Informative Counterfactuals

Adam Bjorndahl[1] and Todd Snider[2]

1 Introduction

We use counterfactuals quite often. We use them to talk about things we know to be false, as well things we are simply unsure of. For example, (1) can be used if the speaker knows that Alice did not attend the party, or if the speaker is unsure whether she did.

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

(2) If the movie had been any good, I wouldn't have fallen asleep.

(3) Even if there hadn't been traffic, Francis still would have been late.

Despite their describing states of affairs which are counter to fact or uncertain, we can and do use counterfactuals to communicate informatively about the actual world. This is done in two ways. First, counterfactuals can encode information about the truth values of events in the actual world. For instance, (1) often implies that Alice in fact did not attend the party, and that Bill did. The *Even...still* construction in (3) communicates that in fact there *was* traffic, and in fact Francis *was* late. This sort of presupposed content can be accommodated in the Stalnaker 1974 sense, thus updating our knowledge of the actual world.

Second, and central to the contribution of this paper, counterfactuals encode information about a "connection" between the antecedent and consequent.

1.1 Intuitively different explanations

There are many different ways that the antecedent and consequent of a counterfactual can be related, all of which make the counterfactual true. For example, (1) encodes some sort of connection between Alice's being at the party and Bob's being at the party, but is silent about the specifics. Perhaps Bob is avoiding Alice, as in (4).

(4) A: If Alice had gone to the party, Bob would have stayed home.
B: Why?
A: He owes her some money.

Of course, there are many explanations for why Bob might be avoiding Alice, of which (4) represents only one. Perhaps Bob dislikes Alice, or he is just shy. Or is Alice Bob's committee chair, to whom he owes a draft? Consider now the discourse in (5).

(5) A: If Alice had gone to the party, Bob would have stayed home.
B: Why?
A: They hate these sorts of functions, so they take turns going.

In this case, Bob isn't avoiding Alice at all, but instead there is some shared reason for their attendance choices. (5) explicitly provides one such interpretation, but again there are many others. Maybe they share a budget and can't both afford to go. This list could go on and on; all told, there seem to be any number of things that the single counterfactual (1) could mean.

How can we capture this multiplicity of meanings? The variety of interpretations available for a single counterfactual, as demonstrated for (1) above, are independent of its truth value. In other words, classifying a counterfactual as true or false isn't enough to determine which interpretation is meant—what the nature of

[1]Cornell University, Department of Mathematics
[2]Cornell University, Department of Linguistics

the connection between antecedent and consequent is. In this paper we present a model for understanding counterfactuals that embraces this explanatory underspecification by explicitly modeling the connections between events.

2 Background

2.1 Informativity

We present an analysis of how we use counterfactuals informatively. But what does it mean to be *informative*?

Normally, when semanticists and philosophers of language call something informative, they mean that it meaningfully reduces the context set—the set of worlds that could plausibly be the actual world. In other words, it eliminates some (but not all) candidate worlds under consideration. A simple utterance of a declarative sentence which encodes the proposition φ, for example, partitions the universe of possible worlds into φ-worlds and $\neg\varphi$-worlds, and then rules out all the $\neg\varphi$-worlds, effectively removing any $\neg\varphi$-worlds still in the context set.

We will be relying on precisely this understanding of informativity, but from a slightly different perspective. Because we aim to explicitly model the connections between events, our analysis relies on a richer notion of possible worlds than is standard. Rather than defining a world as a set of (pairs of) events[3] and their truth values, our worlds will contain events, truth values, and *the dependencies among them*, following Starr (2012) and Briggs (2012). Relying on this richer definition, we have access to an additional means for discriminating among possible worlds. As is the case classically, we can form partitions based on the truth value of a specific event, but with structured possible worlds we can also partition based on the existence (or non-existence) of a particular dependency. That is, if an utterance asserts the existence of some dependency d, we can model that assertion's update as dividing the universe into d-worlds (worlds with that particular dependency) and $\neg(d)$ worlds (worlds without that particular dependency), and then remove worlds in the standard way.

In this way, we rely upon the standard notion of informativity, but with an additional dimension along which to partition. A dependency-conveying assertion is informative if and only if updating by that dependency meaningfully reduces the context set.

