Interventions in Premise Semantics

ZHANG Yu

SEP 24, 2019

Table of Contents

1 Introduction

2 Premise semantics

3 Causal models

4 Filtering semantics for counterfactuals—— first version

5 Filtering semantics for counterfactuals—— refined version

6 Comparation

7 Conclusion

Much work has gone into developing a logic and a semantics for counterfactual conditionals, i.e. conditionals of the form:

(1) If Mary pulled the trigger, her gun would fire.

? ? ?

What will happen when we merge two different lines of theorizing about counterfactuals, with particular attention to the goal of giving a compositional semantics.

Transplanting causal-models-inspired ideas in a possible worlds framework yields a substantially new semantics, which makes systematically different predictions and generates a new logic. The difference is ultimately grounded in different algorithms for handling inconsistent information.

A rough sketch

Premise semantics

We hold fixed a set S of true propositions, which work as covert premises, and we check whether those propositions, together with the antecedent, entail the consequent. Schematically:

$$\ulcorner p \mathrel{\Box\!\!\rightarrow} q \urcorner \text{ is true iff p, together with propositions in set S, entail q}$$

The new semantics adds an extra step.

Rather than using a fixed stock of propositions, we use the antecedent to selectively eliminate some of those propositions from the set. I say that, when this happens, the set S is filtered for the antecedent. Accordingly, I call the new semantics filtering semantics. Here are the new schematic truth conditions:

$$\ulcorner p \mathrel{\Box\!\!\rightarrow} q \urcorner \text{ is true iff p, together with propositions in set S filtered for p, entail q}$$

Ordering semantics

One alternative version of possible worlds semantics exploits, rather than covert premises, a relation of comparative closeness between worlds. Within this framework, filtering amounts to an antecedent-driven shift in what worlds count as closer by or further away—something that is not contemplated by any standard counterfactual semantics.

Tow tasks:

- Constructing a causal models-based semantics;
- Explaining how it differs from classical counterfactual semantics.

Outline

2 Premise semantics

- Ordering semantics
- Modal premise semantics
- Modal premise semantics: counterfactuals
- Causal dependencies and premise semantics

Virtually all contemporary accounts of counterfactuals in the possible worlds tradition start from a simple idea, which is pithily put by Stalnaker:

"Consider a possible world in which A is true, and which otherwise differs minimally from the actual world. 'If A, then B' is true (false) just in case B is true (false) in that possible world." (Stalnaker 1968)

$\preceq_w$

A relation of comparative closeness to compare worlds with respect to their closeness to a benchmark world w.

$w' \preceq_w w''$

w′ is closer to w than w″.

Here is a version of truth conditions for counterfactuals that is often used, and that strikes a middle ground between Stalnaker and Lewis's own accounts:

$\ulcorner$ If ϕ, would ψ $\urcorner$ is true at w just in case all ϕ- world that are closest accouding to $\preceq_w$ are ψ- world.
(Limit assumption: for any antecedent, there is a $\preceq_w$-maximal set of antecedent worlds.)

For Kratzer, modalized claims in natural language state the existence of a relation between the proposition expressed by the embedded clause (the prejacent) and a certain body of information.

(2) David must be the murderer.

(2) is entailed by a body of information, which Kratzer thinks of as a set of covert premises.

Two contextual parameters which jointly determine which propositions are used as premises: the modal base and the ordering source.

Modal base (Assumed consistent)

A function from worlds to sets of propositions.
Including propositions that are, in some relevant sense, settled in the context.

Ordering source

A function from worlds to sets of propositions.
Used to generate a ranking of worlds along some appropriate dimension.

Kratzer's fix: rather than looking at the logical relations between the prejacent and an inconsistent premise set, we consider all the biggest consistent fragments of the premise set.

Three parameters:

- w: a possible world;
- f: a modal base;
- g: an ordering source.

