined.

v. The chances may be vague

Finally, and now letting our two devices for undermining counterfactuals work in tandem, *chances* themselves might be *vague.* For it is plausible that chances are determined by the laws of nature, and if the laws themselves are vague, chances could inherit this vagueness. This would certainly seem to be a live possibility on a Mill-Ramsey-Lewis style account of laws as regularities that appear as theorems in a 'best' theory of the universe, as long as the criteria for what makes one theory better than another are themselves vague. (In Lewis' 1973 theory, for instance, the vagueness may enter in the standards for balancing the theoretical virtues of 'simplicity' and 'strength'.) Then nature may not determine a single best theory, but rather a multiplicity of such theories. Suppose, for example, that these equal-best theories disagree on the chance that a radium atom decays in 1500 years: for each real number r in the interval [1/3, 2/3], there is such a theory that says that the chance is r. Then the chance of this event may be vague over this interval.

There may be even more straightforward ways for the laws of nature to be vague. Perhaps some of the fundamental physical constants are not entirely precise—perhaps, for example, the gravitational constant is only fixed up to 100 decimal places. Or some of the fundamental physical properties might be vague. It would seem that the laws in which such constants or properties figure would then be rendered vague, and that chances, which are determined by the laws, are correspondingly vague. The chance that this coin lands heads on a given toss, for example, may not be a sharp number such as 1/2, but rather a set of such numbers, or perhaps an interval. In that case, even the counterfactual 'If this coin were tossed, it would land heads with chance 1/2' is false—it overreaches, claiming a sharp chance when there is none. The true counterfactuals concerning the coin have consequents that are vague: if I were to toss the coin, it would land heads with chance in such-and-such interval, or in such-and-such set, or with chance fixed to only-so-many decimal places. And such counterfactuals are anything but ordinary, and certainly rarely uttered.

Indeed, they may need to be still more complicated than that in order to be true. Suppose we admit higher-order vagueness, so that not even the endpoints of the intervals, or the number of decimal places, are sharp. We could handle such cases by appealing to still more sophisticated devices, packing them explicitly into the consequent itself. Of course, that means that the counterfactuals are getting still less ordinary: where previously they had probabilistic consequents with sharp probabilities (which made them rarified enough), now the consequents have vague probabilities, suitably characterized by whatever devices for higher-order vagueness are needed. We have left street talk far behind.

In sum, even counterfactuals with chancy consequents can be false, since the chance specified can be incorrect, the antecedent might be too vague to yield a unique chance, there might be higher-order chances, there might be chance gaps, and there might be vague chances. So I may not even have to concede, happily or otherwise, the truth of various counterfactuals with chancy consequents. But I don't mind much either way, because they too are not ordinary. So we don't have to settle these contentious points here.

We have seen earlier that all trivially true counterfactuals can be regarded as those whose antecedents necessitate their consequents. We can either precisify the antecedent, or tailor the consequent to the antecedent (getting the chance right), or vagueify the consequent. We've been looking at technical devices for achieving these things. How do we do that while keeping the counterfactuals ordinary? But most counterfactuals that are ever uttered are ordinary.

We won't see the former device, precisifying, much in ordinary language, so let's go the other way. What vagueifying devices do we have?

6.4 Comparative and qualitative counterfactuals

I *will* concede that there are counterfactuals with consequents that are vaguer still than those I have considered that might count as ordinary—for example, some counterfactuals with *comparative* or *qualitative* consequents concerning probabilities. "If I were to jump, I would be much more likely to fall than not", "if I were to jump I would very likely fall", and so on may count as ordinary. If so, my concession is a tad less happy, but there it is. I did, after all, only claim that *most* counterfactuals are false.

But maybe even this concession is too hasty. I surmise that these counterfactuals would strike the ordinary person as rather odd. They would find puzzling their apparent coyness, their diffidence, in much the same way that they would find puzzling the remark, perhaps from someone who has read a bit of Hume, but not too much: “The sun will *probably* rise tomorrow”. That may well be true, but the Gricean in us wants to hear asserted the stronger claim that we also believe to be true: “The sun *will* rise tomorrow. (Dammit!)” Likewise, that same Gricean wants to hear asserted “if I were to jump, I *would* fall. (Dammit!)”, taking this (rightly) to be a stronger claim than its comparative or qualitative counterparts, and (wrongly) to be true. In short, I am not sure that the ordinary person would find the latter counterfactuals so ordinary. Such counterfactuals that really are ordinary are false, and the corresponding ones that may well be true don’t sound so ordinary after all.

