"If not for" Counterfactuals: Negating Causality in Natural Language*

Robert Henderson
University of California, Santa Cruz

1. Introduction

Using new data from a counterfactual construction in English that has not yet been treated in the literature, this paper presents new arguments for the picture painted by Schulz (2007), where counterfactual antecedents are ambiguous between requiring global and local revision in models enriched with causal relations. Specifically, we investigate the counterfactual construction in (1) (henceforth NC for "not counterfactual"), compared to standard counterfactuals, exemplified in (2).

(1) If not for Mary going to the store, we wouldn't have salsa.

(2) If Mary hadn't gone to the store, we wouldn't have salsa.

While (1) and (2) appear to be good paraphrases of each other, NCs systematically differ from standard counterfactuals. First, unlike standard counterfactuals, which only implicate the counterfactuality of their antecedents, NC antecedents are non-defeasibly counter-to-fact. Secondly, NCs differ from standard counterfactuals in that they do not support certain non-causal epistemic inferences, including those based on backtracking and correlations. What we show is that these properties are precisely what one would expect of an obligatorily ontic counterfactual whose antecedent is interpreted with local revision in a model that incorporates causal relations (Pearl, 2000; Schulz, 2007:i.a.).

The analysis begins in §2 which details the morphosyntactic and semantic properties of NCs, especially where they diverge from standard counterfactuals. Section §3 presents the formal system, which enriches our models with causal relations. It also develops an analysis of the contribution of NC antecedents as local update within such causal models. Section §4 concludes.

2. Introducing a New Counterfactual

The counterfactual construction that is the empirical focus of this work consists of a nominal embedded under *if not for*. Although NCs allow both individual denoting nominals along with more complex eventive nominals, the analysis that follows will focus on NCs with clausal gerunds because doing so makes it easier to draw comparisons to standard counterfactuals which have clausal antecedents.

Beyond the embedded nominal, negation is the most important morphological fact because it is crosslinguistically stable. While the morphology of NCs varies across languages, negation is always present. Example (3), from Kaqchikel (Mayan), shows that there are NCs with nominal antecedents but no preposition, while the example (4) from Spanish has a preposition like English (in this case *porque* 'because'), but allows full clausal complements in place of nominals (which English and Kaqchikel have). In all cases, though, negation is obligatorily present, as is clear from the (b) examples.

(3) KAQCHIKEL

a. *Wi* ***man*** *ta Maria, yi-b'ison ta.*
If **NEG** IRR Maria, INFL-sad IRR
If not for Maria, I would be sad.

*I need to thank Pranav Anand, Scott AnderBois, Ryan Bennett, Adrian Brasoveanu, Daniel Büring, Ivano Caponigro, Sandy Chung, Lily Clark, Cleo Condoravdi, Sam Cumming, Amy Rose Deal, Donka Farkas, Olga Kagan, Ed Keenan, Sven Lauer, Chris Potts, Kyle Rawlins, Magda Sotz Mux, and Luis Vicente for helpful comments on this work. I also need to thank audiences at UC Santa Cruz S(yn/em)-Circle (2008), CUSP I at UCLA (2009), and the N3C Workshop at Stanford (2009). This being said, all mistakes are my own.

Cascadilla Proceedings Project
Completed September 28, 2010

b. *Wi ta Maria, . . .

(4) SPANISH

a. *Si **no** fuera porque María ha ido a la tienda, no tendríamos salsa.*
If **NEG** be.SBJ because Maria had gone to the store, NEG have.COND.PST.1pl salsa
If not for Maria going to the store, we wouldn't have salsa.

b. *Si fuera porque María ha ido a la tienda, . . .

The same is true for English NCs as in (5)

(5) a. If not for Mary, I wouldn't have passed the test.
b. *If for Mary, I would have passed the test.