$[\![\phi]\!]^{\mathrm{w,f,g}}$= semantic value of ϕ relative to parameters w, f, g
$||\phi||_{\mathrm{f,g}} = \{\mathrm{w} : [\![\phi]\!]^{\mathrm{w,f,g}} = 1\}$

A basic version of Kratzer's semantics for modals

A set of propositions S is a maximal consistent superset of S′(modal base) relative to S″(ordering source) iff

(a) S is a superset of S′,

$$[S \supseteq S']$$

(b) S is consistent,

$$[\cap S \neq \emptyset]$$

(c) S is formed from S′ by adding zero or more propositions from S″,

$$[(S - S') \subseteq S'']$$

(d) if any more propositions from S″ were added to S, S would be inconsistent.

$$[\sim \exists p \in S'' : p \notin S \wedge \cap(S \cup \{p\}) \neq \emptyset]$$

The schematic truth conditions of a modal necessity claim are:

(3) $[\![$ must $\phi]\!]^{w,f,g} = 1$ iff for every maximal consistent superset S of f(w) with respect to g(w), $S \vDash \|\phi\|_{f,g}$

For Kratzer, all conditional statements are modal statements of sort.

(4) $[\![$ If ϕ, would $\psi]\!]^{\mathrm{w,f,g}} = 1$ iff, for all maximal consistent supersets S of $\mathrm{f(w)} \bigcup \{||\phi||_{\mathrm{f,g}}\}$ with respect to $\mathrm{g(w)}, \mathrm{S} \vDash ||\psi||_{\mathrm{f,g}}$

Kratzer's proposal: the modal base starts out empty, while the ordering source maps each world to a set of propositions that are true at that world.

In present terms, the limit assumption is that, no matter how we extend the modal base by adding propositions from the ordering source, we always hit on a maximal consistent superset, i.e. one that cannot be further extended without falling into inconsistency.

Consider the following scenario— Coin toss

Alice is about to toss a coin and offers Bob a bet on heads; Bob declines. Alice tosses the coin, which does indeed land heads

Consider the following counterfactual:

(5) If Bob had taken the bet, he would have won.

Natural language counterfactuals track relationships of causal dependence and independence, and this information should be incorporated into premise sets and orderings. The notions of dependence in play may be understood in a broader way. Many general counterfactuals track dependencies of a noncausal nature.Example:

(6) If I had arrived at 2:05, I would have been five minutes late.

Outline

3 Causal models
- The basic framework
- Evaluating counterfactuals

The main ambition of the causal models framework is modeling how events in a causal network are dependent or independent of one another, and how a change in the outcome of one event affects the others.

A causal model consists in an ordered pair of two elements: $< V, E >$.

V : a set of random variables.

A random variable can be thought of as a set of mutually exclusive and jointly exhaustive outcomes for a process.

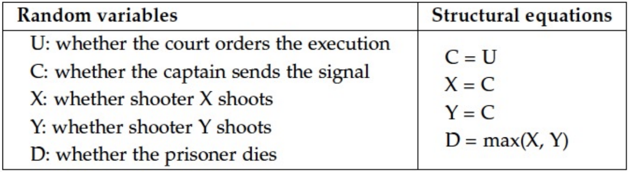
E : a set of structural equations.

Structural equations are mathematical equations that state the relations between different values of random variables.

A classical example from Pearl 2000:

The firing squad. A firing squad is positioned to execute a prisoner. The squad is waiting for a court order. The court issuing the execution order will result in the captain sending a signal to the two members of the squad, X and Y, who will fire and kill the prisoner. The court not issuing the order will result in the captain not sending the signal, the two riflemen not shooting, and the prisoner remaining alive.

A causal model for this scenario:

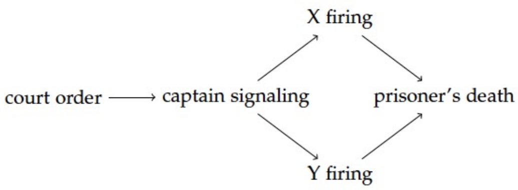


Random variables	**Structural equations**
U: whether the court orders the execution C: whether the captain sends the signal X: whether shooter X shoots Y: whether shooter Y shoots D: whether the prisoner dies	$C = U$ $X = C$ $Y = C$ $D = \max(X, Y)$

Exogenous variables are those whose values are determined by factors external to the model. Such as U in this model.

Endogenous variables are those whose values are determined by factors within the model.

Causal models are usually represented visually by means of directed graphs:



Nodes: random variables;
Arrows: relationships of causal dependence.

Recursive models

Recursive models are the ones in which we can define a relation $\prec$ between random variables such that:
(a) X $\prec$ Y iff the value of X is not dependent on the value of Y;
(b) $\prec$ is a total order.

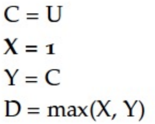
(A total order is a relation R that is antisymmetric, transitive, and connected, i.e. such that, for any x and y, either R(x, y) or R(y, x).)

