**Counterpossibles and Similarity**

David Letterman once asked his audience, ‘Based on what you know about him in history books, what do you think Abraham Lincoln would be doing if he were alive today? One, writing his memoirs of the Civil War. Two, advising the President. Three, desperately clawing at the inside of his coffin.' We are being asked to evaluate the counterfactual

*if Lincoln were alive today, he’d be writing his memoirs*,

and that sounds likely enough, until we hear Letterman’s third option and begin to suspect we’re being asked a rather different question than we had first supposed.

David Lewis has had some important insights into how we are to understand counterfactuals of this sort, and these insights help us see how to think about counterpossibles, i.e., counterfactuals with impossible antecedents, such as

*if Hitler had time-travelled into the past and killed himself as an infant, then World War II would not have occurred*.

A Lewisian semantics for counterfactuals can be extended in such a way that counterpossibles are treated as false or as non-trivially true. The context-sensitivity of counterfactuals and a distinction between certain types of contextual influence, both highlighted by Lewis, take on an even greater importance in the case of counterpossibles. Though Lewis would endorse neither the semantics to be proposed nor its underlying ontology, the proposed account is Lewisian in spirit. In fact, the primary objection to the extended semantics parallels many of the early objections to Lewis’s original semantics. If Lewis was successful in defending his account against these, a semantics of counterpossibles can be defended from similar objections in the same way.

According to Lewis, all counterpossibles are trivially true. If we were to suppose that the unentertainable were true, he argues, we might just as well suppose that anything at all is true. There are counterpossibles that we would not normally assert, such as

*if there were a largest prime p, then there would be six regular solids*

or

*if there were a largest prime p, then pigs would fly*.

But, says Lewis, we would not confidently deny them either. And so he is content to let all counterpossibles come out true.[[1]](#footnote-1)

I will argue for a theory of counterfactual propositions that allows counterpossibles to be false, and so allows others to be non-trivially true. The theory will be extended to address ‘might’ counterfactuals and questions about the relative ‘nearness’ of impossible worlds. Throughout I will argue that an appreciation of counterfactuals’ context-sensitivity enables us to see how the account is best elaborated and defended against objections.

The proposed account uses impossible worlds. Those who find this ontology dubious may yet be able to take advantage of the analysis proposed below. It will be a matter of replacing possible and impossible worlds with consistent and inconsistent maximal classes of propositions, where ‘maximal’ describes a class of propositions containing, for every proposition P, either P or its negation.[[2]](#footnote-2) I will assume for present purposes that each such class corresponds to a world. (Thus the account assumed here differs slightly from the account in [Nolan 1997], in which worlds may have truth value gaps. As in Nolan’s account, however, every necessary falsehood is true in some world or another [1997: 542].)

‘Impossible’ will be taken in a fairly wide sense. The antecedents of the counterpossibles to be considered may be expressed by explicit contradictions or analytic falsehoods; they may be ‘countermathematicals’ or ‘counterphilosophicals’ [Lewis 1973: 24]; or they may be necessary falsehoods of any other sort. The account is thus broader than the one Lewis briefly entertains [1983a: 19] in which the antecedents are ‘not-too-blatantly impossible’.[[3]](#footnote-3)

This is not to deny that there are interesting differences between the kinds of necessarily false propositions, or between the sentences that express them in a given language. The focus here, however, will be on their common impossibility. As in Lewis’s account, the similarity relation between worlds will be the primary determinant of truth-values. I will not argue specifically that each sort of impossibility may be non-vacuously true, but I will defend the general claim that the range of potential similarity relations is very wide.

**False Counterpossibles**

Lewis’s analysis of counterfactuals is equivalent to the following.

(1) A counterfactual is (non-trivially) true iff some possible world in which both the antecedent and the consequent are true (an A C world) is more similar to the actual world than every possible world in which the antecedent is true and the consequent is false (an A ~C world).

Since the antecedent of a counterpossible is not true in any possible world, (1) itself does not return a truth-value of ‘true’ for any counterpossible. Lewis treats counterpossibles as a special case.

Some have argued that a correct account of counterfactuals ought not assign all counterpossibles the value ‘true’, offering examples like

*if I were to create,* ex nihilo*, one duck every morning, then I would create only five ducks each week*,

or

*if the number five grew wings and webbed feet, it would be no different than it is now*.

We might join Lewis in throwing our hands in the air upon an unentertainable supposition, but even obviously impossible propositions can be entertained. And if so, some instances would seem to be properly judged false.

Multiplying purported examples of false counterpossibles is unlikely to convince adherents of the standard account, however. In the next section we will consider more general reasons to regard some counterpossibles as false, reasons which in large measure accede to the objector’s intuitions. We can set the stage by looking at the source of our practices of assertion.

It is true that there are many counterpossibles that we would not confidently deny. However, familiar features of counterfactuals provide a ready explanation of this fact. As the Lincoln example illustrates, counterfactual sentences are especially sensitive to context.[[4]](#footnote-4) Take this well-known pair of sentences from Quine:

(C1) If Caesar had been in command [in the Korean War] he would have used the atom bomb.

(C2) If Caesar had been in command he would have used catapults.

It may be true that if Caesar had been in Korea he would have used catapults--but not if we are discussing how various personalities of history would have deployed the various weapons actually available during the Korean War. Before it can be determined whether a particular counterfactual sentence expresses a truth or not, the evaluator must understand why it is asserted and what features of reality it is meant to point out.

When the context is not entirely clear from what has already been said, a principle of charity takes effect: the one who asserts a counterfactual presumably means it to be true, and so the listener assumes that the context is one which makes the assertion come out true. When someone makes the lone statement, ‘If that guy had been any taller he would have hit his head on the ceiling fan’, the charitable listener assumes that the assertor is primarily concerned with the height of the guy in question, discounting the likelihood that if he had been any taller he would have been more careful about where he was standing. In effect, the listener chooses a similarity relation which, when operative in the analysis (1), makes the statement express a truth.