This "structured possible world" approach (which enriches the structure *within* worlds) stands in contrast to the classical approach to counterfactuals, often called the "similarity" approach. This classical approach relies on a *similarity relation* to encode a notion of "closeness" between worlds, effectively attributing additional structure *between* worlds (Lewis 1973, 1979). Our analysis, taking the internal structure of worlds as a given, relies on no such relation. All of the work done by the similarity relation is accomplished instead by reference to the dependencies among events within individual worlds.

2.2 Structural Equation Modeling

In order to represent these event-relating dependencies within worlds, we make use of the "structural equation modeling" framework. This framework goes back as far as Wright 1921, but we essentially work with the version formalized by Pearl (2000). A *Structural Equation Model* (SEM) can be pictured as a graph consisting of nodes, edges, and labels. Each node, represented as a lettered circle, stands for an event variable. Each edge, represented as an arrow between two nodes, reflects a directed dependency between events. Finally, the labels encode the specific nature of these dependencies. In §4.2 we provide formal definitions; for now, we focus on examples.

[3]These events, or states, are represented as variables. Whether they are all truly events, or whether these variables represent other types of facts as well, is immaterial to this project.

Consider Figure 1, which models two events, A and C, as well a dependency of C on A. The label stipulates that C has the same truth value as A: either they're both true or both false. This is in keeping with a general principle of SEMs, namely that the value of a child node is determined entirely by the value(s) of its parent node(s).

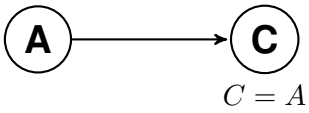

Figure 1: A simple structural equation model.

For the sake of simplicity, in this paper we restrict our attention to SEMs that are two-valued (i.e. variables are either true or false) and deterministic (i.e. the value of a child node is a *function* of the values of its parents, as opposed to a relation specified probabilistically). This framework, however, and indeed our analysis, can be extended to apply to multi-valued and/or probabilistic models.

3 Understanding counterfactuals

With this framework in hand, we can return to the question of what counterfactuals do, and how we use them. Roughly speaking, they assert some sort of connection between the antecedent and the consequent, though the exact nature of this connection need not be specified. Perhaps the simplest way to model such a connection is to posit a *direct sole dependency* of the consequent C on the antecedent A, as depicted in Figure 1. Indeed, this sort of dependency allows for exactly the sort of antecedent-consequent covariance we wish to capture.

In fact, this sort of direct dependency is, generally, conversationally implicated[4] by an utterance of a counterfactual. That is to say, unless otherwise overridden, this sort of connection is the default. We can show that this dependency is conversationally implicated rather than entailed by showing that (like standard conversational implicatures) it can be canceled or strengthened. Consider, for example, the counterfactual in (6).

(6) If I had pushed this button, the rocket would have launched.

Normally we take (6) to mean that pushing the button directly causes the rocket to launch. But this direct dependency can be denied without deriving a contradiction, as in (7), or expressly asserted without being redundant, as in (8).

(7) If I had pushed this button, the rocket would have launched, but pushing this button doesn't directly cause the rocket to launch.

(8) If I had pushed this button, the rocket would have launched, and (in fact) pushing this button directly causes the rocket to launch.

In many cases the story ends here, with the hearer updating their knowledge with the implicated direct dependency. In other cases, however, this direct dependency is problematic, leading the hearer to reject this interpretation of the meaning of the counterfactual.

3.1 Rejecting explanations

There are at least three types of reasons for rejecting a particular candidate interpretation for a counterfactual, including the simplest, conversationally implicated direct dependency described above. First, the

[4]This is not a claim about frequency. Rather, this sort of implicature is a generalized conversational implicature (GCI), in contrast to a particularized conversational implicature (PCI), the latter of which is cued specifically by some contextual factor(s) (Levinson 2000; Simons 2012).

dependency might contradict some prior knowledge. For example, if the speaker of (6) is indicating a button that you know to be an elevator call button, then you are likely to reject the simple direct dependency between pushing that button and launching a rocket. You might go searching for some more elaborate story, in which the elevator carries a military officer who might then go on to launch a rocket, etc.; the sort of model illustrated in Figure 1 is simply not a sufficient explanation.