These data show that negation is a critical part of NCs, both in English, and crosslinguistically (see Ippolito & Su (to appear) and Nevins (2002) for similar data in Mandarin and Tagalog respectively). Negation and nominals play an important role in the analysis that follows because we will take the latter to denote a fact, while negation removes that fact to move to counterfactual contexts for the evaluation of the consequent.

2.1. The Core Semantic Generalizations

This section presents the two core semantic generalizations that receive an account in §3. First, NCs, unlike standard counterfactuals, have antecedents that are non-defeasibly counter-to-fact. The second is that while standard counterfactuals are ambiguous between ontic and epistemic readings, NCs systematically lack epistemic readings.

2.1.1. Non-defeasibly counterfactual antecedents

Both NCs and standard counterfactuals generate inferences that the antecedent and consequent do not hold in the world of evaluation. It is well known that these inferences are only implicatures in standard counterfactuals. The reason is that counterfactuals can be used to argue for the proposition in the antecedent, as well as to conduct informative modus tollens arguments (Anderson, 1957; Stalnaker, 1975). Both of these arguments fail for NCs.

(6) If John were not sick with the measles, he would have the same temperature he does now; therefore, I conclude he doesn't have the measles.

(7) #If it were not for John being sick with the measles, he would have the same temperature he does now; therefore, I conclude he doesn't have the measles.

The consequent in (6) cannot be counter-to-fact because its truth is taken as evidence that the antecedent is also true in the actual world. The next example shows that modus tollens arguments with NCs are infelicitous. This would make sense if the proposition making the antecedent counterfactual is already in the common ground.

(8) If the butler hadn't carved the turkey, the knife would have been sharp. The knife was dull; therefore, the butler carved the turkey.

(9) #If not for the butler having carved the turkey, the knife would have been sharp. The knife was dull; therefore, the butler carved the turkey.

If it is true that NCs are only admissible if the antecedent is counterfactual with respect to a proposition already in the common ground, then it should not be possible to felicitously use an NC when the requisite proposition is missing. Example (10) shows that this is the case. An NC cannot be used in a modus tollens argument if the antecedent is still under discussion.

(10) **A:** John went to the store. **B:** No he didn't.

(11) **A:** (He did!) If John hadn't gone to the store, he would be home right now (and he isn't).

(12) **A:** #(He did!) If not for John going to the store, he would be home right now (and he isn't).

The response in (12) is infelicitous because it ignores the fact that whether or not John went to the store is under discussion. It cannot be in the common ground because of speaker B's denial. To summarize, the generalization is that NCs are only licit if their antecedents are presupposed to be counter-to-fact.

Note that the tests used to diagnose the strong counterfactuality of NC antecedents involve epistemic readings of counterfactuals. For example, both modus tollens arguments and arguing for the antecedent require a *detective* context, where the truth value of the antecedent is not known. In this context, it is possible to ask whether the current epistemic state would match the epistemic state generated by learning the that the antecedent is true; if yes, there is evidence that the counterfactual antecedent is actually true, if not, it is is false by modus tollens. In this way, NCs behave as expected if they were obligatorily ontic. Their antecedents are non-deasibly counterfactual because they presuppose a fact about the world that the antecedent can revise through negation. In the next subsection, we consider more epistemic inferences that are unavailable with NCs, but in these cases, the problem is that there is no causal link between antecedent and consequent.

2.1.2. Missing epistemic inferences

Many classic epistemic readings for counterfactuals arise in situations where the antecedent does not causally antecede the consequent. The reason is that changing the world to accord with the antecedent, as ontic counterfactuals require, does not necessarily affect the consequent since it is not a causal consequence of such a change. In opposition, learning the information in the antecedent changes an agent's epistemic state, which can permit a non-causal inference about the consequent. The two non-causal epistemic inferences to be considered here involve *backtracking* and *correlations*. What we find is that while standard counterfactuals support such inferences, NCs do not.