The key notion is that of an intervention. Two steps:

- Performing an intervention on the model to make the antecedent true.
- Helping ourselves to the modified set of equations and holding fixed the values of the exogenous variables, we recalculate the values of the endogenous variables and check whether the consequent holds.

Example:

(7) If X had fired, the prisoner would have died.



$C = U$

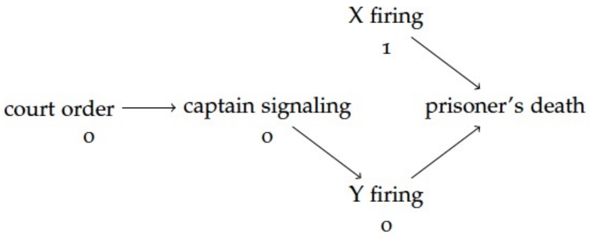
$\mathbf{X = 1}$

$Y = C$

$D = \max(X, Y)$

Notice: Given the way that the procedure is set up, all the values of variables that are upstream with respect to the intervention are guaranteed to remain the same; values of other variables may change.

New graph:



This evaluation procedure restricts to counterfactuals where antecedents are simple sentences—essentially, atomic sentences of the language or conjunctions. One advantage of implementing this algorithm in filtering semantics is that we automatically get a general formal system for handling counterfactuals of any complexity.

Outline

4 Filtering semantics for counterfactuals—— first version

- The goal
- Overview of the semantics
- Directional premises
- Basic filtering
- Summary

There is a conceptual difference between the causal models framework and classical premise semantics: the two frameworks rely on different algorithms for resolving inconsistency. Hence their divergence has to do with the very core of a semantics for counterfactuals.

Our goal:

Exposing this difference;
Building a causal-models-based semantics that captures it.

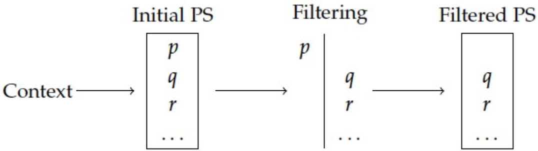
Classical premise semantics handles inconsistent premise sets by considering all maximal consistent subsets of the inconsistent set.

Crucially, the causal-models-based evaluation of counterfactuals operates in a different way. Together with the inconsistency-generating antecedent, we receive instructions to remove some specific piece of information from our previous stock. Hence, together with the addition of information to the existing stock, we have a loss of previously existing information. This solves immediately the problem of inconsistency; there is no need to consider subsets of the premise set.

The main innovation is the filtering operation. On classical premise semantics, recall, the antecedent of a counterfactual is simply added to the (otherwise empty) modal base:

(4)$[\![\text{If } \phi, \text{ would } \psi]\!]^{w,f,g} = 1$ iff, for all maximal consistent supersets S of $f(w) \cup \left\{\|\Phi\|_{f,g}\right\}$ with respect to $g(w), S \vDash \|\psi\|_{f,g}$.

The new semantics adds an extra step: the ordering source is filtered for the antecedent. Hence, while some information is added to the modal base, some other information is removed from the ordering source. In diagram form:



′X | p′ for ′X is filtered for p′. Below is a first-pass new meaning for counterfactuals.

(8)$[\![\text{If } \phi, \text{ would } \psi]\!]^{w,g} = 1$ iff $g(w) \mid ||\phi||_g$ entails $||\psi||_g$

The implementation of filtering requires modifying the format of the ordering source. Recall from §3: interventions crucially exploit the directionality of the equations.To implement a similar algorithm in premise semantics, we need to keep track of direction as well—we need to be able to say what determines what. Hence the premises we use need to be more informative than in standard systems.

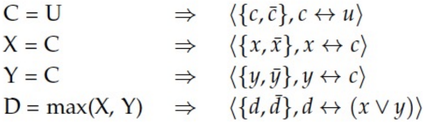
To this end, I treat the members of the ordering source not as propositions, but as pairs of a question denotation and a proposition. For example, the equation $'\mathrm{X} = \mathrm{C}'$ is turned into the pair:

$$\langle \{\{\mathrm{w} : \mathrm{X\ fires\ in\ w}\}, \{\mathrm{w} : \mathrm{X\ doesn't\ fire\ in\ w}\}\}, \{\mathrm{w} : \mathrm{X\ fires\ iff\ C\ gives\ the\ order\ in\ w}\}\rangle$$

For simplicity, I take all questions in play to be binary yes-no questions, though the semantics immediately generalizes to all questions with finite answer sets.