Second, positing a given dependency (or set of dependencies) might violate some law about what makes a good explanation. For example, as far as our current understanding of the world goes, time travel is impossible. As such, if we know the consequent of a counterfactual to be temporally prior to the antecedent, a direct dependency of the consequent on the antecedent is problematic. We should not accept any model of the connections among events in our world which violates this principle, and so any interpretations of a counterfactual which do not conform to this standard must be rejected. The counterfactual in (9), for instance, cannot be explained by positing a direct dependence of Bob's brunch attendance on Alice's party attendance.

(9) If Alice hadn't come to the party this evening, Bob would have attended the brunch earlier today.

Third, an explanation might not meet the contextually-determined standard of specificity required in a given conversation. We seem to have different expectations for what level of detail is required in different contexts. For example, while taking a Physics test, one might expect that a high level of specificity is required (to demonstrate that one understands the material). By contrast, in a conversation among close friends, shared knowledge might well allow interlocutors to gloss over a great deal of detail; to explicitly spell out more than the minimum can even be insulting.[5] As such, if a given interpretation of a counterfactual fails to meet the appropriate standards for specificity in a conversation—either because it is too specific or not specific enough—it might be rejected as a suitable representation of its meaning.

Whatever the reason, rejecting the conversationally implicated interpretation of a counterfactual does not mean rejecting the counterfactual itself; indeed, as we saw above, a single counterfactual can give rise to many possible interpretations. If a hearer takes the speaker to be behaving cooperatively—that is, following normal conversational expectations (Grice 1975, 1978)—then the hearer must come up with an alternative interpretation which maintains the truth of the counterfactual. For an asserted counterfactual to be true, the associated SEM must encode the right kind of relationship between the antecedent and the consequent. What begins, then, is a *search* through the space of appropriate SEMs.

3.2 A typology of explanatory strategies

This process of searching for a suitable SEM seems to have a natural tripartite structure. We identify three primitive "moves" one can make in generating suitable models, and call them *explanatory strategies*. These three strategies, taken alone or in combination, account for the full range of interpretations available for a given counterfactual (beyond the default implicated direct dependency).

We first provide an intuitive sense of the three explanatory strategies; in §4.2, we show that in a precise sense these three classes really do partition the full search space. In order to clearly present these strategies it is convenient to have an example, for which we return to our reliable (1). The SEM representing the implicated direct dependency for (1) is given in Figure 2.

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

[5] Another piece of evidence for such a "contextual specificity" parameter is that we seem to be able to change it on the fly, requesting more detail than an interlocutor might have thought was necessary. Indeed, "Why?" seems to always be a felicitous follow-up in a conversation; one can always raise the bar for the level of detail required. Young children often enjoy taking advantage of this conversational move, once they discover it.

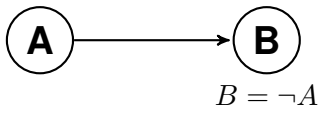

Figure 2: The conversationally implicated interpretation of (1).

Here, the variable A stands for "Alice goes to the party" while B stands for "Bob goes to the party" (and we assume for simplicity that Bob not going to the party is equivalent to Bob staying home).

For the counterfactual in (1), like many others, the SEM in Figure 2 is somewhat unsatisfying. Why should Alice attending the party force Bob not to? What does the speaker actually mean? What is crucially missing here is an *explanation*: why is (1) the case? The three explanatory strategies that follow serve as the basic building blocks by which one can begin to answer such a question, and in so doing, make sense of a counterfactual.

3.2.1 ADDITIONAL CAUSE

One strategy a hearer might employ to explain a counterfactual is to suppose that the consequent is dependent not solely on the antecedent but also on some additional cause. For example, a common interpretation of (1) is that Bob hates Alice, and because of this he avoids parties that she attends. In keeping with this intuition, we can represent Bob's hatred-driven-avoidance of Alice as an extra node H in the model, as in Figure 3.

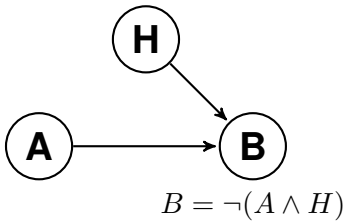

Figure 3: An ADDITIONAL CAUSE interpretation of (1).