As Arregui (2004) notes, backtracking requires a context supporting either a causal or analytic relation between the antecedent and consequent. Backtracking works because information about the consequences of a relation can be used to make epistemic inferences about its cause. For example, if we conceive of the marriage ceremony causally affecting an individual's bachelorhood status through the analytic relation supplied by the predicate *bachelor*, the backtracking inference in (13) is licensed.

(13) If John weren't a bachelor, he would have to have had a marriage ceremony.

The inference in (13) proceeds as follows: Upon learning counterfactually that John is not a bachelor, one can infer that he would have had a marriage ceremony due to the definition of bachelors as unmarried males.

Example (14) presents another case of backtracking, but this time it is supported by the causal connection between rain and the functionality of the car's wiring.

(14) Suppose that the car works perfectly except for after a rain, which always causes the wiring to short out. Suppose it didn't rain and the car started as usual.

a. If the car hadn't started, it would have to have rained.

The reason we want to treat inferences like (14) as epistemic is that intervening to change the car's state does not cause it to have rained, but learning that the car does not work can allow the inference that it had rained. This explains the contrast with (15), which does not license the backtracking reading.

(15) #If I had made the the car not start, it would have to have rained.

When the speaker causally intervenes on the state of the car, the backtracking epistemic reading disappears.

Since backtracking is only possible under epistemic readings of counterfactuals, the possibility of backtracking can be used as a test for the availability of epistemic readings. The crucial observation is that NCs pattern with examples like (15), and not standard counterfactuals like (13-14), with respect to backtracking. This is presented in (16-17).

(16) #If not for John being a bachelor, he would have to have had a marriage ceremony.

(17) #If not for the car starting, it would have to have rained.

Examples (16-17) do not permit a backtracking reading. They can only be true in the implausible situation where John's bachelorhood prevented him from being obliged to marry and the car's starting had an effect on local weather patterns.

The backtracking test shows that NCs do not have an epistemic reading, but are obligatorily ontic. A second type of epistemic inference forces the same conclusion. As discussed by Schulz (2007), standard counterfactuals are systematically ambiguous between ontic and epistemic inferences between two correlated variables. Correlations occur when two effects are anteceded by the same cause. When reasoning epistemically between two such variables, the correlation is maintained; that is, learning something about one variable provides information about the other. This is not the case for ontic interpretation because intervening to change one of the variables destroys the correlation. Schulz (2007) presents dialogues like (18), which show the ambiguity for standard counterfactuals.

(18) Suppose tempests are correlated with a low barometer. Further suppose we took the bridge instead of the ferry because of the low barometer and there was, in fact, a storm.

a. Thank goodness, if the barometer hadn't been low, we would have taken the ferry and we might have all drowned in the storm.

b. No no no, if the barometer hadn't been low, there wouldn't have even been a storm.

Dialogues like (18) are possible because standard counterfactuals are ambiguous between ontic and epistemic readings. Example (18a) presents the ontic reading, where the barometer reading is locally changed, independent of the storm and other correlated variables. The epistemic counterfactual in (18b) allows for the value of the storm to change because we consider, not how the world would be different if the barometer had been forced to be different, but what we would know had we observed a different value on the barometer. In the latter case, we would know that there would have been no storm.

The ambiguity of counterfactuals in the context of correlations provides another test for epistemic readings. If a counterfactual allows both an epistemic and an ontic reading, it should be able to appear in both sides of an argument like that in (18). If it does not have an epistemic reading, it should only be felicitous as the ontic argument. NCs instantiate the latter possibility as seen in (19).

(19) Suppose tempests are correlated with a low barometer. Further suppose we took the bridge instead of the ferry because of the low barometer and there was, in fact, a storm.

a. Thank goodness, if not for the low barometer, we would have taken the ferry and we might have all drowned in the storm.

b. #No no no, if not for the low barometer, there wouldn't have even been a storm.