Still, eventually my resources for denying the truth of counterfactuals will run out, and some of the counterfactuals that remain are—I grudgingly concede—ordinary. Consider a counterfactual whose consequent’s probability is easily and correctly characterized in comparative or qualitative terms, sufficiently informatively that even the Gricean in us is satisfied. The truth of “If Shakespeare had not written Hamlet, very probably nobody else would have” seems secure, as is its ordinariness. Indeed, there will be a spectrum of such cases, corresponding to the range of such qualitative or comparative probabilistic claims: “If Einstein had not come up with the theory of relativity, it is fairly likely that eventually someone else would have”; “if Gore had campaigned harder in Florida, he would have improved his chances of winning the election”, and so on. Come to think of it, we have a way of capturing the whole spectrum at once: “if Kennedy had not been shot in Dallas, there would have been *at least some chance* of his serving his full term as president.” But that sounds like just another way of saying the ‘might’-counterfactual, “if Kennedy had not been shot in Dallas, he *might* have served his full term as president.” (Yes, I know that events of chance zero can happen, and I even exploited their existence in §2. But they require infinitely many trials, something we do not have in the Kennedy example.) ‘Might’ counterfactuals, that is, correspond to ‘would’ counterfactuals with maximally vague probabilistic consequents. And of course I have already conceded the truth of various ‘might’ counterfactuals; in fact, doing so has been the linchpin of my argument!

6.5 Counterfactuals with true antecedents and consequents

It is a consequence of both the Stalnaker and the Lewis semantics for counterfactuals that ‘X → Y’ is true whenever both X and Y are true. This may be plausible on a similarity-based semantics, in which ‘X → Y’ is true iff Y is true at all X-worlds sufficiently similar to the actual world. For no world could be as similar to the actual world as itself—this assumption is sometimes called *centering*—and if X & Y is true at the actual world, we are apparently done. Then it would appear that I need to concede the truth of another large class of counterfactuals: those whose antecedents and consequents are true.

Well, maybe not. This quick argument from considerations of similarity of worlds is a little too quick. It is too quick by my lights, since I have questioned similarity-based accounts of counterfactuals. And it is too quick if the notion of ‘similarity’ at play is *not* simply that of commonsense, but rather some (quasi-) technical notion, as we discussed in §2. Moreover, it is debatable whether conditionals with true antecedents should even count as ‘counterfactuals’. After all, they are not *contrary-to-the-fact-uals*, whereas that’s what is supposedly distinctive about ‘counterfactuals’. This could quickly devolve into a rather boring terminological dispute. It’s not clear why this dispute should be resolved in the less favorable way for me, but let me simply agree to resolve it that way. That still won’t undermine my position.

So, does X & Y imply X → Y, as centering would have it? Here are several reasons for thinking not.

Firstly, it already strains the ear to say that X → Y is true when X and Y are true but are independent of each other. 'If Canberra were the capital of Australia then the moon would have large craters' is more likely to puzzle the common folk than to get their universal assent (not that they are the final arbiters of truth). An extreme case of this one in which X concerns some tiny, localized event, and Y concerns some enormous widespread event, both of which actually occur. 'If I had blinked just now, the entire history of the universe would have been ___' comes out true if we insert into the blank a true statement of the entire history of the universe, for in fact I *did* blink just now.

Secondly, matters are worse when there *is* a connection between X and Y, but of the wrong sort: when the antecedent, if anything, tends to prevent or inhibit the consequent, but despite the truth of the antecedent, the consequent still manages to be true. Then X is evidence *against* Y. Consider the 1989 Australian Rules Football Grand Final between Geelong and Hawthorn. (It helps the example if you don't know what happened; better still if you don't even know what Australian Rules Football *is*.) I tell you: "If Geelong had completely outplayed Hawthorn in the final quarter, they would have won." I expect you will want to read 'they' as referring to Geelong in order to render this counterfactual true. But according to centering, it is true iff we read 'they' as referring to Hawthorn. For as things actually turned out, Geelong *did* completely outplay Hawthorn in the final quarter, and despite that Hawthorn *did* win (their three-quarter time lead proved to be unassailable).

Thirdly, according to centering, X & Y entails not just X → Y, but also Y → X. The conjunction of the two counterfactuals is a biconditional—we might call it a *bicounterfactual,* and symbolize it:

$X \leftarrow\!\rightarrow Y$.