In this model, the consequent (Bob's attendance) is still dependent on the antecedent (Alice's attendance), but not solely so. As we can see from the equation that describes how B inherits its value, the consequent depends on both the antecedent and the additional cause. This avoids the problematic simple dependency that was already rejected, but allows for the continued covariance of A and B (in the right H-conditions).

3.2.2 COMMON CAUSE

Alternatively, a hearer might suppose that the consequent is not dependent upon the antecedent at all. Instead, both the antecedent and consequent depend on some common cause that determines both of their values. Returning to Alice and Bob, one might imagine that they jointly flip a coin to decide who attends the party. We can model the coin flip as an independent event C determining Alice and Bob's party attendance choices, as in Figure 4.

3.2.3 INTERMEDIATE CAUSE

Finally, a hearer might suppose that the consequent depends on the antecedent only by means of some intermediate cause. This allows for the antecedent and consequent to covary in the right way, but without positing a direct dependency between the two events. For example, suppose that Alice brings her cat with

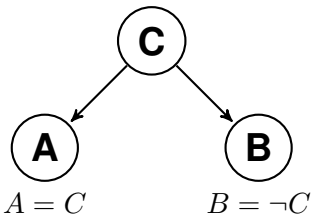

Figure 4: A COMMON CAUSE interpretation of (1).

her wherever she goes, and that Bob is deathly allergic to cats. In this case, Alice's attendance at the party still ends up leading to Bob's nonattendance, but only because of the mediating factor of her cat. We can model this particular scenario, an exemplar of the third explanatory strategy, as in Figure 5.

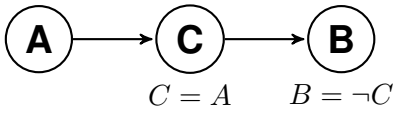

Figure 5: An INTERMEDIATE CAUSE interpretation of (1).

3.2.4 A fourth explanatory strategy?

Before we formalize this typology, it is instructive to consider what appears to be another way to ensure the right kind of covariance between antecedent and consequent: rather than having the consequent depend directly on the antecedent, one could model the antecedent as depending directly on the consequent. In other words, we could simply reverse the arrow, as in in Figure 6. This sort of model is what has been referred

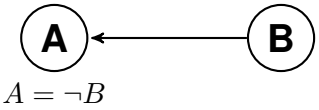

Figure 6: A classical backtracking explanation.

to in the classical philosophical literature as a *backtracker* (Lewis 1979): the consequent is the cause while the antecedent is the effect. And it might seem perfectly plausible; indeed, if we are happy to represent Bob as avoiding Alice, why not also admit interpretations in which Alice is avoiding Bob? Importantly, though, this model is unacceptable as an interpretation of (1): in any context where it is understood that Alice is avoiding Bob, (1) is simply infelicitous.

This kind of interpretation does, however, become available (and even preferred) under the right syntactic conditions, namely the double-auxiliary construction in (10).

(10) If Alice had gone to the party, Bob would have had to have stayed home.[6]

Our emphasis on understanding counterfactuals by searching for explanations in the form of SEMs proves its value here in two ways. First, it provides us with a natural class of semantic objects—SEMs like the one in Figure 6—that correspond to an explicit syntactic construction—the double-auxiliary as in (10). Second, it cleanly and clearly separates two notions that seem to be conflated in the literature. As noted above, in many classical philosophical accounts a "backtracker" describes a counterfactual conditional in

[6]Or, said differently, "For Alice to have gone to the party, it would have to have been the case that Bob stayed home."

which the consequent causally or temporally precedes the antecedent. By contrast, in more recent psychological literature, the term is applied to any counterfactual that relies on "upstream" reasoning, that is, that invokes an explanation involving something causally or temporally prior to the antecedent (Edwards and Rips 2012; Rips and Edwards 2013). Thus, Figure 4 depicts a canonical "psychological" backtracker, while Figure 6 depicts a "philosophical" backtracker. The two behave quite differently, as evidenced by the fact that the former, but not the latter, is an acceptable interpretation of (1). To our knowledge, this distinction has not been spelled out before now.

4 Formalizing explanation

4.1 Underspecification, not ambiguity

The multiplicity of interpretations for a given counterfactual is best analyzed not as *ambiguity*, but rather as *semantic underspecification*. Ambiguity is identifiable through the standard VP ellipsis test (see Asher, Hardt and Busquets 2001, and the many references therein): genuine ambiguity, as in words like *bank* (river bank vs. financial institution), is impossible under VP ellipsis. In the sentences in (11), for example, Janine and Kevin must both be at the same sort of bank.