The reply in (19b) can only be true under the strange situation where the low barometer *caused* the storm, that is, they are no longer merely correlated. Once again, we find that NCs and standard counterfactuals are mutually paraphrasable under an ontic reading (18a-19a), but that NCs have no epistemic reading. When placed in a context where such a reading is forced, they are either false or aberrant depending on whether a causal relation can be constructed between antecedent and consequent, allowing the possibility of an ontic reading.

3. The Semantics of NCs

This section uses the nonstandard morphology of NCs to build a semantics that captures the generalizations presented in the previous section. Specifically, it argues that the eventive nominal embedded in an NC antecedent denotes a set of situations that exemplify a proposition (Kratzer, 2002). Due to a presupposition contributed by the preposition *for*, one of these situations is presupposed to be contained in the world of evaluation, which explains the non-defeasible counterfactuality of NC antecedents. Finally, the obligatory and eponymous negation of NCs is treated as constituent negation that removes the fact denoted by the nominal. Since this "fact removal" negation is interpreted as a local update, we can account for the fact that NCs reject certain non-causal epistemic inferences after enriching our models with causal laws.

3.1. Worlds, Situations, and Causal Models

The analysis employs a possible world semantics. Instead of treating possible worlds (situations) as indices for an interpretation, they will be (partial) interpretations themselves.

(20) **Definition. (Worlds and Situations)**
Let $\mathcal{P}$ be a finite set of atomic sentences.

i. A *world* w is a function from $\mathcal{P} \rightarrow \{0,1\}$.
ii. A *situation* s is a partial function from $\mathcal{P} \rightarrow \{0,1\}$.

We work with the language $\mathcal{L}$ that is the closure of $\mathcal{P}$ under negation and conjunction, giving the usual recursive truth definitions for complex formula.

(21) **Definition. (Models and Propositions)**

i. A *model* M for $\mathcal{L}$ is a set of worlds W. For any formula ϕ of $\mathcal{L}$, we write $M, w \vDash \phi$ in case ϕ is true with respect to M and w.
ii. Write ϕ^M for the set of worlds $w \in W$ s.t. $M, w \vDash \phi$, and call this the *proposition* that ϕ.

The discussion of the distinction between ontic and epistemic counterfactuals in the previous section showed the importance of the causal structure of the context, which we model via *Causal Structures*. This is a partition of the atomic formula into dependent and independent sets, along with a set of functions that allows the values of the dependent atoms to be determined based solely on the values of the independent atoms. These functions can be thought of as encoding the causal relations.

(22) **Definition. (Causal Structures and Models)** (Schulz 2007, pg. 141)
Given a finite set of atomic sentences $\mathcal{P}$, a model is a set of worlds W along with a causal structure $C = \langle B, E, F \rangle$ where:

i. $B \subseteq \mathcal{P}$ are *exogenous* variables.
ii. $E = \mathcal{P} - B$ are *endogenous* variables.
iii. F is a function mapping elements Y of E to tuples $\langle Z_Y, f_Y \rangle$, where Z_Y is an n-tuple of $\mathcal{P}$ and f_Y is a partial truth function $f_Y : \{0,1\}^n \longrightarrow \{0,1\}$. F is *rooted* in B.

The structure of the causal relations is given by the final clause (22iii). Every endogenous atom is associated with a set of atoms that it causal depends on and a partial interpretation that gives its value depending on the values of those atoms that it causally depends on. Rootedness ensures that when looking at the variables associated with an endogenous variable by F, if they are not exogenous, their F relations can be followed all the way back to exogenous variables (see Schulz (2007, pg. 104 and 141) for more discussion of the rootedness property).

The truth value of a formula from the set of background variables is simply its value at the world of evaluation as seen in (23a). The new case is presented in (23b). If ϕ is endogenous, then its truth value is a function of the causal laws f_ϕ applied to the causal antecedents of ϕ, Z_ϕ.