Lewis puts the point this way.

That is how it is in general with dependence on complex features of context. There is a rule of accommodation: what you say makes itself true, if at all possible, by creating a context that selects the relevant features so as to make it true [1986: 251].

When a counterpossible sentence is uttered, there is often a plausible context that would make it true, especially when the antecedent and consequent do not appear to be relevant to each other. If one heard the serious assertion that if there were a largest prime p then pigs would fly, one would tend to take this as a way of saying a largest prime is out of the question, i.e., that the antecedent is, for present purposes, ‘unentertainable’.

Since such an interpretation is often plausible, we hesitate to deny counterpossibles. We do not want to deny what is true on a plausible interpretation. And if we think too long about a given counterpossible sentence, we may begin to lose sight of its original context, the source of the reasons for thinking it false. So we do hesitate. Nevertheless, *contra* Lewis, our reluctance does not justify assigning the value ‘true’ to all of them.

**Objection: Strict Implications Entail the Corresponding Counterfactuals**

But if P entails Q, isn’t that sufficient for the truth of the counterfactual that has P as its antecedent and Q as its consequent? In other words, the objector may say, isn’t

(R) P  Q

 P  Q

a logically valid rule of inference? Some take the inference to be self-evidently valid. If so, then every counterpossible is true, despite the fact that there are a variety of counterpossible statements we may want to deny. According to this objection, the proposition

*if every natural number had three divisors, then no number would have more than one divisor*

is true. Whatever inclination we might feel to deny it is either simply mistaken or else a confused apprehension of the fact that the its opposite,

*if every natural number had three divisors, then some number would have more than one divisor*,

also true, is more enlightening in some way.

Of course, we may deny that every necessary falsehood entails every proposition, i.e., that entailment should be understood as strict implication. It might also be helpful to deny that a contradiction entails everything. Then the rule of inference would be better stated as

(R\*) P e Q

 P  Q

and we would not be forced to admit that its premise is true whenever P is necessarily false. In this way paraconsistent (and perhaps other non-classical) logics may come to the aid of a theory of non-trivial counterpossibles.

However, I would prefer not to rely on this strategy, since the case for non-trivial counterpossibles does not require it. As I argue below, even classical logicians have strong reasons to agree that some counterpossibles are false. The theory of counterpossibles under consideration could be taken as part of an attempt to show that classical logic can deal usefully with inconsistent premises. Alternately, it could be taken as further evidence that classical logician’s usual attitude toward hypothetical reasoning is misguided, and that the entire approach should be traded for something more fruitful [Nolan 1997: 557-8]. I would like the present theory to appeal to as many as possible. Best, then, not to take a stand. If the objection can be defeated without appeal to the nature of entailment, so much the better.

To that end, let us consider two elements of Lewis’s theory: the influence of context and the rule of accommodation.

Lewis mentions Quine’s Caesar sentences to illustrate how two kinds of contextual influence might be called upon to explain the apparent truth of apparently conflicting counterfactuals. On one theory context serves to supply a part of the antecedent that has been left implicit. The antecedents of the pair might implicitly specify that Caesar have the latest technology available to him, so that (C1) expresses a truth and (C2) does not. In another context, the antecedents might specify that Caesar have the weapons that were in fact available to him, so that (C1) expresses a falsehood and (C2) expresses a truth.

Lewis prefers the theory that context has another kind of influence.

[I]nstead of using context to restore the real antecedent from the explicit part of the antecedent, I could say that the explicit antecedent *is* the real antecedent and call on context rather to resolve part of the vagueness of comparative similarity in a way favorable to the truth of one counterfactual or the other. In one context, we may attach great importance to similarities and differences in respect of Caesar’s character and in respect of regularities concerning the knowledge of weapons common to commanders in Korea. In another context we may attach less importance to these similarities and differences, and more importance to similarities and differences in respect of Caesar’s own knowledge of weapons [1973: 67].

So there are several ways in which we might plausibly see context as resolving a tension among our intuitions about counterfactual pairs like the above. Whether we prefer one kind contextual influence or the other or some combination of the two, sentences (C1) and (C2) each express one proposition in one context and a different proposition in another.

Further, there is the rule of accommodation. Assert that Caesar would have used the atom bomb, and you speak the truth. Or deny that he would have used the bomb, saying instead that he would have used catapults, and again you speak the truth. In the case of counterfactuals, context affects not only standards of precision [Lewis 1986: 251-2] but also the relative weights of various respects of comparison. Since we charitably tend to select the similarity relations that make a counterfactual true, a false-making context will be harder to find. The pertinacious but habitually charitable objector who seizes upon a true-making context may not readily see that the same sentence can also express a falsehood.

Naturally contextual influence and the rule of accommodation apply to counterpossibles. When someone says that the sentence ‘if there were a largest prime p, then pigs would fly’ is vacuously true since, if the unentertainable were true, anything you like would be true, they’re right. Here the assertion that the antecedent is unentertainable, along with the conventional use of the ‘pigs would fly’ clause, fixes a similarity relation that makes the counterpossible trivially true. For classical logicians: If someone says the sentence is true because its antecedent entails (strictly implies) its consequent, they’re right, too. In so saying, they create a context in which the entailment of consequent by the antecedent has a significant weight, a context in which worlds with the same entailment relations are *ipso facto* more similar to the actual world than others.