(11) a. Janine went to the bank, and so did Kevin.
b. Janine went to the bank, and Kevin did, too.

We can apply the same test to counterfactuals, as in (12).

(12) a. If Alice had gone to the party, Bob would have stayed home, and so would Eve (have).
b. If Alice had gone to the party, Bob would have stayed home, and Eve would've, too.

Note that in both versions of (12), Bob and Eve can have very different reasons for not attending the party, even reasons that cut across distinct explanatory strategies. We can, for instance, have a COMMON CAUSE explanation for Bob (e.g., a coin toss) and an INTERMEDIATE CAUSE explanation for Eve.[7] As such, we formalize our analysis of counterfactuals using a single, underspecified semantics, rather than appealing to ambiguity.

4.2 Formalism

Here we provide a rigorous development of the SEM framework we employ, and use it to formalize our proposed semantics for counterfactuals. A **structural equation model** $\mathcal{M}$ is:

- a finite set of *variables* Var;
- a subset $\mathsf{End} \subseteq \mathsf{Var}$ of *endogenous* variables;
- for each $X \in \mathsf{End}$, a Boolean expression φ_X over Var such that $X \notin dom(\varphi_X)$.

Here, a *Boolean expression over* Var is any expression built from the variables in Var by closing under the Boolean connectives in the standard way; more precisely, it is any expression generated by the grammar

$$\varphi ::= X \mid \top \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \vee \psi \mid \varphi \rightarrow \psi,$$

[7] Notably, though, it seems quite difficult to get a reading of (12) where Bob is avoiding Alice but Alice is avoiding Eve—that is, mixing a classical backtracker, as in Figure 6, with a non-backtracking interpretation. That is to say, the classical backtracking interpretation is a different *reading*.

where $X \in \mathsf{Var}$. We write $dom(\varphi)$ to denote the set of all variables occurring in the expression φ; the **parents** of $X \in \mathsf{End}$ are precisely those variables in $dom(\varphi_X)$. The **ancestor** relation is the transitive closure of the parent relation, and we write $Y \prec X$ to denote that Y is an ancestor of X. Note that the presence of the "constant" formula $\top$ (*true*) means there are expressions with empty domain, and as such, there may be endogenous variables with no parents. Following many others (e.g., Hiddleston 2005, Briggs 2012, Kaufmann 2013) we restrict our attention to *recursive* SEMs, where the parent relation is acyclic.

A **truth assignment for** $\mathcal{M}$ is a function $v \colon \mathsf{Var} \to \{\mathsf{true}, \mathsf{false}\}$ such that for every $X \in \mathsf{End}$, $v(X) = v(\varphi_X)$, where by (a minor) abuse of notation we allow v to also denote the standard recursive extension of v to all Boolean expressions over Var. Given a Boolean expression φ over Var, we write $\mathcal{M} \models \varphi$ just in case for all truth assignments v for $\mathcal{M}$, we have $v(\varphi) = \mathsf{true}$.

We wish to define the semantic interpretation of the general counterfactual conditional in (13).

(13) If it had been the case that A, it would have been the case that B.

In keeping with the discussion in this paper, what we are looking for is an appropriate *space of explanations*. To this end, we begin by considering the following class of structural equation models:

$$[\![A \to B]\!] \coloneqq \{\mathcal{M} \,:\, \mathcal{M} \models A \to B\}.$$

The models in this class can be thought of as explanations for why B should be true whenever A is, since they specify (in terms of relations between variables as given by Boolean expressions) mechanisms by which the truth of A guarantees the truth of B. Indeed, the requirement that $\mathcal{M} \models A \to B$ says precisely that whenever $\mathcal{M}$ satisfies A, it also satisfies B.

