**Backtracking Counterfactuals and iterated modalities**

*Ana Arregui - University of Ottawa*

Backtracking counterfactuals are counterfactuals that claim that if things had been different at some time **t**, they would also have been different at some earlier time **t'**. They appear to go against the view that the future depends on the past, and not vice-versa [as defended in Lewis 1979]. I propose an analysis of backtracking counterfactuals according to which they are examples of iterated modality: they are counterfactuals with a modalized consequent clause. The second layer of modality is introduced by a modal *have-to* auxiliary. Backtracking counterfactuals provide us with insights into the combinatorial possibilities of different types of modality: I argue that they may embed deontic modality, but not epistemic modality.

**Similarity à la Lewis 1979.** According to a Stalnaker-Lewis style analysis, a counterfactual conditional of the form 'if A, B' is true in the actual world iff the most similar A-worlds are also B-worlds. The interpretation of conditional statements involves quantification over possible world, and the quantificational domain of the modal is made up of the most similar worlds in which the antecedent is true [adopting the *Limit Assumption* for simplicity].

Lewis 1979 argued that the notion of similarity relevant to the evaluation of counterfactuals is subject to several constraints. The interaction between the constraints in the actual world ensures that similarity at earlier times counts for more than similarity at later times (similarity is 'past-sensitive'). By placing constraints on similarity, Lewis sought to counter Fine 1975's objections. Fine pointed out that (a naïve view of) similarity predicted that a conditional like (1) is false, whereas most people would judge (1) true.

(1) If Nixon had pushed the button, there would have been a nuclear holocaust. [Fine 1975]

Worlds in which somebody cut the electricity cable before Nixon pushed the button are (naively) more similar to the actual world than worlds in which the signal got through and a holocaust followed. Lewis's constraints conspire to make sure that the quantificational domain of the modal is made up of worlds in which the past is kept "the same" as in the actual world at times preceding the eventuality described by the antecedent clause.

**The puzzle: backtracking counterfactuals.** Lewis's proposal appears to be challenged by backtracking counterfactuals. Lewis notes that backtracking counterfactuals are often helped by 'special syntax': a combination of *would* and *have to* in the consequent clause, as in (2).

(2) (Jim and Jack had a big quarrel yesterday, and Jack is still very angry.)
*If Jim were to ask Jack for help today, there would have to have been no quarrel yesterday.*[Lewis 1979: 33]

I provide an analysis of backtracking counterfactuals that respects Lewis's view that similarity is past-sensitive. I argue that the second auxiliary introduces a second layer of modality that is not past-sensitive. My analysis explains the following generalizations:

**Generalization 1**: Backtracking counterfactuals with *regular* syntax can be judged straightforwardly true if the relation between antecedent and consequent is not contingent (similarity plays no role).

(3) a. If he were a bachelor, he wouldn't have married.
b. If she had a twin sister, her mother would have had at least two children.
c. If she had sold a horse, she would have owned a horse.

**Generalization 2:** Backtracking counterfactuals with special syntax can be judged (more or less) straightforwardly true if some salient law-like generalization establishes a relation between antecedent and consequent.

(4) A: It's lucky the guard didn't push the alarm button. It would have been a false alarm.
B: a. ?Actually, he is a very intelligent man. *If he had pushed the button, something serious would have happened.*

b. Actually, he is a very intelligent man. *If he had pushed the button, something serious would have to have happened.*

We judge the conditional in (4a) to be odd or false. But the conditional in (4b) makes us think about the fact that the guard would never push the button unless something serious had happened. The use of *have to* invokes this regularity.

**Generalization 3:** In cases in which there is no (contextually salient) law or regularity to be invoked by *have*, special syntax does not help.

(5) The bridge wasn't completed, and the driver came to a sudden stop.
   a. ?*If the driver had kept going, the bridge would have been completed.*
   b. ?*If the driver had kept going, the bridge would have to have been completed.*

There isn't a salient relation between the condition that the driver kept going, and the completion of the bridge. Drivers may choose to drive off uncompleted bridges. Backtracking is unsuccessful, even in the presence of special syntax.

**The role of special syntax:** Stowell (2003) refers to *have-to* in (7) as a 'semi-modal'. I treat modal $have_{mod}$ as a universal quantifier over possible worlds (see also von Fintel & Iatridou 2005). Statements with $have_{mod}$ claim that some proposition holds in every possible world for some restricted domain. As with *would*, the quantificational domain of $have_{mod}$ is established by a contextually salient accessibility relation. Different contexts make salient different accessibility relations. Some examples are given in (6):

(6) a. She has to know. [epistemic modality]
   b. She had to pay a fine. [deontic modality]

The type of relation associated with $have_{mod}$ differs from that associated with *would*. The worlds quantified over by $have_{mod}$ are not identified purely on the basis of overall similarity:

(7) $[[have_{mod}]](P_{<i, <s,t>>>})(t)(w) = 1$ iff
$\forall w' [R_{<i, <s, <s, t>>>>}(t)(w)(w') \rightarrow \exists t' [(t=t' \text{ or } t<t') \& P(t')(w')]$,
where R is some contextually salient temporally sensitive accessibility relation.

According to (7), $have_{mod}$ combines with a property of times, locating it at some non-past time (depending on aspectual properties of the antecedent clause eventuality). The temporal argument of *have* provides the evaluation time for the accessibility relation. If we are dealing with deontic modality, the temporal argument sets the time at which the obligation held. [Condoravdi 2002 discusses the relation between type of modality and temporal parameters].

**Two layers of modality.** Backtracking counterfactuals are not counterexamples to Lewis's claim that we give the past special prominence when evaluating similarity. They should be understood as a combination of two types of modality. The standard semantics for *would* takes us to the most similar worlds in which the antecedent is true (in Lewis's sense of *similar*). The consequent clause then makes a modal claim about such worlds: if the law were obeyed, some earlier time would have been different. I provide arguments supporting the view that $have_{mod}$ in backtracking counterfactuals is sensitive to law-like regularities, and does not embed epistemic modality. The difference between *would* and $have_{mod}$ is that the similarity relation responsible for identifying the quantificational domain of *would* is past-sensitive, whereas the accessibility relation responsible for identifying the quantificational domain of $have_{mod}$ is not.

Condoravdi, C. 2002, Temporal interpretation of modals, in David Beaver et al, eds.,
Fine, K. 1975. Critical notice of Lewis 1973. *Mind* 84: 451-458.
Von Fintel, K & S. Iatridou, 2005, Anatomy of a modal, ms. (forthcoming LI)
Lewis, D. 1973. *Counterfactuals*. Harvard University Press, Cambridge, MA.
Lewis, D. 1979. Counterfactual dependence and time's arrow. *Noûs* 13: 455-476.
Stalnaker, R. 1968. A theory of conditionals. In *Studies in Logical Theory*, ed. N. Rescher.
Stowell, T. 2003. *Tense and Modals* (ms.)