But that context is a rather different one than we might expect to find, say, in a mathematics classroom. In such a context, no connection between the antecedent and consequent is salient. If we can ignore the usual function of ‘pigs would fly’, the proposition expressed will strike us as false (and a little odd). The proposition expressed in this case is entirely different than the one(s) expressed in the earlier contexts. Because of the rule of accommodation this context shift is easily accomplished.

It would be startling indeed if counterpossibles were immune to this context-sensitivity. The claim that counterpossibles must be trivially true, then, is like the steadfast contention that ‘If that guy were much taller, he would hit his head on the ceiling fan’ is false. It is an insistence on one way of measuring similarity, a failure to recognize that other ways may be more appropriate on certain occasions.

Early objections to Lewis’s account of counterfactuals were much the same. Typically they would propose a counterexample, say, a clearly true counterfactual whose consequent was apparently not true in antecedent-worlds much like the actual world. In reply, Lewis pointed to the ‘extreme shiftiness’ of similarity. The way of measuring similarity manifested in our snap judgments is rarely the one best suited to evaluating counterfactuals, but the objectors seemed to use that method exclusively.[[5]](#footnote-5) ‘It is all too easy to make offhand similarity judgments and then assume that they will do for all purposes [Lewis 1983b: 42].’

**Entailment in Context**

One might complain that although shifts in context can bring certain counterfactual sentences to express truths on some occasions and falsehoods on others, there are limits [Lewis 1973: 93]. Some sentences express truths in any context--any context, at least, that we are likely to produce. And, the classical complainant may add, entailment relations are important enough to us that counterpossible sentences are each of this sort. The fact that an antecedent entails its consequent will never be outweighed by some more important respect of comparison; entailment is the metaphysical bottom line.

We may grant that in some sense entailment is the bottom line, but that is not to say that it is always a maximally important feature of every comparison. Here I want to suggest that we have two general reasons for thinking that it is not always so.

First, which similarity relation governs a given counterfactual depends on the relative importance of various respects of comparison. Their relative importance, in turn, depends on the intentions of the assertor, the immediate conversational context of the assertion (if any), and the broader social context that surrounds all conversation. And our intentions, the ways in which we might wish to emphasize one aspect of the world over another, vary rather widely indeed. If an assertor can shape her intentions and, by them, the immediate context so that some respect of comparison outweighs the entailment (in whatever sense) of one proposition by another, nearby worlds may preserve similarity in the favoured respect at the cost of dissimilarity in respect of the entailment. The entailing proposition may be true, and the entailed proposition false, in a world that is relatively similar to the actual world in the relevant respects.[[6]](#footnote-6)

Second, we sometimes do explicitly what our intentions accomplish implicitly. We could say, ‘If that guy were much taller (disregarding the difference in his location) he would hit his head on the ceiling fan’, even if the parenthetical phrase is not really needed. If we think analytic falsehoods entail everything, we can still say, ‘If a married bachelor were Prime Minister (never mind that this is contradictory and entails everything), then a man would be Prime Minister’. The parenthetical clause points out explicitly that the entailment of consequent by antecedent is not meant to be a respect of comparison between antecedent worlds and the actual world. But it does not need to be pointed out explicitly; without the parenthetical clause it is clear enough which respects of comparison are relevant.

We now have another reason not to rely on paraconsistent logic as a means of criticizing the argument that all counterpossibles must be true. Even the paraconsistent kind of entailment may fail to be a maximally important respect of comparison. Similarity is in general independent of entailment, no matter how entailment is understood. The appeal to paraconsistent logic does undermine the argument based on the rule of inference (R) but is beside the present point.

I contend, then, that rule (R) and the generic (R\*) are invalid. This sets the theory against the idea that counterfactual implication is an intermediate between strict implication and material implication, i.e., that P  Q entails P  Q and that P  Q entails P  Q [Lewis 1973: 23; Pollock 1975: 75]. Rule (R) gets its appeal from consideration of a natural but incomplete range of cases.

**A Modified Account**

A satisfactory account of counterfactuals, then, will allow a nontrivial analysis of counterpossibles. Fortunately, impossible worlds can be used to modify Lewis’s analysis slightly to do just that. Consider this revision:

(2) A counterfactual is true iff some (possible or impossible) A C world is more similar to the actual world than every A ~C world is.

(2) does allow for nontrivial analysis of counterpossibles, as desired. A counterpossible is false just in case no A C world is more similar to the actual world than every A ~C world. Otherwise the counterpossible is true.

Analysis (2) has at least two advantages of over analysis (1): counterpossibles may turn out to be either true or false, and counterpossibles are assigned truth values on the same basis that all other counterfactuals are assigned truth values--no separate analysis or justification is required. In most other respects, (1) and (2) are on a par. (2) does not make the dubious limit assumption Lewis avoids, that is, the assumption that there is always some collection of antecedent worlds closest to the actual world and never an infinite series of closer and closer antecedent worlds [1973: 19-21]. And (2) satisfies as well as (1) does the motivation behind Lewis’s account.

However, our analysis needs a little chisholming. The formulation of (2) subtly made use of the assumption that in no world are both P and ~P true. But we cannot rely on that assumption once we introduce impossible worlds. Consider this sentence:

(F) If France were a monarchy and France were not a monarchy, then France would be a monarchy.

(F) seems to be true; France could not possibly be both a monarchy and a non-monarchy, but if it were both of those things, a monarchy is one of the things it would be. But the sentence is false according to (2). The antecedent worlds most similar to the actual world are worlds in which both ‘France is a monarchy’ and ‘France is not a monarchy’ are true. (For the antecedent to be true without both of its conjuncts being true would be a gratuitous departure from actuality in this case.) But then no A C world is more similar to the actual world than any A ~C world is, because the nearest A C worlds are A ~C worlds. (See Diagram 1, which takes advantage of the custom of treating the worlds most similar to the actual world as nearby worlds).



