COUNTERFACTUAL CONDITIONALS AND DYNAMIC LAWS

KATRIN SCHULZ

ILLC
University of Amsterdam
k.schulz@uva.nl

In this paper we will argue that in order to account for the meaning of counterfactual conditionals we need to refer to and distinguish between static (f.i. analytical) laws and dynamic (f.i. causal) laws. We will propose an approach that combines premise semantics for counterfactuals in the style of Veltman 2005 with a representation of causal dependencies based on Pearl 2000.

1. Introduction

How to describe the truth conditions of counterfactual conditionals? Lets simplify matters and assume as logical form of such conditionals '$A > C$', where A is the antecedent, C the consequent and $>$ the conditional connector. An answer to the question that is on first view very attractive is the strict conditional approach. According to this approach $A > C$ is true with respect to model M and evaluation world w_0, if the set of possible worlds in M where the antecedent is true (denoted by $[\![A]\!]^M$) is a subset of the set of possible worlds where the consequent holds ($[\![C]\!]^M$). It is well-known that this approach is too strong. We do not want to demand that the consequent is true on all antecedent-worlds. There are other facts true of the evaluation world that can be used in inferring the consequent. The central challenge of the semantics of counterfactuals is to characterize these facts and the way they play a role in the derivation. Authors agree that general laws have to be part of these facts, but that also some singular facts of the evaluation world can be used. Furthermore, we also know that the relevant facts depend not only on the evaluation world but also on the antecedent. A well-known and successful way to describe this dependence is *premise semantics* for counterfactuals (Veltman 1976, Kratzer 1979). This proposal distinguishes two sets of relevant facts, called *premises*: the set $P_1(w_0)$ of general laws taken to hold in the evaluation world w_0, and a particular subset $P_2(w_0)$ of singular facts of w_0. Let U be the set of possible worlds where all elements of $P_1(w_0)$ hold. The truth conditions of a counterfactual can then be formalized by (i) defining an order that compares worlds with respect to how many of the premises in $P_2(w_0))$ they make true, and (ii), demanding that the consequent of the counterfactual has to be true all worlds in $[\![A]\!]^M \cap U$ minimal with respect to the order.

Definition 1 *(Truth conditions according to premise semantics)*[1,2]
$w_1 \leq^{M,w_0} w_2$ *iff* $\{\psi \in P_2(w_0) \mid M, w_1 \models \psi\} \subseteq \{\psi \in P_2(w_0) \mid M, w_2 \models \psi\}$,
$M, w_0 \models A > C$ *iff* $Min(\leq^{M,w_0}, [\![A]\!]^m \cap U) \subseteq [\![C]\!]^M$.

This approach leaves open how to define the functions P_1 and P_2. A recent proposal made by Veltman 2005 is to take as P_2 the *basis* of the evaluation world, which is defined as a minimal set of facts of the evaluation world from which everything else true about this world can be derived (using the general laws in $P_1(w_0)$).[3] This approach makes correct predictions for many traditionally hard examples for such conditionals. However, in some cases the predictions made are not in accordance with intuitions. In this paper we will concentrate on one particular type of such mispredictions.

2. A problem: causal counterfactuals

Suppose there is a circuit such that the light is on (L) exactly when both switches are in the same position (up or not up: $(S1 \wedge S2) \vee (\neg S1 \wedge \neg S2)$). At the moment switch one is down $(\neg S1)$, switch two is up $(S2)$ and the lamp is out $(\neg L)$. Now consider the following counterfactual conditional:

(1) If switch one had been up $(S1)$, the lamp would have been on (L).