The previous two problems are only exacerbated. For example, relations of counter-support often go in both directions. ‘If Gore had won the popular vote in the 2000 election, then Bush would have won the election’ is a case in point. Yet by centering, it is true that

Gore wins the popular vote ←→ Bush wins the election.

Fourthly, centering faces a problem with ‘might’ counterfactuals, assuming them to be duals of ‘would’ counterfactuals. ‘If I had tossed the coin, it might have landed tails.’ Arguably, this is true even if, in fact, I did toss the coin and it landed heads. But by centering, the ‘would’ counterfactual is true, so this ‘might’ counterfactual is false.

Finally, centering commits us to a dubious inequality concerning the probability of a counterfactual:

$$P(X \rightarrow Y) \geq P(X \& Y),$$

(a special case of the theorem of probability theory that $P(U) \geq P(V)$ whenever V implies U). The inequality is sufficiently dubious, indeed, *Lewis himself* seems to doubt it (1986, 22). He gives an example involving a chancy counterfactual of the form $A \rightarrow C$, and he concludes that $P(A \rightarrow C) \approx 0$. He appeals to much the same intuition as the one I began with regarding coin tossing: since $A \Diamond\!\rightarrow \neg C$ is very probably true, $A \rightarrow C$ is very probably false. Yet in Lewis’s example, $P(A \& C)$ may be as high as 0.97.

Thus, I am not convinced that I must automatically grant the truth of all counterfactuals with true antecedents and consequents. In fact, I’m tempted to say that those that are true are —you guessed it—those with some sort of necessary connection between antecedent and consequent, in which case we have just more examples of case i). Then the true counterfactuals are a very special breed indeed—they are really those whose truth is secured by corresponding strict conditionals!

But even if I concede that centering secures the truth of some counterfactuals, I still think they represent a very small proportion of the counterfactuals that we actually utter. Much of the *point* of uttering counterfactuals, after all, is to convey information about a hypothetical scenario, one known or believed or assumed to be non-actual. If for nothing but Gricean reasons, we usually don't assert 'X → Y' if we know or believe or assume X and Y both to be true; we simply assert their conjunction instead, which (assuming centering) is more informative. To be sure, we sometimes accommodate an interlocutor who disagrees with us on the truth of X by asserting the less informative counterfactual. But this often happens only as a concession *after* disagreement over the conjunction has emerged; and often it doesn't even happen then.

* * * * * * *

That's where my concessions end—and some of them were not really concessions after all. Even granting as true all of the counterfactuals in §§6.1 – 6.5, it is somewhat disquieting that they were all underwritten by corresponding strict conditionals or conjunctions—none of them seem to get to the heart of *counterfactuality*. All other counterfactuals, I claim, are false, and they form the vast majority. And so I conclude again, now I hope with even more justification than before: *most counterfactuals are false.*

7. Rethinking the logic of counterfactuals

Certain argument forms, which are valid for the material and the strict conditional, are said to be invalid for the counterfactual. Here are some examples:

Transitivity:

$A \to B$

$\underline{B \to C}$

$\therefore\ A \to C$

Stalnaker (1968) gives the famous counterexample:

If J. Edgar Hoover had been born a Russian, then he would have been a Communist

If he had been a Communist, he would have been a traitor.

∴ If he had been born a Russian, he would have been a traitor.

Here, supposedly the premises are true but the conclusion false. Or consider

Strengthening the antecedent:

$\underline{A \to C}$

$\therefore\ (A\&B) \to C$

Lewis (1973) gives the following counterexample:

If I had struck that match, it would have lit.

∴ If I had struck that match, and it had been soaking in water, it would have lit.

Again, supposedly the premise is true but the conclusion is false.

Contraposition also putatively fails for counterfactuals; and so on. See Lewis (1973, §1.8) for more examples and discussion.

Contra Stalnaker, Lewis, and apparently received wisdom, I claim that we don't really have counterexamples to these inference patterns, because the premises are in fact not true. So there is no evidence from the failure of these patterns that counterfactuals obey some special logic.