However, it is not hard to see that there are models in this class that do not, intuitively, correspond to explanations for the counterfactual in (13). One problematic case consists in those models $\mathcal{M}$ for which A *cannot* be true under any truth assignment; in this case, $\mathcal{M} \models A \to B$ holds vacuously. Indeed, the impossibility of A seems quite at odds with (13), and certainly should not count as an explanation for it. As such, we restrict our attention to models where this does not occur, which we denote by:

$$[\![A > B]\!] \coloneqq \{\mathcal{M} \,:\, \mathcal{M} \models A \to B \text{ and } \mathcal{M} \not\models \neg A\}.$$

Another issue, perhaps more central to the theme of this paper, consists in those models $\mathcal{M}$ where B is an ancestor of A, a canonical example being the SEM pictured in Figure 6. We have already discussed (§3.2.4) the unacceptability of this kind of model as an explanation for a standard counterfactual conditional (i.e. one without additional syntactic licensing, as in (10)). We therefore add a further restriction excluding these cases, denoting the remaining models by:

$$[\![A \rhd B]\!] \coloneqq \{\mathcal{M} \,:\, \mathcal{M} \models A \to B,\ \mathcal{M} \not\models \neg A,\ \text{and } B \not\prec A\}.$$

The class $[\![A \rhd B]\!]$ is what we take to be the semantic interpretation of (13). It is not hard to see that the SEMs pictured in Figures 2, 3, 4, and 5 all fall into the class $[\![A \rhd \neg B]\!]$, and by construction, the SEM pictured in Figure 6 does not.[8]

Observe that the requirement by which we obtained $[\![A \rhd B]\!]$ from $[\![A > B]\!]$—that B not be an ancestor of A—is *not* the negation of the requirement that B is an ancestor of A. In fact, there are three possibilities in total: either $A \prec B$, or $B \prec A$, or neither $A \prec B$ nor $B \prec A$. If we denote this latter condition by $A \sim B$, then we have

$$[\![A \rhd B]\!] = \big([\![A > B]\!] \cap \{\mathcal{M} \,:\, A \prec B\}\big) \sqcup \big([\![A > B]\!] \cap \{\mathcal{M} \,:\, A \sim B\}\big).$$

[8] Note that $\neg B \not\prec A$ is technically undefined, since $\prec$ is a relation on variables, not Boolean expressions. However, to avoid more cumbersome notation, we will suffer this minor abuse, identifying a negated variable with the variable itself for the purposes of assessing ancestorship.

Thus, the space of explanations for (13) is naturally divided into two components: those where A is an ancestor of B, and those where neither is an ancestor of the other. A canonical example of an SEM that would lie in the second component is the one pictured in Figure 4, corresponding to the COMMON CAUSE explanatory strategy. In other words, essentially the same mechanism that excises classical backtrackers from the denotation of a counterfactual also serves to demarcate the class of COMMON CAUSE explanations. On the other hand, both ADDITIONAL CAUSE and INTERMEDIATE CAUSE explanations must lie in the set $[\![A > B]\!] \cap \{\mathcal{M} : A \prec B\}$; a precise formulation of the boundary between these two classes is the subject of ongoing research.

4.3 Extensions

The class $[\![A > B]\!]$ is useful as more than just a stepping-stone to defining $[\![A \rhd B]\!]$. Recall that what $[\![A > B]\!]$ admits beyond $[\![A \rhd B]\!]$ are models $\mathcal{M}$ where $B \prec A$, our classical backtrackers. As we saw in §3.2.4, these models are acceptable provided the counterfactual is embedded in the right syntactic environment, namely the double-auxiliary construction. There seem to be times, then, when the $B \not\prec A$ requirement must be relaxed. We therefore propose to take the class $[\![A > B]\!]$ as the semantic interpretation of double-auxiliary counterfactuals.

Note that this denotation does not *exclude* non-backtracking models, since $[\![A > B]\!]$ is a superset of $[\![A \rhd B]\!]$. But in fact, this turns out to be precisely what we want, as it accounts for two different but related facts about these double-auxiliary constructions: they seem to bias classical backtracking interpretations while still allowing (albeit with some difficulty) non-backtracking interpretations. We can account for the bias by appeal to general Gricean reasoning principles (Grice 1975): if the speaker is going out of her way to use the strictly weaker $[\![A > B]\!]$ rather than $[\![A \rhd B]\!]$, then she must be doing so because the interpretations allowed by $[\![A \rhd B]\!]$ are not sufficient. But while this biasing effect makes other interpretations harder to get, they are still available, which accords with the fact that $[\![A > B]\!]$ includes all the models that $[\![A \rhd B]\!]$ does.