The approach of Veltman wrongly predicts that the conditional (1) is false in the given context. The relevant law of this example that defines the set U is $(S1 \leftrightarrow S2) \leftrightarrow L$. Because the state of the lamp can tell you something about the position of the switches (as much as the position of the switches gives you information about the state of the lamp), there are bases containing the fact $\neg L$. In consequence, among the worlds making the antecedent $S1$ true and maintaining a maximal subset of a basis of w_0 are also worlds that make $S1$, $\neg S2$, and $\neg L$ true. In other words, among the antecedent worlds on which the consequent has to be true there are worlds where the lamp is out and, thus, switch two down. Because of these worlds the conditionals (1) is evaluated to be false.

3. Solution: an ontic notion of basis

Notice that if bases that contain $\neg L$ were not considered, the approach would have made the correct predictions. Only one basis would have been left ($B(w_0) = \{\neg S1, S2\}$) and the minimal worlds according to the order would have been the worlds making $S1$, $S2$ and L true. Therefore, we propose that the origin of the mispredictions lays in the way Veltman defines a basis. The notion of basis Veltman

[1] For sake of simplicity we assume here that minimal elements exist.
[2] $Min(\leq, S) = \{s \in S \mid \neg\exists s' \in S : s' \leq s \;\&\; s \not\leq s'\}$.
[3] A world can have more than one basis.

employs involves **epistemic** reasoning with laws. $\neg L$ is predicted to be part of a basis, because it gives you *information* about the position of the switches. However, the (dominant) reading of (1), according to which the sentence is true, is not epistemic, i.e. not about the conclusions you would derive given that you believed the antecedent to be true, but **ontic**, i.e. about how the world would have evolved if switch one had been up.[4] We propose that the notion of basis has to take **causal dependencies** into account. To arrive at the right definition of a basis we need an ontic concept of what can be derived from laws. In order to formulate this concept we have to distinguish two types of laws. First, there are **static laws** like analytical laws and logical laws. With static laws ontic reasoning works in the standard deductive manner. But besides static laws there are also **dynamic laws**, like **causal laws**. Characteristic for these laws is that they come with a direction, like the direction from cause to effect. Ontic reasoning with these laws has to follow their direction. A basis of the evaluation world is then again defined as a minimal set of basic facts from which everything else about this world can be derived – but now the ontic notion of derivation is applied.

In order to formalize this idea one needs an appropriate notion of a model that keeps track of the relevant static and dynamic laws. While static laws can be encoded as restrictions on acceptable possible worlds, for dynamic laws we need a more complex representation that also holds information about the direction of the law. Here we use a representation building on the **causal models** of Pearl 2000. For a finite set of proposition letters $\mathcal{P}$ we define a model M as a tuple $\langle C, U\rangle$, where C is a causal structure and U a set of possible worlds, the worlds where all laws hold. A causal structure is a tuple $\langle B, F\rangle$. B is a subset of the proposition letters called the *background variables*. F is a function mapping all elements of $\mathcal{P} - B$ to tuples $\langle Z_Y, f_Y\rangle$ where Z_Y is an n-tuple of elements of $\mathcal{P}$ and f_Y an n-ary partial truth function. F provides for every non-background variable the propositional variables on which they directly causally depend (Z_Y) and the character of the dependency (f_Y). We demand additionally that the relation XRY that can be defined by the condition XRY iff $X \in Z_Y$ is acyclic and that all its minimal elements are members of B. This realizes the idea that causal dependencies are not cyclic and that the background variables have in the relevant context no causal history. One can think of them as chance events. The figure below sketches the model of example (1).

Based on this definition of a model we can now define our ontic notion of basis. We first introduce the *law closure* of a partial interpretation function i. This is the extension of i with the interpretation of proposition letters that can be derived by laws from i. Crucial here is that only derivations from causes to effects are allowed. The basis of a world w will be defined as the union of all smallest subsets of w (thus, partial interpretation functions) for which w is the law closure.[5]

[4] I argue elsewhere (Schulz 2007) that also a epistemic reading of conditional sentences has to be distinguished.