Or consider another putative logical principle governing counterfactuals, *conditional excluded middle:*

(CEM) $(A \to C) \vee (A \to \neg C)$

Stalnaker subscribes to (CEM) – indeed, its status is the chief point of disagreement between him and Lewis. But by my lights, (CEM) is no logical principle, and indeed it fails for all cases except those conceded in the previous section. For example, 'if I were to jump, I would come down' is false (because I might not come down), and 'if I were to jump, I would not come down' is false (because I might come down). Thus both disjuncts, and hence the disjunction, is false:

(I jump → I come down) v (I jump → ¬ I come down)

This should come as no surprise, given my argument that all true counterfactuals are underwritten by either strict conditionals or conjunctions. For (CEM) is false if we replace the '→'s by either fish-hooks, ampersands, or one of each.

If I'm right about this, it undercuts an argument in favor of similarity-based accounts: the fact that they give an elegant account of the failures of these inference rules.

8. How is our practice of uttering counterfactuals vindicated?

And yet we go on cavalierly uttering counterfactuals. How, then, does our practice survive?

Here is my best hypothesis: In the neighborhood of the ordinary but false counterfactuals that we utter, there are closely related counterfactuals that are true but not ordinary. They are counterfactuals with appropriate probabilistic or imprecise consequents. Where the consequents are probabilistic, "appropriate" means: the chance stated in the consequent is sufficiently high. If the chance is sharp, the consequent states it correctly. If the chance is vague, the consequent states the region of vagueness correctly. Where the consequents are not probabilistic, "appropriate" means: the antecedent is sufficiently precise to accommodate the precision of the consequent. Either way, the counterfactual is made true by a corresponding strict conditional.

We are guilty, then, of what we might call *rounding errors*—we treat high chance propositions as certainties, low chance propositions as impossibilities, within the scope of these counterfactuals. IMPRECISION? But this is a very minor crime, at least when committed by the folk in ordinary contexts. (It may not be so minor when committed by philosophers in the extraordinary contexts that they create—more on that shortly.) It is reasonable for us to do so when the rounding errors are small, as they often are. We say, for instance: 'if I were to jump, I would come down', which strictly speaking is false; but doing so may be justified pragmatically by the truth of a neighboring counterfactual, as it might be: 'if I were to jump, I would come down with chance 0.99999993'.

Or perhaps, as I argued earlier, this isn't quite right either, for I gave some reasons for skepticism even about counterfactuals with probabilistic consequents. So perhaps our practice is vindicated by counterfactuals that combine two of the devices that I offered as helping to secure their truth, probability, and vagueness:

'if I were to jump, I would come down with chance 0.99999993 ± 0.00000001';

or vaguer still:

'if I were to jump, I would come down with chance at least 0.97;

or even vaguer still:

'if I were to jump, I would come down with very high chance';

So there are true counterfactuals closely related to the ones we assert that support our practice, at least when the prevailing standards for asserting counterfactuals are somewhat forgiving, as they typically are on the street. So we can legitimately assert various counterfactuals. Still, most of them remain false.

9. Is this merely philosophical pedantry?

You may grant all this, but not be disturbed by it. You may say that we already knew that various things we say are not strictly speaking true, but they are close enough to true to serve our purposes well enough. As Unger (1975) would argue, according to strict philosophical standards, nothing would count as *flat.* Kansas is not *really* flat—why, it varies in elevation by several feet! The most carefully crafted pool table is not really flat—why, even the naked eye can discern tiny bumps in the felt! One could point out the lack of flatness of Kansas, or the table, but in most contexts doing so would simply be tiresomely pedantic. Speaking the way we do about their flatness serves our purposes well enough. So you may say that I'm just being tiresomely pedantic here. Our rounding errors are as harmless as saying "It's 2 o'clock" when your watch reads 1:58.

So counterfactuals with probabilistic consequents support our practice, at least when the prevailing standards for asserting counterfactuals are somewhat forgiving, as they typically are on the street. Or so you may say.

My reply has three parts: the street is not always forgiving; even when it is, falsehood is merely forgiven rather than eradicated; and we are not always on the street. Earlier I attempted some damage control. Now it is time for some:

10. More damage

10.1 The street is not always forgiving

Even the person on the street is not immune to the effects that I have highlighted, for even on the street contexts can be created in which improbable possibilities are salient. I have not bought a ticket in this week's million-ticket lottery. I say: "If I were to buy a ticket, it would lose". You can make me take that back by having me concede that there's a one-in-a-million chance that any given ticket wins, and thus if I were to buy a ticket, it *might* win. After all, why else do people buy tickets? The very act of buying a ticket makes salient a possibility that is very improbable: *this very ticket's winning*.