We might also entertain a similar analysis of the *Even...still* construction mentioned in §1. These constructions seem to bias models where the truth of the consequent is fixed. One could argue that models $\mathcal{M}$ satisfying $\mathcal{M} \models B$ should be excluded from our basic semantic interpretation of (13), and the *Even...still* construction viewed as relaxing this restriction in just the same way that the double-auxiliary relaxes the $B \not\prec A$ restriction. In particular, this would imply that the *Even...still* construction biases interpretations where the truth of B is fixed, which accords well with intuition. However, it is not clear that all the standard interpretations are still available (even with effort) when this construction is used, and moreover, some speakers report being able to interpret B as fixed even without the *Even...still* construction. We leave the suitability of the $\mathcal{M} \not\models B$ restriction as an open question.

5 Further work

5.1 Biasing strategies

We have established a typology of explanatory strategies available for interpreting counterfactuals, but we have made no claims about when one explanation might be preferred over another. It is entirely possible, for instance, that some explanatory strategies tend to take precedence, or that explanations which posit fewer nodes are preferred to more complex SEMs. If such preferences exist, some might be universal (due perhaps to cognitive constraints), while others might be cued by certain (linguistic or non-linguistic) contextual factors. Two examples along these lines were discussed in §4.3; we leave a systematic study of such biases to future research.

5.2 Predictions for other languages

The typology we describe is motivated mathematically, not by any particular facts about English counterfactuals, and so it naturally extends to uses of counterfactuals crosslinguistically. We predict that any language with counterfactuals should be able to make use of all three explanatory strategies (including, of course, the simple direct dependency). However, there is no *a priori* reason to predict that all languages use a single underspecified form for all types of interpretations, nor that the simple direct dependency is conversationally implicated in all languages. Similarly, the classical backtracking explanation may not be differentially marked in all languages. Our analysis does, however, predict that a language which can use counterfactuals to describe COMMON CAUSE explanations should also be able to describe ADDITIONAL CAUSE and INTERMEDIATE CAUSE explanations. A crosslinguistic approach to the study of counterfactuals promises to be a rich area of investigation.

References

Asher, N., Hardt, D. and Busquets, J. (2001), 'Discourse parallelism, ellipsis, and ambiguity', *Journal of Semantics* **18**(1), 1–25.

Briggs, R. (2012), 'Interventionist counterfactuals', *Philosophical studies* **160**(1), 139–166.

Edwards, B. J. and Rips, L. J. (2012), Explanations of counterfactual inferences, *in* 'Proceedings of the Cognitive Science Society'.

Grice, H. P. (1975), 'Logic and conversation', *Syntax and Semantics* **3**, 64–75.

Grice, H. P. (1978), 'Further notes on logic and conversation', *Syntax and Semantics 9: Pragmatics* pp. 113–127.

Hiddleston, E. (2005), 'A causal theory of counterfactuals', *Noûs* **39**(4), 632–657.

Kaufmann, S. (2013), 'Causal premise semantics', *Cognitive Science* **37**(6), 1136–1170.
URL: *http://dx.doi.org/10.1111/cogs.12063*

Levinson, S. C. (2000), *Presumptive meanings: The theory of generalized conversational implicature*, MIT Press.

Lewis, D. K. (1973), 'Counterfactuals and comparative possibility', *Journal of Philosophical Logic* **2**(4), 418–446.

Lewis, D. K. (1979), 'Counterfactual dependence and time's arrow', *Noûs* **13**, 455–476.

Pearl, J. (2000), *Causality: Models, Reasoning and Inference*, Cambridge Univ Press.

Rips, L. J. and Edwards, B. J. (2013), 'Inference and explanation in counterfactual reasoning', *Cognitive Science* .

Simons, M. (2012), 'Conversational implicature', *Semantics: An international handbook of natural language and meaning* **33**.

Stalnaker, R. (1974), Pragmatic presuppositions, *in* P. K. U. Milton K. Munitz, ed., 'Semantics and Philosophy', New York: New York University Press, pp. 197–214.

Starr, W. B. (2012), Structured possible worlds. Ms. Cornell University.

Wright, S. (1921), 'Correlation and causation', *Journal of Agricultural Research* **20**(7), 557–585.