The new ordering source, mirroring causal models, will incorporate information of two kinds:
(a) information about causal dependencies and independencies between relevant events (corresponding to structural equations);
(b) information about some background facts (corresponding to the values of exogenous variables).

For illustration, this is how the equations in the execution model get transposed into premises:



$$\begin{array}{lll} \mathrm{C} = \mathrm{U} & \Rightarrow & \langle \{c, \bar{c}\}, c \leftrightarrow u \rangle \\ \mathrm{X} = \mathrm{C} & \Rightarrow & \langle \{x, \bar{x}\}, x \leftrightarrow c \rangle \\ \mathrm{Y} = \mathrm{C} & \Rightarrow & \langle \{y, \bar{y}\}, y \leftrightarrow c \rangle \\ \mathrm{D} = \max(\mathrm{X}, \mathrm{Y}) & \Rightarrow & \langle \{d, \bar{d}\}, d \leftrightarrow (x \vee y) \rangle \end{array}$$

Assuming that the court does not issue the order, we get:

$$\mathrm{U} = 0 \quad \Rightarrow \quad \langle \{\mathrm{u}, \overline{\mathrm{u}}\}, \overline{\mathrm{u}} \rangle$$

On this basic version of the semantics, a premise is filtered just in case the antecedent settles the answer to its question.

Answer

A proposition p is an answer to a premise S iff $S = \langle Q, r \rangle$ and $p \in Q$.

Settle

A proposition p settles a premise S iff, for some answer q to $S, p \vDash q$.

Filtering

A filtering of a premise set Π relative to proposition p (formally: $\Pi \mid p$) is a premise set Π such that,
(i) $\langle \{p, \overline{p}\}, p \rangle \in \Pi'$
(ii) for all premises $S \in \Pi$, if p doesn't settle $S, S \in \Pi'$;
(iii) no other premises are in Π'

Proposition set

The proposition set of a premise set Π is the set Prop_{Π} such that:

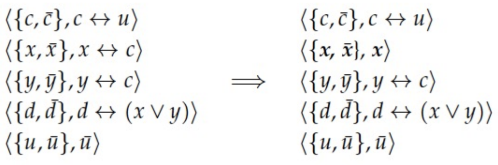
$$\mathrm{Prop}_{\Pi} = \{p \mid \exists S \in \Pi : \text{ for some } Q, S = \langle Q, p\rangle\}$$

Here is a semantics for counterfactuals (minimally different from the first-pass statement in (8)):

(9) $[\![$ if ϕ, would $\psi]\!]^{w,g} = 1$ iff the proposition set of $g(w) \mid \|\phi\|_g$ entails $\|\psi\|_g$.

Example:

(7) If X had fired, the prisoner would have died.



$$\begin{array}{lcl}
\langle\{c,\bar{c}\}, c \leftrightarrow u\rangle & & \langle\{c,\bar{c}\}, c \leftrightarrow u\rangle \\
\langle\{x,\bar{x}\}, x \leftrightarrow c\rangle & & \langle\{x,\bar{x}\}, x\rangle \\
\langle\{y,\bar{y}\}, y \leftrightarrow c\rangle & \Longrightarrow & \langle\{y,\bar{y}\}, y \leftrightarrow c\rangle \\
\langle\{d,\bar{d}\}, d \leftrightarrow (x \vee y)\rangle & & \langle\{d,\bar{d}\}, d \leftrightarrow (x \vee y)\rangle \\
\langle\{u,\bar{u}\}, \bar{u}\rangle & & \langle\{u,\bar{u}\}, \bar{u}\rangle
\end{array}$$

On classical semantics, we evaluate a counterfactual by adding the antecedent to our stock of information, and we check all ways of making that stock consistent; On filtering semantics, we also remove some information from our existing stock.

Classical semantics employs a 'global' strategy for solving inconsistency ("check all ways to make the premise set consistent"); Filtering semantics a 'local' strategy ("check some ways to make the premise set consistent, specifically the ones that ignore information about the causal links upstream from the antecedent")

Outline

5 Filtering semantics for counterfactuals—— refined version
- Minimally different models
- Permissible filterings
- Formal semantics

There may be multiple ways to filter a set of premises for an antecedent. To see this, consider once more the prisoner scenario and take the counterfactual:

(10) If rifleman X or rifleman Y had shot, the prisoner would have died.