**Diagram 1.**

Since we would like a counterfactual to be assigned ‘true’ whenever the closest antecedent worlds are all consequent worlds, we need to distinguish the ~C worlds from the nonC worlds, i.e., those which are not C worlds. Thus:

(3) A counterfactual is true iff some (possible or impossible) A C world is more similar to the actual world than every A nonC world.

This amendment of (2) makes no difference in the evaluation of most counterfactuals, but it makes (F) true, and so avoids an infelicity of (2). Account (3) shares the advantages of (2) over (1) and will serve nicely as our final analysis.

**‘Would’ Implies ‘Might’**

The meanings of the words ‘would’ and ‘might’ seem to be such that ‘would’ implies ‘might’. That is, a ‘might’ counterfactual[[7]](#footnote-7) of the form ‘If A were the case, then C might be the case’ follows from the corresponding ‘would’ counterfactual, ‘If A were the case, then C would be the case’. Let us call this principle WIM.

One somewhat awkward consequence of Lewis’s account is the falsehood of WIM. Unfortunately, the problem is not easily solved with impossible worlds. Even an account that admits false and non-trivially true counterpossibles must choose between mutually inconsistent but plausible theses.

Lewis offers a rather intuitive definition of the ‘might’ counterfactual in terms of the ‘would’ counterfactual. A ‘might’ counterfactual ‘If A were the case, then C might be the case’ is true if and only if the negation of ‘If A were the case, ~C would be the case’ is true. Using ‘’ as the ‘would’ connective and ‘◊’ as the ‘might’ connective,

DEF: A ◊ C = ~(A  ~C).[[8]](#footnote-8)

It follows immediately from DEF and the trivial truth of all counterpossibles that every ‘might’ counterfactual with an impossible antecedent is false. When A is impossible, the definiens is never satisfied, and so neither is the definiendum. The awkward result is the falsehood of WIM. Sometimes a ‘would’ counterfactual is true and the corresponding ‘might’ counterfactual is false.[[9]](#footnote-9) One must affirm ‘If there were a largest prime p, pigs would fly’ but deny ‘If there were a largest prime p, pigs might fly’.

In some ways, though, Lewis’s semantics reflects the intuition that ‘would’ counterfactuals imply the corresponding ‘might’ counterfactuals. The account of ‘would’ counterfactuals and the interdefinability of ‘would’ and ‘might’ together entail that each non-vacuous ‘would’ counterfactual does imply the corresponding ‘might’ counterfactual. Lewis treats the failure of vacuously true counterfactuals to conform to WIM as a somewhat counterintuitive result that is to be accepted on the strength of the principles that entail it.

Sadly, the awkwardness appears even if we do not assume the trivial truth of all counterpossibles. For suppose that DEF and WIM are true. Is it ever the case that both of A  C and A  ~C are true? If it is, then by DEF the ‘might’ counterfactuals A ◊ ~C and A ◊ C are both false. However, according to WIM, both A ◊ C and A ◊ ~C are true. So it follows from DEF and WIM that A  C and A  ~C are not both true. But there seem to be counterexamples to this conclusion, such as

(F) If France were a monarchy and France were not a monarchy, then France would be a monarchy

and its opposite

(F\*) If France were a monarchy and France were not a monarchy, then France would not be a monarchy.

Whether or not all counterpossibles are true, (F) and (F\*) certainly look that way. Our unpleasant options are to say that (F) and its ilk are false, to reject WIM, or to reject the plausible DEF. Whichever we choose, we owe some explanation of the resulting oddness. The choice will be a matter of judgment rather than conclusive argument, but still I can present my judgment and say why it is least counter-intuitive to me. I reject DEF.

Rejecting WIM is very unattractive. Intuitively it seems true, and I have been unable to find any promising counterexamples. Here I can do little more than point to what seem to be the pertinent meanings of the words ‘would’ and ‘might’.

Only a bit less unattractive is denying either (F) or (F\*). If I try, I can generate some confusion about the antecedent. ‘How am I to envision such a thing?’ I ask myself. ‘What would it be like for France to be and not to be a monarchy?’ But none of this confusion helps me escape the conclusion that if, impossibly, France both were and were not a monarchy, France would be a monarchy (and, furthermore, it would not be a monarchy).

One might argue that the explicitly contradictory antecedent brings to mind a logically chaotic state of affairs--and who knows whether the two conjuncts will be true in it? A relatively ambitious view about what kinds of impossible worlds there are (such as the one assumed here) might have the result that for any proposition P, there are impossible worlds in which the conjunction[[10]](#footnote-10) P & ~P is true but the conjunct P is not. (On such a view, the truth value of P & ~P in an impossible world would not in general be a function of the truth values of P and of ~P, on the grounds that the conjunction’s non-truth-functionality is one of the impossibilities.) But the plausibility of (F) and (F\*) make it difficult to deny that in at least some contexts such worlds are not the ones most similar to the actual world.

DEF, on the other hand, seems to make a doubtful presupposition. Making A ◊ C and A  ~C exclusive of each other, it in effect presupposes that whatever the antecedent, the nearest antecedent worlds will never be ones in which both the consequent and its negation are true. Of course, for the vast majority of antecedents this is so, but we cannot rely on this assumption when antecedents may be necessary falsehoods or even formal contradictions. Lewis regards technical convenience as factor to be weighed when choosing a theory, and cites the interdefinability of ‘would’ and ‘might’ counterfactuals as a mark in DEF’s favour.[[11]](#footnote-11) For my part, I doubt technical convenience is any strong indicator of truth, and not enough in any case to outweigh the considerations against it.