(23) a. $\phi^{M,w} = w(\phi)$, if $\phi \in B$
b. $\phi^{M,w} = f_\phi(Z_\phi)$, if $\phi \in \mathcal{P} - B$

What the causal models just defined provide is a more fine grained notion of law against which counterfactuals can be evaluated. We can then follow Veltman (2005) in distinguishing laws from facts via the notion of *basis*, which is the smallest set of facts that along with the laws, derives all the other facts that characterize a world. Schulz (2007) defines the basis for a causal world as follows:

(24) **Definition. (Basis)** (Schulz 2007, pg. 144)

i. The basis b_w of a world $w \in W$ is the union of all interpretation functions $b \in I$ that fulfill the following two conditions:
a. $b \subseteq w \subseteq \bar{b}$

b. $\neg\exists b' : b' \subseteq w \subseteq \bar{b'}$ & $b' \subset b$, where $\bar{b}$ is the closure of b under the causal laws.

A basis can be thought of as encoding the facts that characterize a world. It is the smallest interpretation function that, along with the causal laws, can derive all there is to know about a world. When interpreting a counterfactual, the consequent must be verified only in those worlds that (i) satisfy the antecedent, (ii) have minimally different bases, and (iii) are minimally different in the laws that hold. Example (25) defines a notion of similarity that fulfills these requirements.

(25) **Definition. (Similarity)** (Schulz 2007, pg. 145)

i. Define $\leq$ mapping w to the order: for $w_1, w_2 \in W$: $w_1 \leq_w w_2$ iff

a. $b_{w1} \cap b_w \supseteq b_{w2} \cap b_w$

b. if $b_{w1} \cap b_w = b_{w2} \cap b_w$, then $b_{w1} - b_w \subseteq b_{w2} - b_w$.

The order defined by $\leq_w$ says that a world w_1 is more similar to w than w_2 iff its basis overlaps more with w or, in case they are equal in this respect, the part of its basis that does not overlap is smaller than that of w_2. The fact that it is smaller means that more of the causal laws are left intact since fewer stipulated facts are needed to characterize the world. Since similarity is defined in terms of a world's basis, and the basis is dependent on the flow of causal laws, locally changing upstream variables counts more for the purposes of similarity than those downstream. This is precisely the distinction we need to account for the lack of epistemic readings with NCs.

3.2. The Evaluation of NCs

Following an idea from Kratzer (2002), eventive nominals in NCs will denote situations that *exemplify* a proposition. Such situations contain all and only the information relevant for the truth of a proposition. In order to define exemplification clearly, a few other useful notions must be defined, starting with *forces* (Veltman, 2005).

(26) **Definition. (Forces and Minimally Forces)**

i. Say that s *forces* ϕ iff $\forall w(s \subseteq w \rightarrow w \in \phi)$

ii. Say that s *minimally forces* ϕ iff

a. s forces ϕ and there is no $s' \subseteq s$ such that s' forces ϕ

Now we can define a situation that exemplifies a proposition as a situation that forces a proposition containing no superfluous information. We filter off this information as in (27).

(27) s *exemplifies* ϕ iff for all $s' \subseteq s$ such that s' does not force ϕ, there is an s'' such that $s' \subseteq s'' \subseteq s$ and s'' minimally forces ϕ.

We can now define the Kratzerian minimalization operator $\downarrow$ that takes a formula ϕ and maps it to the set of situations that exemplify the proposition ϕ. In what follows, the analysis assumes (though it is surely a oversimplification), that eventive nominals uniformly denote situations that exemplify a proposition. Thus, the operator $\downarrow$ can be thought of as the semantic counterpart to a nominalization operator, and along these lines, the nominalization of a proposition ϕ will be translated as $\downarrow\phi$.