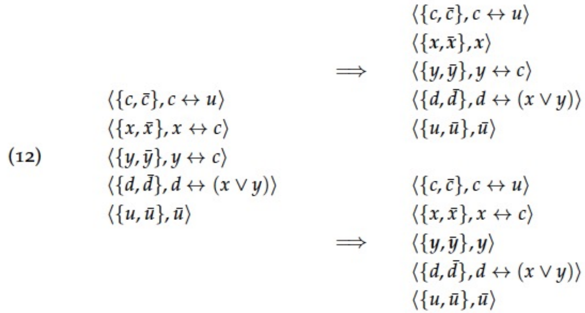
The antecedent of (10) doesn't trigger any filtering. Recall the premise set I've been using:

(11)
$$\begin{array}{l} \langle\{c,\bar{c}\}, c \leftrightarrow u\rangle \\ \langle\{x,\bar{x}\}, x \leftrightarrow c\rangle \\ \langle\{y,\bar{y}\}, y \leftrightarrow c\rangle \\ \langle\{d,\bar{d}\}, d \leftrightarrow (x \vee y)\rangle \\ \langle\{u,\bar{u}\}, \bar{u}\rangle \end{array}$$

The problem is obvious: there are (at least) two ways to filter the premise set. The antecedent doesn't settle how to do it. Hence the naïve filtering mechanism I considered above would predict that the premise set doesn't change. This is not the result we want.

The key idea behind filtering is that we modify the background information that we use to evaluate a conditional. Our first-pass attempt simply assumes that each conditional antecedent settles how this information should be modified. This is too simplistic. Conditional antecedents may be too unspecific to determine exactly how the relevant information changes. The natural suggestion is that we consider multiple ways of modifying the background information in the light of the antecedent.

For illustration, let me anticipate the result of the proposal for (10). The semantics considers the following two filterings—one for each of the disjuncts:



(12)

$$
\begin{array}{lcl}
 & & \langle\{c,\bar{c}\}, c \leftrightarrow u\rangle \\
 & & \langle\{x,\bar{x}\}, x\rangle \\
 & \Longrightarrow & \langle\{y,\bar{y}\}, y \leftrightarrow c\rangle \\
\langle\{c,\bar{c}\}, c \leftrightarrow u\rangle & & \langle\{d,\bar{d}\}, d \leftrightarrow (x \vee y)\rangle \\
\langle\{x,\bar{x}\}, x \leftrightarrow c\rangle & & \langle\{u,\bar{u}\}, \bar{u}\rangle \\
\langle\{y,\bar{y}\}, y \leftrightarrow c\rangle & & \\
\langle\{d,\bar{d}\}, d \leftrightarrow (x \vee y)\rangle & & \langle\{c,\bar{c}\}, c \leftrightarrow u\rangle \\
\langle\{u,\bar{u}\}, \bar{u}\rangle & & \langle\{x,\bar{x}\}, x \leftrightarrow c\rangle \\
 & \Longrightarrow & \langle\{y,\bar{y}\}, y\rangle \\
 & & \langle\{d,\bar{d}\}, d \leftrightarrow (x \vee y)\rangle \\
 & & \langle\{u,\bar{u}\}, \bar{u}\rangle
\end{array}
$$

We call the premise sets resulting from this procedure permissible filterings of the original premise sets. Hence the new schematic truth conditions of a counterfactual are:

(13) $[\![$ if ϕ, would $\psi]\!]^{w,g} = 1$ iff for every Π s.t. Π is a permissible filtering of g(w) for ϕ, the proposition set of Π entails that ψ is true.

Basic idea:

We use something like the converse of the filtering algorithm we had in §4. There we checked whether the antecedent of a conditional settled the answer to any questions in the premise set. Now we check which answers or combinations of answers in the premise set entail the antecedent. In particular, we check which minimal combinations of answers (for some suitable way of understanding minimality) will make the antecedent true. This will capture the idea that filterings are minimal ways of modifying the premise set that make the antecedent true.

Question set

The question set of a premise set P is simply the set of all questions appearing in the premise set:

(14) $\Sigma_\Pi = \{Q : \exists P \in \Pi : \text{ for some } p, P = \langle Q, p \rangle\}$

Answer set

The answer set is just the set of all the answers appearing in the question set.

(15) $A_\Pi = \cup \Sigma_\Pi$

Notice:
The question set is a set of sets of propositions;
the answer set is a set of propositions.