How shall we understand the ‘might’ counterfactual? I propose:

(4) A ◊ C = no A ~C world is more similar to the actual world than every A C world

or equivalently

(4\*) A ◊ C = no A nonC world is more similar to the actual world than every A C world.

The equivalence of (4) and (4\*) depends on the maximality of worlds, i.e., on the absence of truth-value gaps.

The above differs slightly from the account derived from analysis (3) of the ‘would’ counterfactual and DEF. That account can be stated as follows:

(5) A ◊ C = no A ~C world is more similar to the actual world than every A world that is not a ~C world.

(4) and (5) yield different truth-values in cases where all the A worlds most similar to the actual world are ~C worlds, and some are also C worlds. Here (4) returns the value true and (5) will in general return the value false. (See Diagram 2.) It is never the case that (5) gives the value true when (4) gives the value false, so (4) is a strictly weaker account than (5). According to (4), the right-hand side of DEF, ~(A  ~C), entails the left, A ◊ C, but not vice versa.



**Diagram 2.**

If all the A worlds most similar to the actual world are C and ~C worlds (as illustrated in Diagram 1), then (3) assigns the value ‘true’ to A  C and (5) assigns the value ‘false’ to A ◊ C. I have argued that this is the very situation we find in the case of counterfactuals like (F) and (F\*). This illustrates the falsehood of WIM assuming the truth of (3) and DEF.

If (4) is correct we do lose the technical convenience of defining the ‘might’ connective as in DEF. Edward Wierenga has offered a substitute account [1998] for those who insist that ‘would’ implies ‘might’.

DEF\*: A ◊ C iff ~(A  ~C)  (A  C)

Wierenga suggests that the awkwardness in Lewis’s account can thereby be eliminated without recourse to the thesis of false counterpossibles. If in fact there are no impossible worlds, the revised definition may serve. But if ever we encounter the scenario illustrated in Diagram 2, where both C and ~C are true in some of the nearest antecedent worlds, DEF\* will fail.

**The Unentertainable**

Above I distinguished between the impossible and the unentertainable and argued that, sometimes at least, not just anything would be true if the impossible were true. But what of those propositions, possible or impossible, that really are unentertainable? Does each of them counterfactually imply every proposition?

We should pause first to clarify the notion of entertainability. Each conversation proceeds on certain assumptions, and some propositions violate those assumptions[[12]](#footnote-12), whether by being inconsistent with them or by being unlikely on their basis or in some other way. Presumably an entertainable proposition is one that can be supposed or considered, given the assumptions currently in effect. Thus entertainability, like similarity, is context-relative. Perhaps in some context it is unentertainable that the earth reverse its rotation, so that ‘If the earth reversed its rotation, then pigs would fly’ is trivially true and any counterfactual which shares its antecedent, such as ‘If the earth reversed its rotation, then the sun would rise in the west’, is also trivially true in that context. But in other contexts the former is false and the latter is non-trivially true.

We may be inclined to make something like Lewis’s claim about the unentertainable[[13]](#footnote-13). Let’s call it (U).

(U) For any sentence Q and any sentence P that expresses a proposition that is unentertainable in a given context, P  Q is true in that context.

Lewis points out that a similar claim about ‘might’ counterfactuals could be made.

(U\*) For any sentence Q and any sentence P that expresses a proposition that is unentertainable in a given context, P ◊ Q is true in that context.

(U\*) is a result of Lewis’s alternate account of ‘might’ and ‘would’, offered to those who find it dubious that every counterpossible is true. The alternate ‘might’ could be motivated with a shrug like that which motivated (U). ‘If that were so, anything you like might be true!’ [1973: 25]. This alternate account assigns each ‘would’ counterpossible the value ‘false,’ and so is no more plausible than Lewis’s preferred account. However, one might well accept (U\*) without accepting all of the alternate account.

I have a mild preference for (U\*) over (U). If nothing else, we may note that if WIM is true, then (U\*) follows from (U), and so (U) puts us further out on a limb. But in any case both (U) and (U\*), as well as the claim that neither is true, are consistent with the impossible world theory outlined above. In both accounts the truth-values of counterfactuals with entertainable antecedents will be unaffected.

Following Stalnaker [1968], let us call the impossible world in which every proposition is true ‘’. If (U) is true, then  is the unique closest world in which an unentertainable antecedent A is true. For suppose that some other antecedent world W is as similar or more similar to the actual world given the relevant similarity relation. Then some proposition P is true in  but not in W. But then by (3), A  P is false, contrary to (U).

Several types of world-orderings are compatible with (U\*). One kind of world-ordering that is sufficient for the truth of (U\*) is that which makes all worlds in which the unentertainable antecedent is true equally and maximally dissimilar to the actual world. In this case all A C and A ~C worlds are equally similar to the actual world, and so no A ~C world is nearer than every A C world, and by (4) A ◊ C is true.

The theory permits us to say that a possible antecedent can be unentertainable. This seems to be a conceptual possibility, since entertainability is a matter of what we language-users are prepared to suppose in a given context. It seems quite likely that some logically possible proposition is or might be unentertainable in some context or another. If (U) is true, the question stands or falls with the question whether there is a context and a possible antecedent A such that every conditional of the form A  C is true in that context; the affirmative answer has the consequence that at least one impossible world, , is more similar to the actual world than some possible worlds in the given context. If (U\*) is true, the question stands or falls with the question whether there is a context and a possible antecedent A such that every conditional of the form A ◊ C is true in that context; this has the consequence that some impossible world is at least as similar to the actual world as some possible world (since some among the consequents counterfactually implied by A would be impossible).