[5] For more details on the definitions see Schulz 2007.

$$\begin{aligned}\mathcal{P} &= \{S_1, S_2, L\},\\ M &= \langle\langle B, E, F\rangle, U\rangle,\\ U &= \text{all interpretation}\\ &\quad\ \text{functions for } \mathcal{P},\\ B &= \{S_1, S_2\}.\end{aligned}$$

$F(L):$	S_1	S_2	L
	0	0	1
	0	1	0
	1	0	0
	1	1	1

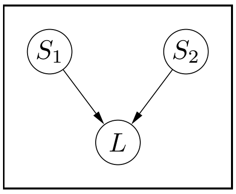

Figure 1: A model for example (1)

Definition 2 *(Law closure)*
Let i be a partial interpretation of $\mathcal{P}$. The law closure $\bar{i}$ of i is the minimal i' with $i \subseteq i'$ fulfilling the following conditions.

(i) $i' = \bigcap\{w \in U \mid i' \subseteq w\}$ *(closure w.r.t. static laws),*

(ii) for all $P \in \mathcal{P} - B$ with $Z_P = \langle P_1, ..., P_n\rangle$ such that $i(P)$ is undefined it holds that if for all $k \in \{1, ..., n\}$: $i'(P_k)$ is defined and $f_P(i'(P_1), ..., i'(P_n))$ is defined, then $i'(P)$ is defined and $f_P(i'(P_1), ..., i'(P_n)) = i'(P)$ (closure w.r.t. dynamic laws).

Definition 3 *(Basis)*
The basis b_w of a possible world $w \in U$ is the union of all partial interpretation functions b that fulfill the following two conditions: (i) $b \subseteq w \subseteq \overline{b}$ and (ii) $\neg\exists b' : b' \subseteq w \subseteq \overline{b'}\ \&\ b' \subset b$.

4. Another problem: causal backtracking

If we use this notion of basis set as the premise function P_2 in premise semantics, we predict the counterfactual (1) to be true. However, there is a different but related problem of Veltman 2005 that we do not solve this way. This approach, as well as the one defined above, predicts causal backtracking to be valid. That means that in the described scenario the following counterfactual comes out as true.

(2) If the lamp had been on, the switches would have been in the same position.

There is general agreement in the literature that while backtracking using static laws is fine, causal backtracking is highly marked if not even impossible (cf. Lewis 1979 and many others). We propose that according to the dominant ontic reading of counterfactuals that we aim to model here, causal backtracking is not possible, that means, counterfactual (2) should come out as false.[6] We will go even a step further

[6] We do not want to claim this way that causal backtracking is in general not possible, but rather defend the position that it is excluded by the dominant ontic reading of counterfactuals. There is an marginal epistemic reading that does allow for backtracking, but this reading is not subject of the present paper (for more details on the epistemic reading see Schulz 2007).

and propose that we even want strong exclusion of backtracking, that means that we want be able to conclude that the causal history of the antecedent remains unchanged (cf. Lewis 1979, Pearl 2000 and others). In other words, the counterfactual (3) should come out as true.

(3) If the lamp had been on, the switches would not have changed their position.

5. Solution: a new kind of minimization

Following an idea of Lewis 1979 we model strong exclusion of backtracking by allowing for violations of causal laws. We take U to be the set of worlds that follow the static laws, but not necessarily also the dynamic laws encoded in the causal structure. That means, for instance, that for example (1) among others also the world w_1 where both switches are up, but the lamp still is off is an element of U. However, if we use this new definition of U in premise semantics the counterfactual (1) no longer comes out as true. The problem is that we do not only allow for law violations that cut the antecedent from its causal history, but that all kinds of law violations are possible - for instance also law violations preventing the causal effects of the antecedent to hold (as in w_1).