Given a counterfactual antecedent p, we single out the minimal subsets of the answer set A_Π that entail p.

Filter set

The filter set of a premise set Π relative to proposition p is the set $\Phi_{\Pi,p}$ of all minimal subsets S′ of the answer set A_Π such that $S' \vDash p$

$$(16) \quad \Phi_{\Pi,p} = \{S' \subseteq A_\Pi : S' \vDash p \text{ and } \neg\exists S'' : S'' \subset S' \text{ and } S'' \vDash p\}$$

Informally, a permissible filtering of a premise set Π relative to a proposition p is the result of
(a) picking a set member of the filter set;
(b) filtering out all and only the premises whose questions are answered by that set of propositions, while letting in the premise corresponding to p.

Answer

A proposition p is an answer to a premise S iff $S = \langle Q, r\rangle$ and $p \in Q$.

Permissible filtering

A permissible filtering of a premise set Π relative to proposition p is a premise set Π_p such that:

(i) $\langle\{p, \overline{p}\}, p\rangle \in \Pi_p$;

(ii) for some set of propositions S in the filter set $\Phi_{\Pi,p}$ and for all $P \in \Pi$:

- if P is not answered by any proposition in $S, P \in \Pi_p$;
- if P is answered by some q in $S, \langle\{q, \overline{q}\}, q\rangle \in \Pi_p$;

(iii) nothing else is in Π_p .

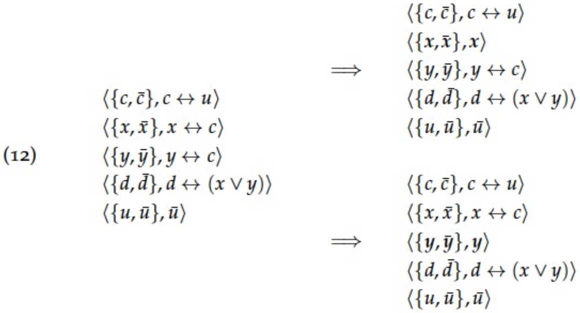
Example:

(10) If rifleman X or rifleman Y had shot, the prisoner would have died.

Question set: (17) $\Sigma_{g(w)} = \{\{u, \overline{u}\}, \{c, \overline{c}\}, \{x, \overline{x}\}, \{y, \overline{y}\}, \{d, \overline{d}\}\}$.

Answer set: (18) $A_{g(w)} = \{u, \overline{u}, c, \overline{c}, x, \overline{x}, y, \overline{y}, d, \overline{d}\}$

Filter set: (19) $\Phi_{g(w), x \vee y} = \{\{x\}, \{y\}\}$



(12)

$$\begin{array}{lcl} & & \langle\{c,\bar{c}\}, c \leftrightarrow u\rangle \\ & & \langle\{x,\bar{x}\}, x\rangle \\ & \Longrightarrow & \langle\{y,\bar{y}\}, y \leftrightarrow c\rangle \\ \langle\{c,\bar{c}\}, c \leftrightarrow u\rangle & & \langle\{d,\bar{d}\}, d \leftrightarrow (x \vee y)\rangle \\ \langle\{x,\bar{x}\}, x \leftrightarrow c\rangle & & \langle\{u,\bar{u}\}, \bar{u}\rangle \\ \langle\{y,\bar{y}\}, y \leftrightarrow c\rangle & & \\ \langle\{d,\bar{d}\}, d \leftrightarrow (x \vee y)\rangle & & \langle\{c,\bar{c}\}, c \leftrightarrow u\rangle \\ \langle\{u,\bar{u}\}, \bar{u}\rangle & & \langle\{x,\bar{x}\}, x \leftrightarrow c\rangle \\ & \Longrightarrow & \langle\{y,\bar{y}\}, y\rangle \\ & & \langle\{d,\bar{d}\}, d \leftrightarrow (x \vee y)\rangle \\ & & \langle\{u,\bar{u}\}, \bar{u}\rangle \end{array}$$

Here is the new semantics for counterfactuals:

(20) $[\![\text{if } \phi, \text{ would } \psi]\!]^{\mathrm{w,g}} = 1$ iff for every premise set $\Pi_{||\phi||_{\mathrm{g}}}$ s.t. $\Pi_{||\phi||_{\mathrm{g}}}$ is a permissible filtering of g(w) relative to $||\phi||_{\mathrm{g}}$, the proposition set of $\Pi_{||\phi||_{\mathrm{g}}}$ entail $||\psi||_{\mathrm{g}}$.