**Are Impossible Worlds Ever Closer Than Possible Worlds?**

Even if we are disinclined to adopt (U), we might wonder whether any counterfactuals with possible antecedents and impossible consequents are true.[[14]](#footnote-14)

Again we have an issue that may be resolved either way given the above theory. The proposed analysis does not tell us very much about the character of the similarity relations generated by the contexts our conversation can produce. However, the variability of what is conversationally important suggests that impossible worlds sometimes are more similar to the actual world, in relevant respects, than some possible worlds are.

Suppose a student, working on a computer program which carries out proofs in a formal system, knows that the system’s Gödel sentence is of special significance, and so includes a command to print ‘I have proved my Gödel sentence!’ if ever such a proof is found. The instructor points out that the command is superfluous: ‘The Gödel sentence says, in effect, “This sentence cannot be proved in this system,” so your program isn’t going come up with a proof of it. The system would have proved something it couldn’t if it executed the print command’. On the face of it, the instructor’s last remark expresses a true counterfactual with a possible antecedent and an impossible consequent. If so, then some impossible antecedent world (one in which the consequent is true) is closer to the actual than any possible antecedent world.

Such examples will not persuade those who believe that context influences a counterfactual only by filling in a largely implicit antecedent. If context acts as a parenthetical addendum to the counterfactual in the present example, we have something like:

*If the computer were to print ‘I have proved my Gödel sentence!’ (and the computer were working exactly according to its flawless program without malfunction or outside interference), then it would have proved something that it cannot*.

If this is the way we think of the counterfactual, it is no longer a counterfactual with a possible antecedent, such as is needed to support the claim that impossible worlds are sometimes closer to actuality than some possible worlds.

We have some reason to believe that a large part of context’s influence (but not all) is of the second sort Lewis mentions, the specification of a similarity relation rather than the explication of an antecedent. In this case, a counterfactual proposition’s antecedent is more or less what it seems to be. We could express a counterfactual whose antecedent is the one apparently expressed, and would presumably do so with the same words. Why suppose we have not done so? True, we often do leave implicit some of what we intend to express, and so it would be surprising if we did not sometimes leave implicit something of counterfactuals’ antecedents. But it would be even more surprising if we were unable to express possible antecedents at all in the presence of impossible consequents. The most natural understanding of sentences like the instructor’s remark gives them possible antecedents, and takes them to express truths.

Both the character of counterfactuals’ antecedents and relevant respects of similarity are shaped by convention, and so have great flexibility. Lewis’s insights into counterfactuals thus furnish us both with a strong defence of counterpossible semantics against objections and also with the makings of a case for the relative nearness, in some rare contexts, of impossible worlds.

**Impossibilities Too Close to Home?**

Nevertheless, it might be thought that similarity is the Achilles’ heel of a possible/impossible world semantics. (The objection is inspired by Lewis’s comments about relatively nearby ‘impossible limit worlds’ [1983a: 14-5], though Lewis proposes nothing like it himself.) Far from claiming that impossible worlds are always more distant than the possible worlds, the objection alleges that if the antecedent of a counterfactual is false, the antecedent world most similar to the actual one will invariably be an impossible world. Among the impossible worlds, one differs from the actual world only by affirming a single extra proposition, the antecedent. If such worlds are always the worlds most similar to the actual one, then our semantics is in serious trouble. For if both the antecedent and consequent of a counterfactual are false, and if, in the antecedent world most similar to ours, the consequent remains false, the counterfactual itself will come out false. But many counterfactuals with false antecedents and consequents are clearly true. If I had jumped from my 13th story window this morning, I would have fallen to my death. Furthermore, since such worlds affirm both the antecedent and its negation, the sentence ‘If I had jumped from my 13th story window this morning, I would not have jumped from my 13th story window this morning’ will come out true, and clearly it should not.

Are we compelled to judge that ‘one-proposition difference’ worlds are always the antecedent worlds most similar to actuality? Here are two methods of analyzing similarity that answer in the affirmative, both of them faulty. Each can be shown to produce similarity relations very unlike the ones normally operative in a possible worlds semantics. If the central idea of Lewis’s semantics is right and there is a notion of overall similarity between possible worlds that allows the semantics to work even reasonably well, then there is no reason to suppose that the similarity notion of a possible/impossible world semantics must be like those that fall prey to the objection.

Let us call the first way of analyzing similarity the cardinality method. The idea is simply that the smaller the cardinality of the class of propositions on which the two worlds differ, the more similar they are to each other. (Two worlds *differ* on a proposition whenever the proposition is true in one of them, but not in the other. The class of propositions on which two worlds differ is their *difference*.) The maximum of similarity is the minimum of difference; i.e., two worlds that differ on no propositions are maximally similar. They are, in fact, identical. (The cardinality method thus entails Lewis’s assumption that each world is more similar to itself than any other world is [1973: 14-5].) A world differing from the actual on four score propositions will be more similar to the actual world than one differing on a countable infinity of propositions, which in turn will be more similar than a world differing on continuum-many propositions. Worlds that differ from the actual world by the same number of propositions will be equally similar to it. Apparently on the cardinality method a world differing from the actual only by the (false) antecedent of some counterfactual will be the nearest antecedent world.

But the cardinality method does not generate a notion of similarity that is at all useful in comparing the nearness of possible worlds. Two observations are in order. First, there are pairs of possible worlds that differ on as many propositions as there are propositions. Since there are at least as many propositions as there are cardinal numbers, and there is no cardinality of the collection of cardinal numbers, the differences between possible worlds cannot always be measured by their cardinalities as the cardinality method implies. Second, even if there were enough cardinal numbers to measure the differences between possible worlds, the cardinality method would not yield a usable similarity relation because the cardinality of the differences would always be equal to that of all propositions.