Section §2 showed that NCs cannot be used in informative modus tollens arguments or in arguments for the antecedent. It was suggested that this was due to a presupposition of counterfactuality, but now that we have a semantics for the nominals in NC antecedents, the locus of this presupposition can be found and its introduction modeled explicitly. Specifically, we propose that the preposition *for* imparts a factive presupposition to its nominal complement, which is supported by the fact that NCs in other languages also employ factive morphology.[1] For instance, Spanish NCs, as exemplified in (4), contain

[1] In considering Mandarin NCs, Ippolito & Su (to appear) argue that the factive presupposition is contributed by light negation (Schwarz & Bhatt, 2006). For the purposes of this paper, both approaches work equally well, and so we will not try to distinguish them here. Their approach potentially better generalizes across languages since all NCs contain negation, though the fact that NCs sometimes contain "reason" prepositions, which are often factive, is still telling and left unexplained in their analysis, though not here.

the preposition *porque* 'because', which imposes a factive presupposition on its clausal complement. In the framework developed here, the presupposition would project past negation and require a situation in the denotation of the gerund be contained in the world of evaluation for the sentence to be admissible. Since every situation in the denotation of the gerund exemplifies a proposition, this proposition will be presupposed to be true, accounting for the non-defeasible counterfactuality of NC antecedents.

This discussion prefigures the analysis of negation, which brings together the treatment of eventive nominals in NC antecedents and the implementation of premise semantics using causal models. First, note that negation in English NCs cannot be the standard propositional operator; there is no proposition for it to operate on. Instead, negation must provide a bridge from the denotation of eventive nominals to possible worlds. The reason is that only a set of worlds can entail the consequent, and since NC consequents are indistinguishable from the consequents of standard counterfactuals, we should make the standard assumption that the truth of a *would*-counterfactual depends on whether the consequent is entailed by the context after it is updated with the antecedent. With these concerns in mind, NC negation is interpreted as the model update function in (28) that takes a world and a fact that exemplifies a proposition and returns the most similar worlds where that fact has been removed. Similarity is calculated with the help of the function $Sim_{w,\leq}$ in (28b), which takes a set of worlds and gives the subset that is the most similar with respect to a world w.[2]

(28) **Definition. (Remove$_{\mathbf{M}}$)**

a. $Remove_M(w, \downarrow\phi) = Sim_{w,\leq}(\{w' \mid \forall s(s \in \downarrow\phi^M \rightarrow s \nsubseteq w')\})$
where given a set of worlds U:

b. $Sim_{w,\leq}(U) = \{w' \in U \mid \neg\exists w''(w'' \in U \ \ \& \ \ w'' \leq_w w')\}$

The definition in (28a) says that *Remove*$_M$ is a function from the facts exemplifying ϕ at w to a set of worlds without these facts that are, due to the definition of $\leq_w$, maximally similar to w in exogenous facts (from which all others can be derived by the laws). Note that the update defined above is inherently local. It revises a world to remove facts.

Since NCs obligatorily contain light negation and a nominalization, NC antecedents will necessarily require the local update defined in (28). We define their interpretation now. For simplicity, we will not allow for arbitrarily embedded NCs.

(29) **Definition. (Language with *would*-NCs)**
Give a set of propositional letters $\mathcal{P}$ and our language $\mathcal{L}$, which is the closure of $\mathcal{P}$ under negation and conjunction, the language $\mathcal{L}^{\succ}$ is the union of $\mathcal{L}$ with the set of sentences of the form $\downarrow\phi \succ \psi$ for all $\phi, \psi \in \mathcal{L}$.

We interpret the language of NCs in our causal models as defined in (22).

(30) **Definition. (Interpretation of *would*-NCs)**
Let M be a causal model and $w \in W$ a possible world.

a. $M, w \vDash \downarrow\phi \succ \psi$ iff $Remove_M(w, \downarrow\phi) \vDash \psi$

It is now possible to see how this correctly predicts the restricted class of readings that are available with NCs. All epistemic readings that require an inference against the flow of causality will be necessarily suppressed.