Outline

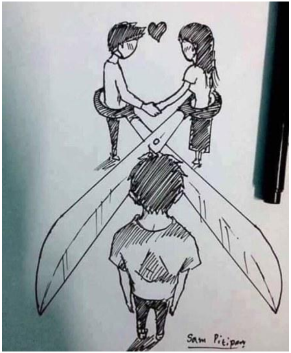
6 Comparation

- A problem
- A loop
- Remarks on filtering logic

♡ Love triangle ♡

Andy, Billy, and Charlie are in a love triangle. Billy is pursuing Andy; Charlie is pursuing Billy; and Andy is pursuing Charlie. Each of them is very annoyed by their suitor and wants to avoid them.

A party is taking place and all three were invited. None of them went, but each of them tracked whether the person they liked was going. Each of them wanted an occasion to spend time with their beloved and without their suitor. Having an occasion of this kind would have been sufficient for each of them to go.



Forward loop counterfactuals:

✓(22) $A \mathrel{\Box\!\!\rightarrow} B$ If Andy was at the party, Billy would be at the party.

✓(24) $C \mathrel{\Box\!\!\rightarrow} A$ If Charlie was at the party, Andy would be at the party.

✓(25) $B \mathrel{\Box\!\!\rightarrow} C$ If Billy was at the party, Charlie would be at the party.

Backward loop counterfactuals:

✗(23) $B \mathrel{\Box\!\!\rightarrow} A$ If Billy was at the party, Andy would be at the party.

✗(26) $A \mathrel{\Box\!\!\rightarrow} C$ If Andy was at the party, Charlie would be at the party.

✗(27) $C \mathrel{\Box\!\!\rightarrow} B$ If Charlie was at the party, Billy would be at the party.

⊗ Problem

The problem is simple: it is impossible to accommodate these judgments in existing kinds of ordering or premise semantics. The proof is particularly quick for Stalnaker's ordering semantics, which assumes that the $\preceq_w$ relation is a strict total order (i.e. all worlds are comparable, and there are no ties: for all w', w'', exactly one of $w' \preceq_w w''$ and $w'' \preceq_w w'$ holds). Here it is:

Since $\preceq_w$ is a strict total order, there is a unique closest world to w that is an A-world, a B-world, or a C-world. Call this world w^*. Without loss of generality, suppose w^* is an A-world. Since $A \Box\!\!\rightarrow B, w^*$ is also a B-world. Since $B \Box\!\!\rightarrow C$, and since $w^*$is the closest B-world, w^* is also a C-world. But then, since the (only) closest A-world is also a C-world, $A \Box\!\!\rightarrow C$ is true. QED.

- Context shift:

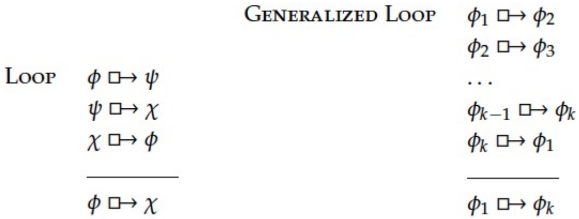
  Forward loop counterfactuals would be evaluated with respect to one ordering source; Backward loop counterfactuals with respect to another.

- Lewis:

  It consigns to the wastebasket of contextually resolved vague- ness something much more amenable to systematic analysis than most of the mess in that wastebasket (1973a, p. 13).

  In the case of (22)–(27), we have no independent reason to think that there is a context shift. Indeed, it would seem extraordinary that context should systematically shift just when we evaluate backward loop counterfactuals.

Here is a more general way of stating the problem. Consider the following inference rule:



Generalized Loop

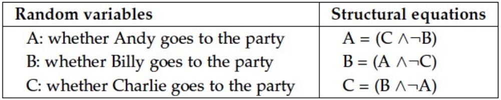
$$\begin{array}{l} \phi_1 \mathrel{\Box\!\!\rightarrow} \phi_2 \\ \phi_2 \mathrel{\Box\!\!\rightarrow} \phi_3 \\ \ldots \\ \phi_{k-1} \mathrel{\Box\!\!\rightarrow} \phi_k \\ \phi_k \mathrel{\Box\!\!\rightarrow} \phi_1 \\ \hline \phi_1 \mathrel{\Box\!\!\rightarrow} \phi_k \end{array}$$