This second point may be seen as follows. Let us pretend that the class of all propositions has a cardinality, , and that each of its subclasses has a cardinality as well. Now consider two distinct possible worlds, W and W\*, which differ at least on some proposition P. For every proposition Q which W and W\* share, there is a distinct proposition on which they differ, viz., Q  P. So if d is the cardinality of the class of propositions on which W and W\* differ and s that of the class of propositions they share, s  d. Because W and W\* either differ on or share each particular proposition, s + d = . Finally, since any infinite cardinal added to itself yields itself, d = d + d ≥ s + d = . Clearly d is not greater than , so d = . So for any two possible worlds, the cardinality of the class of propositions on which they differ is equal to the cardinality of the class of all propositions.

But then the cardinality method evidently fails to distinguish nearby possible worlds from those very unlike the actual world. Each possible world is like any other, as far as this method can tell us. It cannot be that the cardinality method is the manner of measuring similarity implicitly at work in a successful Lewis-style counterfactual semantics.

Will a related method do the job? The cardinality method attempts to correlate great similarity with small differences and to measure the differences by cardinality. But there is another way of comparing the sizes of sets and classes, the way that tells us there are more natural numbers than even numbers, despite the fact that these sets have the same cardinality. Let us call this way the subclass method. The idea is that whenever the class of propositions on which worlds W1 and W2 differ is a proper subclass of the class of propositions on which W1 and W3 differ, W2 is more similar to W1 than W3 is. Of course, many pairs of worlds will be incomparable, neither class of differences being a subclass of the other, but we may take the subclass method to tell us something about comparative similarity in those cases where it does apply. If an impossible world differs from the actual world only by the false antecedent of some counterfactual, that world will be the unique, nearest antecedent world on the subclass method.

The problem this time is that no two possible worlds are comparable. For suppose each of the propositions on which possible worlds W1 and W2 differ is also one on which W1 and possible world W3 differ, and that W1 and W3 differ on additional propositions as well. Let P be one of the propositions true in W2 and W3 but not W1, and let Q be true in W3 but not W1 or W2. Then the material conditional P  Q is true in W1, since W1 is possible and P is false there, and in W3, since both antecedent and consequent are true there, but not in W2. In W2 it is ~(P  Q) which is true instead. But then P  Q is a proposition on which W1 and W2 differ, but on which W1 and W3 agree. Contrary to supposition, the differences between W1 and W2 do not form a proper subclass of the differences between W1 and W3.

So the subclass method, like the cardinality method, fails to show that any possible world is more or less similar to the actual world than any other. Neither method will serve as an analysis of the similarity notion of Lewis’s semantics. And so, assuming Lewis’s account is not a failure from the beginning, there is a notion of comparative overall similarity between worlds which has nothing whatever to do with the sizes of the classes of propositions on which worlds differ, whether we judge the size of a class by its cardinality or by its sub- and superclass relations (or by both). And if such a notion is available, why shouldn’t we be able to use it in a possible/impossible world semantics for counterfactuals?

There is another reason why the cardinality and subclass methods cannot govern a counterfactual semantics. The point belongs to Lewis: It is fitting that we analyze counterfactuals with a vague notion, since counterfactuals themselves are vague in just the way similarity is. However, neither cardinality nor the subclass relation is vague; neither depends in any way on context. Hence the cardinality and subclass methods are inflexible where similarity is flexible, and the former will not serve to analyze the latter.

The above objection, then, gives us no reason to think that similarity is more problematic in a possible/impossible world counterfactual semantics than it is in a possible world semantics. Unless Lewis’s possible world account is much more seriously flawed than has generally been recognized, we need not suppose that the nearest non-actual worlds are always impossible. One lesson to be learned from this is that overall similarity between worlds is generally a matter of much more than the size of their difference. Among other things, it is a matter of the modal properties of the states of affairs involved--something which may have nothing to do with how many or few propositions they differ on.[[15]](#footnote-15)

**References**

Barwise, Jon and John Perry 1983. *Situations and Attitudes*, Cambridge: The MIT Press.

Cohen, Daniel H. 1988. The Problem of Counterpossibles, *Notre Dame Journal of Formal Logic* 29: 91-101.

Lewis, David 1973. *Counterfactuals*, Cambridge: Harvard University Press.

Lewis, David 1983. Counterfactuals and Comparative Possibility, in *Philosophical Papers* II, New York: Oxford University Press: 3-31.

Lewis, David 1983. Counterfactual Dependence and Time’s Arrow, in *Philosophical Papers* II, New York: Oxford University Press: 32-66.

Lewis, David 1983. Scorekeeping in a Language Game, in *Philosophical Papers* I, New York: Oxford University Press: 233-49.

Lewis, David 1986. *On the Plurality of Worlds*, New York: Blackwell.

Mares, Edwin D. 1997. Who’s Afraid of Impossible Worlds?, *Notre Dame Journal of Formal Logic* 38: 516-526.

Nolan, Daniel 1997. Impossible Worlds: A Modest Approach, *Notre Dame Journal of Formal Logic* 38: 535-572.

Perszyk, Kenneth J. 1993. Against Extended Modal Realism, *Journal of Philosophical Logic* 22: 205-214.

Pollock, John 1975. Four Kinds of Conditionals, *American Philosophical Quarterly* 12: 51-59.

Schlossberger, Eugene 1978. Similarity and Counterfactuals, *Analysis* 38: 80-82.

Stalnaker, Robert 1968. A Theory of Conditionals, in *Studies in Logical Theory*, ed. N. Rescher, Oxford: Blackwell.

Strand, Jonathan 2001. Can a Possibility ‘Counterfactually Imply’ an Impossibility? (unpublished).

Tichy, Pavel 1976. A Counterexample to the Stalnaker-Lewis Analysis of Counterfactuals, *Philosophical Studies* 29: 271-273.