Recall that NCs reject backtracking inferences, like that in (14). The standard counterfactual allows an inference about the state of rain at some previous time based on the state of the car (14a). This reading is necessarily false with the NC (17); it seems odd because it can only be true under the strange reading where the car starting prevented it from raining in the past. The background situation in (14) is such that the rain R and whether the car starts S are causally connected. This state of affairs can be represented as in (31). The model will associate the variable S with the function f_S that gives the value of S as a function of R.

[2] $Sim_{w,\leq}$ is just a particularized version of Schulz's (2007) more general minimization function (*Min*) built for the situation at hand.

(31) R causally antecedes S

$$f_S(R) : \begin{cases} 1 \to 0 \\ 0 \to 1 \end{cases}$$

The truth table in (32) presents the universe that covers the logically possible truth values of R and S. Recall that the $\boxed{\text{box}}$ indicates the basis of each world given the causal structure. For instance, the basis of w_2 is $\langle R, 1\rangle$ since we can derive $\langle S, 0\rangle$ from the causal laws. World w_4 where the causal laws are broken must concomitantly have a larger basis. The pair $\langle S, 1\rangle$ has to be represented in the basis because it cannot be derived from the laws due to the fact that this set of facts violates the laws.

(32)

	R	*S*
w_1	$\boxed{0}$	$\boxed{0}$
w_2	$\boxed{1}$	0
$@w_3$	$\boxed{0}$	1
w_4	$\boxed{1}$	$\boxed{1}$

In the scenario above, world w_3 corresponds to the actual world because it did not rain and the car, in fact, started. To interpret the counterfactual, one must consider maximally similar worlds where the car did not start. There are two such worlds, namely w_1 and w_2. Notice that w_2 supports the backtracking inference, while w_1 does not. Under an epistemic reading, the evaluation of the antecedent would pick out world w_2 as the most similar. Under the ontic reading, the consequent would be interpreted in world w_1. Crucially, the analysis of NC antecedents correctly predicts that w_1 is the closest. The reason is that w_1 retains more facts from the basis of w_3, per clause (25i-a) in the definition of similarity.

Specifically, interpreting the antecedent $Remove_M(w_3, \downarrow S)$ returns the set of worlds that are closest to w_3 with respect to $\leq_{w_3}$ that do not contain the situation exemplifying S, namely $\langle S, 1\rangle$. The worlds where this holds are ranked as in (33).

(33) Similarity: $w_1 <_{w_3} w_2$

The derivation in (34) shows that the backtracking inference cannot go through with NCs.

(34) If not for the car starting, it would have to have rained^{M,w_3} = 1 iff

a. $M, w_3 \vDash \downarrow S \succ R$ iff
b. $Remove_M(w_3, \downarrow S) \vDash R$ iff
c. $Sim_{w_3,\leq}(\{w' : \forall s(s \in \downarrow S^M \to s \nsubseteq w'\}) \vDash R$ iff
d. $Sim_{w_3,\leq}(\{w_1, w_2\}) \vDash R$ iff
e. ✗ $w_1 \vDash R$

The analysis not only predicts that backtracking inference is necessarily unavailable, it also predicts that the only true reading is the strange case where the car starting causally antecedes it raining. Notice that switching the causal relationship between R and S generates the new bases in (35).

(35)

	R	*S*
w_1	$\boxed{0}$	$\boxed{0}$
w_2	1	$\boxed{0}$
$@w_3$	0	$\boxed{1}$
w_4	$\boxed{1}$	$\boxed{1}$

The only difference is that now knowing whether the car started is sufficient to completely characterize worlds that follow the causal laws, namely w_2 and w_3. While only a minor change, it makes all the difference for the interpretation of a local revision function, like that contributed by NC antecedents. Now the minimally different world without the fact $\downarrow S$ is w_2.