And even on the street, we can make bizarre possibilities salient. Hollywood could easily make a movie whose plot turns on various characters quantum tunneling—in fact, I'm surprised that as far as I know, such a movie hasn't been made already! Having been told about quantum tunneling, you cannot immediately ignore it. A context has been set up in which quantum tunneling is a live, salient possibility. In that context, it is not tiresome pedantry to balk at various ordinary counterfactuals whose truth is sabotaged by the possibility of quantum tunneling. On the contrary, it is required of a cooperative, attentive audience.

10.2 On the street, falsehood is merely forgiven, not eradicated

Granted, on the street we get away with uttering various counterfactuals whose truth I have questioned. Still, that is no proof that they *become true* on the street. Earlier I expressed my doubts about a contextualist account according to which they do. And the linguistic data that we have about street-talk could equally be explained by an account according to which the counterfactuals become *assertable*. I suggest, then, that high, but less-than-1, probabilities of their consequents suffice for assertability, but not for truth. How high the probabilities need to be may be context-dependent, in a way that the falsehood of the counterfactuals is not. I can thus concede that much to the friends of context-dependence regarding counterfactuals, without conceding the context-dependence of their truth-values.

I think there is an important disanalogy here between 'flatness' and counterfactuals—and here I part company with Unger. I think 'flat' (and indeed all other adjectives) is tacitly *comparative*. While it looks like a one-place property, it is really a two-place relation. When we say ‘x is flat’, we mean ‘x is flatter than __’, where the ‘__’ is filled in by a salient alternative to x—perhaps a neighboring case to x, or the average of a set of cases to which x belongs. Thus, when I say that Kansas is flat, I may be comparing its degree of overall flatness to the surrounding landscape, or to other states, or … Moreover, once context makes clear the second relatum, what I say may be *true*—pace Unger. Flatness of landscapes comes in degrees; relative to some alternatives a given landscape may have a higher degree of flatness, while relative to others it has a lower degree.

But I don’t think that truth comes in various degrees—pace degree-of-truth theorists. I can make no sense of the notion that a given counterfactual is ‘truer than’ another, while ‘less true than’ a third one. Truth comes in two degrees. A miss is as good as a mile; and most counterfactuals miss.

So at best, various counterfactuals that we utter on the street—which are almost always ordinary—are assertable (although of course many of them are not even that). And as followers of Adams have emphasized regarding indicative conditionals, assertability does not imply truth. He went further, and thought that indicative conditionals don’t have truth values at all—not that that stops us from uttering them. I, by contrast, am happy to grant that counterfactuals *have* truth values. In many cases, however, they are not what we offhand took them to be—not that that stops us from uttering them.

10.3 We are not always on the street

I have claimed that we are guilty of various tiny rounding errors, ignoring the tiny probabilities that under various counterfactual suppositions, various consequents turn out false. (Sometimes this is the even tinier rounding error of treating a possibility with zero probability as it is impossible.) Moreover, I agree with the spirit of ‘your’ objection in section

9: perhaps it *is* tiresomely pedantic to worry about tiny rounding errors. But it is a philosopher's job to be tiresomely pedantic. And I don't just mean that while I'm presenting a philosophical paper such as this, I can reasonably raise standards of precision above what they would be on the street. That's true (unlike most counterfactuals), but not all that surprising, especially in these post-Ungerian, post-Lewisian times. The problem runs deeper than that. *For various philosophers employ counterfactuals in their philosophical positions, or in their conceptual analyses.* Counterfactuals are, as I said at the outset, a philosophical staple these days: they figure in influential analyses of causation, perception, knowledge, personal identity, laws of nature, rational decision, confirmation, dispositions, free action, explanation, and so on. There, the standards of precision are high, the context unforgiving. And when they are high, philosophers can't so easily plead the person on the street's excuse. The danger is that the various counterfactuals that are supposed to underpin causation, perception, knowledge, personal identity, … turn out to be false, instead of true, when their analyses *require them to be true*. (We can safely assume that these counterfactuals are not on my list of concessions in §6, although you may like to check that for your favourite counterfactual-laden analysis.) By analogy, if the analysis of some concept dear to us required some things to be *perfectly flat,* and it turned out that there were no such things, then either the analysis, or the concept, would be in trouble. (This last sentence, while a counterfactual, is trivially true.) Fortunately, we do not run into this problem with flatness.

