Informative Counterfactuals

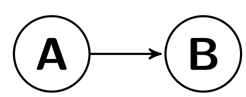
Synopsis. There is a rich body of literature devoted to the *truth conditions* for counterfactual conditionals, but very little work on the *informative use* of such expressions. How does the felicitous utterance of a counterfactual update the common ground? How do we understand a counterfactual like (1), which can have a variety of possible explanations? We use structural equation models (SEMs) to capture both the truth conditions and the update mechanism of counterfactuals in a way that embraces their explanatory underspecification.

(1) If Alice had gone to the party, Bob would have stayed home.

Background. In recent years, causal modeling has risen to prominence in the analysis of counterfactuals. SEMs represent causal (in)dependencies graph-theoretically, with the resulting structure playing a crucial role in the determination of "counterfactual shifts". While there is still disagreement as to the exact implementation of such shifts (Pearl 2000; Hiddleston 2005), the fundamental theory has engendered much interest in the cognitive sciences generally and, more recently, in linguistics specifically (Kaufmann 2013).

Counterfactual update and explanation. We adopt the basic SEM framework, though we treat the attendant graph-theoretic structure not as a given, but as a *goal*. In particular, we model updating with a counterfactual as a two-stage process: first, a set of "partial" SEMs are constructed that, in a precise sense, capture (various explanations for) the given counterfactual; second, these partial SEMs are used to pare down (an enriched version of) the common ground. Thus, we separate the task of *understanding* a counterfactual from that of *integrating* this understanding with an extant body of knowledge.

Consider again the counterfactual conditional in (1), which asserts a certain kind of covariance but provides no explicit mechanism to explain it; perhaps the most straightforward way to capture it with a SEM is shown in Figure 1, which encodes a direct causal dependence of B on A. But this representation is unsatisfying: Alice's attendance at the party need not be considered the immediate and sole cause of Bob's staying home in order to felicitously utter (1), as evidenced by the follow-ups in (2)–(4).



$B = \neg A$

Figure 1: Direct dependence.

(2) Bob hates Alice, and therefore avoids her.

(3) Alice and Bob have a young son Doug, and always arrange in advance (by flipping a coin, say) for at least one of them to be home to watch him.

(4) Alice brings her cat wherever she goes, and Bob is deathly allergic to cats.

In other words, there are many possible *explanations* for why Alice's attendance results in Bob's absence. Formally speaking, we associate with a counterfactual like (1) a set of SEMs denoted $[A \triangleright \neg B]$, each of which encodes a relationship between A and B that is compatible with an utterance of (1). In this regard, we propose a tripartite typology of explanation: ADDITIONAL CAUSE, COMMON CAUSE, and INTERMEDIATE CAUSE, exemplified respectively by the SEMs in Figures 2–4, which in turn correspond to the follow-ups (2)–(4). Context or pragmatics might favor one type of explanation over another, but in general, hearers may entertain multiple explanations for a given counterfactual.

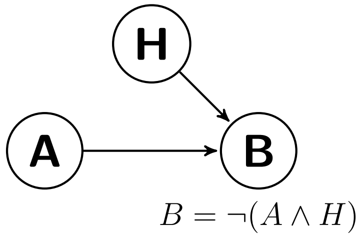


Figure 2: Additional cause.

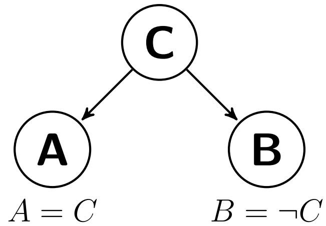


Figure 3: Common cause.

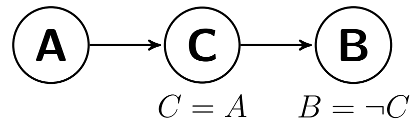


Figure 4: Intermediate cause.

Backtracking. Note that *reversing* the causal relationship of the direct dependency of Figure 1, as in Figure 5 (so that A depends causally on B), is rejected as an interpretation of (1), though it can be licensed with a double-auxiliary construction as in (5).

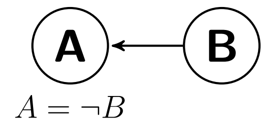


Figure 5: A classical backtracking explanation.

(5) If Alice had gone to the party, Bob would have had to have stayed home.

This suggests that the set of explanations $[A \rhd \neg B]$ should exclude SEMs in which B is an ancestor of A (i.e., SEMs in which a directed path can be traced from B to A). This in turn yields a simple formalization of the effect of the double-auxiliary: namely, the relaxation of this exclusion. This implementation also explains the (marked) availability of a non-backtracking interpretation (e.g., that of Figure 1) for a double-auxiliary construction like (5): relaxing the restriction does not exclude other interpretations, but it does make the double-auxiliary a strange way to encode such a meaning, from a pragmatic perspective.

The connection between such "backtracking" counterfactuals and the double-auxiliary construction has been studied (Iatridou 2000; Arregui 2005); notably, however, even though COMMON CAUSE is also typically taken to be a case of backtracking (as it requires reasoning about things "upstream" of the antecedent), it is available without the double-auxiliary, suggesting an important distinction between the backtracking in Figure 5 (where B is an ancestor of A) versus that in Figure 3 (where it is not). Only by investigating the internal structure of explanations does this distinction become apparent.

Jointly incompatible counterfactuals. Finally, consider the counterfactual in (6).

(6) If Alice had gone to the party, Doug would have been home alone.

Each of (1) and (6) is acceptable on its own (at least in a context where it is understood that Alice and Bob have a son Doug), but taken together, they are problematic. We show that our framework predicts both their individual tenability and their joint infelicity, the latter a consequence of how counterfactual explanations generated in the first stage of the update process are integrated into the common ground.

References. Arregui, Ana. 2005. Layering modalities: The case of backtracking counterfactuals. Accessed December, 2013 • Hiddleston, Eric. 2005. A causal theory of counterfactuals. *Noûs* 39(4). 632–657 • Iatridou, Sabine. 2000. The grammatical ingredients of counterfactuality. *Linguistic Inquiry* 31(2). 231–270 • Kaufmann, Stefan. 2013. Causal premise semantics. *Cognitive Science* 37(6). 1136–1170. doi: 10.1111/cogs.12063 • Pearl, Judea. 2000. *Causality: Models, Reasoning and Inference*. Cambridge Univ Press