Vander Laan, David 1997. The Ontology of Impossible Worlds, *Notre Dame Journal of Formal Logic* 38: 597-620.

Wierenga, Edward 1998. Theism and Counterpossibles, *Philosophical Studies* 89: 87-103.

Yagisawa, Takashi 1988. Beyond Possible Worlds, *Philosophical Studies* 53: 175-204.

Zagzebski, Linda 1990. What If the Impossible Had Been Actual?, in *Christian Theism and the Problems of Philosophy*, Michael Beaty, Notre Dame: University of Notre Dame Press.

1. Though Lewis does not think that we want to assert that any counterpossibles are false, he does admit that his reasons are less than decisive and offers an alternative analysis which makes all ‘would’ counterpossibles false and all ‘might’ counterfactuals true (1973: 25)--a point we will take up later. [↑](#footnote-ref-1)
2. Edwin Mares, Daniel Nolan, and I conceive of impossible worlds as abstract objects [Mares 1997; Nolan 1997; Vander Laan 1997]. Mares treats impossible worlds as structures like Barwise and Perry’s states of affairs [Barwise and Perry 1983], and remains neutral about the nature of the individuals and relations involved. The account should thus have broad appeal. Nolan observes that believers in abstract possible worlds, fictionalists, and others will have no greater ontological difficulty with impossible worlds than possible ones, and may in fact already be committed to them. Takashi Yagisawa defends a theory of concrete impossible worlds [1988]. Concrete impossible worlds are criticized by several others [Lewis 1986: 7n; Nolan 1997; Perszyk 1993; Vander Laan 1997]. [↑](#footnote-ref-2)
3. The account is also broader than Zagzebski’s account [1990], where only counterfactuals whose antecedents are ‘Interesting Impossible Propositions’, i.e., those whose antecedents are not self-contradictory, are under consideration. It is somewhat difficult to see what it is for a *proposition* to be self-contradictory, though. A *sentence* may be called a contradiction because of its form, but what is the form of a proposition? If it has a unique form, we cannot be sure it is the form of the sentence that expresses it, since a proposition may be expressed by sentences in many languages with many forms. To speak of what can be derived in a given formal system is again to treat a sentence rather than a proposition.

   I will cheat by helping myself to the assumption that counterpossible propositions have antecedents and consequents, and that these are the propositions expressed by the antecedent and consequent of any natural language sentence that adequately expresses the counterpossible proposition. [↑](#footnote-ref-3)
4. Or perhaps not context per se, but something that frequently varies with context, such as the intentions of the speaker or the speaker’s assumptions about which aspects of reality are now most relevant. I will continue to use the word ‘context’ to maximize the fit with Lewis’s terminology, but this may sometimes require us to take the word in a fairly broad sense. For example, on this usage it is possible for context to change suddenly in the middle of a conversation with a slight change of subject or emphasis. [↑](#footnote-ref-4)
5. One particular objection of this sort, the ‘future similarity’ objection, was raised by at least eight authors [Lewis 1983b: 43; Tichy 1976]. Some objectors were explicit in rejecting similarity’s ‘shiftiness’. Thus Eugene Schlossberger: ‘For similarity is otiose here unless either the relevant kinds of similarity and their appropriate weights are clearly demarcated in advance, or one can trust one’s ordinary intuitions about similarity.... Otherwise one might as well drop reference to similarity and rest content with an unanalysed ordering relation instead [1978: 82]’ . [↑](#footnote-ref-5)
6. Nolan makes a similar point without reference to similarity [1997:557-8]. I also want to call attention to his insightful suggestion there, which should be helpful to classical and paraconsistent logicians alike. To rephrase it somewhat, hypothetical reasoning should be understood via counterfactual conditionals rather than by the conditionals (of whatever sort) one normally uses in nonhypothetical reasoning. [↑](#footnote-ref-6)
7. When the text does not specify, it should be assumed that the word ‘counterfactual’ refers to ‘would’ counterfactuals. [↑](#footnote-ref-7)
8. This is Lewis’s preferred account [1973: 21]. Later (pp. 25-26) he considers an alternate, stronger ‘would’ conditional (A  C) which is never vacuously true, and so is false whenever the antecedent is impossible. This stronger ‘would’ conditional corresponds to a weakened ‘might’ conditional (A ◊ C) which is vacuously true when the antecedent is impossible. These two conditionals are interdefinable in the same way as the preferred pair. [↑](#footnote-ref-8)
9. Zagzebski takes note of this awkwardness [1990: 176]. [↑](#footnote-ref-9)
10. Assuming it is sensible to say that propositions have forms in the way that sentences do. See note 3. [↑](#footnote-ref-10)
11. The interdefinabililty is weighty enough to motivate Lewis to reject the mixed pair of ‘’ and ‘◊’ as connectives which together satisfy the intuitions one might have about the unentertainable [Lewis 1973: 25-6]. [↑](#footnote-ref-11)
12. Cf. Lewis [1983c], who is thinking primarily of utterances rather than propositions. [↑](#footnote-ref-12)
13. If the claim is correct, then adding impossible worlds to one’s semantics does not, or need not, eliminate vacuity [Cohen 1988:100]. [↑](#footnote-ref-13)
14. Mares says the supposition that all possible worlds are closer than any impossible world seems reasonable [1997: 521]; Nolan says it has a ‘fair bit of intuitive support’ but suspects that it has some exceptions [1997: 550]. Jonathan Strand defends the relative closeness of impossible worlds with some very natural-sounding examples [2001]. [↑](#footnote-ref-14)
15. I owe Patricia Blanchette thanks for some very helpful comments on an ancestor of this section. [↑](#footnote-ref-15)