However, we *often* run into this problem with counterfactuals. Either the various philosophical analyses that appeal to them are mistaken, or we live in a world devoid of causation, devoid of perception, devoid of knowledge, devoid of persisting persons, and so on. Assuming that our world is not so impoverished, it is the analyses that are under threat.

In §5.1, I countenanced the possibility that the counterfactuals I was considering were *indeterminate* rather than false. I said there that I would not yet budge on my more radical thesis that they are false. But let me now budge, for the sake of the argument. Rewrite this section, if you like, replacing the word "false" by "indeterminate". This should hardly comfort the relevant philosophers, and it will hardly save their analyses. For the philosophers think that various claims of causation, of perception, and so on, are *true*. This is not the case if their analyses are right, and if the counterfactuals that figure in those analyses are indeterminate. So it is not true that we live in a world with causation, perception, knowledge, persisting persons, and so on. Either our commonsense beliefs about the world require wholesale revision, or, more likely, these philosophical positions do.

Nor did I feel I had to budge to allow that the truth values of most counterfactuals are context-dependent (§5.2), but let us grant that now for the sake of the argument. Then the counterfactuals that figure in the analyses of causation, perception, knowledge, personal identity, and so on are context-dependent. This in turn means that if these analyses are correct, or at least correct insofar as they employ counterfactuals (even if their exact details are incorrect), the analysanda are context-dependent, unless somehow the context-dependences 'cancel out'.[15] To see how this is possible, suppose we analyze 'bachelor' as 'either a tall unmarried male or a non-tall unmarried male'. Each disjunct of the analysis is context-dependent, since 'tall' is. However, the context-dependence of each disjunct compensates for the other, taking up its slack: as we move to a context that is more demanding for 'tall' (say, by discussing NBA basketballers), we *ipso facto* move to a context that is less demanding for 'non-tall', and vice versa. As a result, 'bachelor' proves not to be context-dependent after all. Good news for 'bachelor'! However, I very much doubt that there will be such good news for 'causation', 'perception' and so on. While we would need to run

[15] I thank Alex Byrne for this observation.

through their analyses on a case-by-case basis, I bet that in each case the context-dependence of their counterfactuals won't cancel out.

The threat, then, is that if these analyses are correct, then much that we hold dear turns out to be context-dependent. Perhaps in some cases that may not be such a worry; in fact, perhaps we already had reason to believe it. Contextualism about some those concepts already has some currency—see e.g. Schaffer 2006 on causation, or Lewis 1996 on knowledge, or van Fraassen 1980 on explanation. Note, however, that the context-dependence of counterfactuals may reveal a *further*, perhaps unintended, unwanted, and hitherto unappreciated source of context-dependence in these concepts. Moreover, contextualism is rather less appealing for some of the other concepts—for example, for laws of nature, or perception, or personal identity. Imagine a defence lawyer telling the jury that it is *context-dependent* whether the defendant is the same person as the murderer! And do you think that whether or not you are the same person as you were when you began reading this paper is, despite its considerable length, *context-dependent*?!

To be sure, these philosophical analyses are given *in a particular context*—a philosophical context. That may alleviate the problem of context-dependence: once and for all, the counterfactuals in the analyses should be evaluated by the standards of that context. But this only makes worse the problem of the *falsehood* of the relevant counterfactuals. After all, philosophical contexts are demanding: bizarre possibilities are fair game. We can legitimately entertain possibilities in which billiard balls quantum tunnel to China or in which human bodies vaporize; indeed, if we are doing our job properly, we should be so imaginative. This brings us back full circle to my original argument for the falsehood of most counterfactuals—such were the possibilities that I entertained when arguing for the truth of the various 'might not' counterfactuals that undermined the corresponding 'would' counterfactuals. I gave that argument in a particular context—a philosophical context.

Well may we wonder, then, how the slogan "counterfactuals are context-dependent" can be repeated so blithely when philosophers so often reach for counterfactuals in their conceptual analyses. Or perhaps we should wonder, rather, how philosophers can reach so blithely for counterfactuals when that slogan is repeated so often! Either way, something is amiss. I can't accept that the facts about the laws of nature, or perception, or personal identity, are context-dependent. Nor can I accept that we should be eliminativists about these things because our favorite philosophical analyses of them involve counterfactuals that are false. I suspect, rather, that the fault lies with the analyses. Again, this should really be addressed on a case-by-case basis, and our verdicts may differ across the cases. Still, I hope that *my* slogan regarding counterfactuals—that most of them are false—will alert us to the dangers that one courts when trafficking in them while philosophizing. A similar caution carries over to their use in the sciences and the social sciences.