(36) Similarity: $w_2 <_{w_3} w_1$

The reason is that both w_1 and w_2 are equally bad with respect to preserving the basis of the actual world, but w_2 is more similar than w_1 to the actual world because its basis is smaller with respect to the actual world, that is, more of the causal laws that hold in the actual world are preserved (see (25i-b) for the formal calculation). The inference now goes through.

(37) If not for the car starting, it would have to have rained^{M,w_3} = 1 iff

a. $M, w_3 \vDash \downarrow S \succ R$ iff

b. $Remove_M(w_3, \downarrow S) \vDash R$ iff

c. $Sim_{w_3,\leq}(\{w' : \forall s(s \in \downarrow S^M \rightarrow s \nsubseteq w'\}) \vDash R$ iff

d. $Sim_{w_3,\leq}(\{w_1, w_2\}) \vDash R$ iff

e. ✓$w_2 \vDash R$

The prediction is that NCs in backtracking context can be true, only if the context is restructured so that the antecedent causally antecedes the consequent, eliminating the backtracking. The NC will be felicitous in so much as such this causal structure is plausible. The prediction is borne out, and (37) is infelicitous because it can only be true if the state of the car's engine can causally affect the weather. Although there is not the space to consider the example with correlations, the analysis extends. Just like with the backtracking, locally altering a fact makes that fact causally independent of the rest of the model. It is exactly this that blocks backtracking, correlations, and other epistemic inferences with NCs.

4. Conclusions

We showed that English counterfactuals are not monolithic. There exists a species of counterfactual, namely the NC, which systematically differs from standard counterfactuals in resisting non-causal epistemic inferences. This lead us to propose a semantics for the obligatory negation accompanying NCs that would ensure that their antecedents could only be evaluated with local revision in a causal model. Specifically, NC antecedents presuppose a fact that negation removes by minimally altering the world. First, since NC antecedents carry a factive presupposition, we can account for the fact that their antecedents are non-defeasibly counterfactual. Second, since minimally altering the world is different than minimally altering a belief state, we can capture the fact that NCs have only a subset of the readings available with standard counterfactuals if the latter are ambiguous between both types of update, as argued for by Schulz (2007). When there is an antecedent that is counter-to-fact, and an inference that does not violate any causal laws, NCs and standard counterfactuals will be mutually paraphrasable, which is exactly what the data show.

References

Anderson, Alan Ross (1957). A theory of subjunctive conditionals. *The Journal of Symbolic Logic* 22:3.

Arregui, Ana (2004). *On the accessibility of possible worlds*. Ph.D. thesis, UMass Amherst.

Ippolito, Michela & Julia Su (to appear). Counterfactuals, negation, and polarity. *Proceedings of NELS*.

Kratzer, Angelika (2002). Facts: Particulars or information units? *Linguistics and Philosophy* 25:5, 655–670.

Lewis, David (1973). *Counterfactuals*. Harvard University Press.

Lewis, David (1979). Counterfactual dependence and time's arrow. *Noûs* 13:4, 455–476.

Nevins, Andrew Ira (2002). Counterfactuality without past tense. Hirotani, Masako (ed.), *Proceedings of North East Linguistic Society 32*, vol. 2, 441–450.

Pearl, Judea (2000). *Causality: Models, reasoning, and inference*. Cambridge UP.

Ramsey, Frank P. (1950). Foundations of Mathematics and other Logical Essays, chapter General Propositions and Causality. *New York* 237–257.

Schulz, Katrin (2007). Minimal models in semantics and pragmatics: Free choice, exhaustivity, and conditionals. *ILLC Dissertation Series* .

Schwarz, Bernhard & Rajesh Bhatt (2006). Light negation and polarity. *Crosslinguistic research in syntax and semantics: negation, tense, and clausal architecture* p. 175.

Stalnaker, Robert (1975). Indicative conditionals. *Philosophia* 5:3, 269–286.

Veltman, Frank (2005). Making counterfactual assumptions. *Journal of Semantics* 22:2, p. 159.