I propose that to account for this problem we have to change the order with respect to which minimal worlds are selected. Instead of maximizing the overlap with the basis of the evaluation world, we have to minimize the law violations happening in a world. Because law violations result in extensions of the basis, this can be formalized by minimizing additional basis elements (clause (1) of def. 4). Using this order instead of the order of premise semantics already allows us to achieve weak exclusion of backtracking while at the same time predict examples like (1) to be true. To get additionally strong exclusion of backtracking I propose that also non-basis facts count for the truth conditions of counterfactuals, but to a lesser degree. We define a second order that maximizes the overlap with derivable facts of the evaluation world (clause (2) of def. 4). After minimizing with respect to the first order the worlds where the antecedent and the static laws hold, in a second step we minimize with respect to this second order (clause (3) of def. 4). These are the worlds on which we claim the consequent of the counterfactual to hold as well.

Definition 4 *(The ontic reading of counterfactuals)*

(1) $w_1 \leq_B^{M,w_0} w_2 \textit{ iff } B(w_1) - B(w_0) \subseteq B(w_2) - B(w_0),$

(2) $w_1 \leq_D^{M,w_0} w_2 \textit{ iff } D(w_1) \cap D(w_0) \subset D(w_2) \cap D(w_0),\ where\ D(w) = w - B(w),$

(3) $M, w_0 \models A > C \textit{ iff } Min(\leq_D^{M,w_0}, Min(\leq_B^{M,w_0}, [\![A]\!]^M \cap U)) \subseteq [\![C]\!]^M.$

Based on this approach to the truth conditions of counterfactuals we predict the examples (1) and (3) to be true, while (2) is predicted to be false – as intended.

6. Conclusion

In this paper we have proposed an approach to the truth conditions of counterfactual conditionals that is based on premise semantics, but deviates from standard premise semantics (i) in assuming an ontic premise function, and (ii) in the way we propose that the premises influence the truth conditions. We propose that the consequent of a counterfactual has to be true on those antecedent worlds that (i) make all static laws true, (ii) add the least number of law violations, and (iii) with respect to facts derived from the basis are as similar as possible to w_0. This approach improves on Veltman 2005 in being able to account for causal counterfactuals, without loosing Veltman's appealing predictions. Furthermore, the approach predicts strong exclusion of causal backtracking. That means that it does not only prevent reasoning from effect to cause, but it supports the even stronger prediction that making the antecedent true leaves its causal history unchanged.

There are some issues left open for future research. First, similar to Veltman 2005 the present approach improves on standard premise semantics in providing a way to calculate the premises for concrete examples. However this calculation takes as input the set of general laws considered valid. As does Veltman 2005, we still need a theory for what should count as general laws. We do need even a bit more: we have to be able to make the distinction between static and dynamic laws. This is a problem that has to be investigated in future work. Second, here we studied the semantic meaning of counterfactuals on a very abstract level: on the level of propositional logic. We did not consider the linguistic fine structure of antecedent and consequent, as for instance the meaning of modals occurring there. This is a limitation that should be overcome in future work. In Schulz 2007 the present proposal has been extended with a more compositional approach to the meaning of conditionals is made. This work has to be continued in the future.

Bibliography

Kratzer, A.: 1979, Conditional necessity and possibility, in R. Bäuerle, U. Egli, and A. von Stechow (eds.), *Semantics from different points of view*, pp 387–394, Springer, Berlin/Heidelberg/New York

Lewis, D.: 1979, Counterfactual dependence and time's arrow, *NOÛS* 13, 455–476

Pearl, J.: 2000, *Causality. Models, reasoning, and inference*, Cambridge University Press, Cambridge

Schulz, K.: 2007, *Minimal models in semantics and pragmatics: Free choice, exhaustivity, and conditionals*, Ph.D. thesis, University of Amsterdam, Amsterdam

Veltman, F.: 1976, Prejudices, presuppositions and the theory of counterfactuals, in J. Groenendijk et al. (eds.), *Amsterdam Papers in Formal Grammar*, Vol. 1, Centrale Interfaculteit, Universiteit van Amsterdam

Veltman, F.: 2005, Making counterfactual assumption, *Journal of Semantics* 22, 159–180