*   *   *   *   *   *   *

If had more time, I would say still more about all of this. But that's a counterfactual, and it's doubtless false.[16]

[16] My first debt is to David Lewis – anyone familiar with his work should be able to see how much it has influenced me. I wrote a short first draft of this paper when I was a visiting fellow in 2002 at the Center for Philosophy of Science at the University of Pittsburgh, which provided a very stimulating research environment. My thanks to all who helped me there. I am grateful to many people over the years since then for helpful discussions and/or comments on this material that led to improvements therein, especially: Robert Adams, Karen Bennett, Eliza Block, Alex Byrne, Jim Cargile, David Chalmers, Fiona Cowie, Vladan Djordjevic, Kenny Easwaran, Antony Eagle, Andy Egan, Adam Elga, Justin Fisher, Hilary Greaves, Chris Hill, Chris Hitchcock, Carrie Jenkins, Jenann Ismael, Martin Jones, Jeff King, Boris Kment, Douglas Kutach, Igal Kvart, Marc Lange, Stephan Leuenberger, Aidan Lyon, Daniel Nolan, Denis Robinson, Jonathan Schaffer, Wolfgang Schwarz, Ted Sider, Nico Silins, Robert Stalnaker, Jason Stanley, Brian Weatherson, Duncan Watson, Lee Walters, Jim Woodward, Lyle Zynda, and audiences at Indiana University South Bend, Melbourne University, Monash University, Singapore Management University, University of New South Wales, Oberlin College, Arizona State University, University of North Carolina at Chapel Hill, McGill University, the December 2006 meeting of the APA in Washington, and the 2007 meeting of the AAP in Armidale. If this paper had not benefited from their comments, it would be poorer.

Philosophy Program
Research School of Social Sciences
Australian National University
Canberra, ACT 0200
Australia

REFERENCES (to be continued and filled in)

Adams, Robert Merrihew (1977): “Middle Knowledge and the Problem of Evil”, *American Philosophical Quarterly* 14, No. 2, 109 – 117.

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

Bigelow, John and Robert Pargetter (1990): *Science and Necessity*, Cambridge: Cambridge University Press.

DeRose, Keith (1999): “Can It Be That It Would Have Been Even Though It Might Not Have Been?”, *Philosophical Perspectives* 13, Epistemology, 385 -413.

Elgin, Catherine (1988): “The Epistemic Efficacy of Stupidity.” *Synthese* 74: 297-311.

Hájek, Alan (1997): "*'Mises Redux'—Redux:* Fifteen Arguments Against Finite Frequentism", *Erkenntnis*, Vol. 45, 209-227. Reprinted in *Probability, Dynamics and Causality – Essays in Honor of Richard C. Jeffrey,* D. Costantini and M. Galavotti (eds.), Kluwer.

Hájek, Alan (2003): “What Conditional Probability Could Not Be”, *Synthese,* Vol. 137, No. 3, 273-323.

Halliday and Resnick (19xx): Fundamentals of Physics, Vol. 2, John Wiley & Sons.

Hawthorne

Kvart, Igal (1986): *A Theory of Conditionals*, Indianapolis: Hackett.

Lewis, David (1969): *Convention: A Philosophical Study,* Harvard University Press.

Lewis, David (1973): “Counterfactuals and Comparative Possibility”, *Journal of Philosophical Logic* 2; reprinted in Lewis 1986.

Lewis, David (1986): *Philosophical Papers,* Vol. II, Oxford University Press.

Lewis, David (1994): “Humean Supervenience Debugged”, *Mind* 103, 473-490.

Lewis, David (1996): “Elusive Knowledge”, *Australasian Journal of Philosophy* 74, No. 4, 549-567.

Plantinga, Alvin (1974): *The Nature of Necessity,* Oxford University Press.

Pollock, John (1976): *Subjunctive Reasoning*, Boston: Reidel.

Schaffer, Jonathan (2006): “Contrastive Causation”, *Philosophical Review.*

Sorensen, Roy (1988): *Blindspots,* Oxford University Press.

Unger, Peter (1975): *Ignorance: A Case For Skepticism,* Oxford: Clarendon Press.

van Fraassen, Bas (1980): *The Scientific Image,* Oxford: Clarendon Press.

Williams, R