Loop

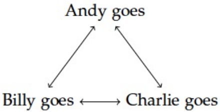
$$\begin{array}{l} \phi \mathrel{\Box\!\!\rightarrow} \psi \\ \psi \mathrel{\Box\!\!\rightarrow} \chi \\ \chi \mathrel{\Box\!\!\rightarrow} \phi \\ \hline \phi \mathrel{\Box\!\!\rightarrow} \chi \end{array}$$

Loop is a valid rule in the logics generated by classical premise semantics, as well as in all standard counterfactual logics. All rules that are instances of Generalized Loop are valid in classical counterfactual semantics. Loop and Generalized Loop show the point of divergence between filtering and classical premise semantics. While they are valid in standard premise semantics, they are invalid in filtering semantics.

Let’s consider a simple causal model and see how filtering semantics invalidates Loop:



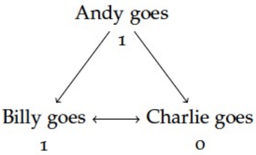
Random variables	Structural equations
A: whether Andy goes to the party	$A = (C \wedge \neg B)$
B: whether Billy goes to the party	$B = (A \wedge \neg C)$
C: whether Charlie goes to the party	$C = (B \wedge \neg A)$



The model has a unique solution, the one on which all the variables have value 0. And the model yields exactly the intuitive verdicts when it is used to evaluate the relevant counterfactuals. See how (22) is evaluated:

(22) If Andy was at the party, Billy would be at the party.

$A = 1$
$B = (A \wedge \neg C)$
$C = (B \wedge \neg A)$



As the graph shows, in the modified model B must have value 1 and C value 0.

It’s easy to see that we get analogous results on filtering semantics. Forward loop counterfactuals ((22), (24), and (25)) are predicted to be true; backwards loop counterfactuals ((26), (23), and (27)) are predictedto be false.

(28)— the initial premise set;

(29)— the only one permissible filtering that the antecedent of (22)) and (26) generates.

(28)
$$\begin{gathered}\langle\{a,\bar{a}\}, a \leftrightarrow (c \wedge \neg b)\rangle \\ \langle\{b,\bar{b}\}, b \leftrightarrow (a \wedge \neg c)\rangle \\ \langle\{c,\bar{c}\}, c \leftrightarrow (b \wedge \neg a)\rangle\end{gathered}$$

(29)
$$\begin{gathered}\langle\{a,\bar{a}\}, a\rangle \\ \langle\{b,\bar{b}\}, b \leftrightarrow (a \wedge \neg c)\rangle \\ \langle\{c,\bar{c}\}, c \leftrightarrow (b \wedge \neg a)\rangle\end{gathered}$$

The propositions in the premises in (29) entail b and c̄, thus yielding the intuitively right predictions for (22) and (26).

Since Loop is valid on standard ordering/premise semantics, this shows that filtering semantics gives rise to a different logic. Let's point out where the new logic departs from standard counterfactual logic.

(A4) $((\phi \mathrel{\Box\!\!\rightarrow} \chi) \wedge (\psi \mathrel{\Box\!\!\rightarrow} \chi)) \supset ((\phi \vee \psi) \mathrel{\Box\!\!\rightarrow} \chi)$

Kraus et al. 1990 point out that (A4), together with some very basic assumptions, allows the derivation of Loop. Hence it's unsurprising that the axiom is invalid in the new semantics. Counterexample:

✓ (25) If Billy was at the party, Charlie would be at the party.

✓ (30) If Andy was at the party and Billy wasn't at the party, Billy would not be at the party.

✗ (31) If Andy was at the party or Billy was at the party, either Billy would not be at the party or Charlie would be at the party.

Outline

7 Conclusion

Three things we have done:

- Implementing causal-models-inspired ideas in a possible worlds semantics for counterfactuals;
- Focusing on the algorithm for resolving inconsistency;
- Implementing this algorithm yields a new kind of possible worlds semantics, which generates a new logic.

Two issues we haven't convered:

- The current version of filtering semantics uses information about causal dependencies and independencies. Hence it's unclear how it would handle noncausal counterfactuals.
- Backtracking counterfactuals, for example:
  (32) If the prisoner had died, one of the two riflemen (or both) would have shot.

References

Santorio (2019). ”Interventions in Premise Semantics.” Philosophers’ Imprint.

Briggs, Rachael (2012). “Interventionist Counterfactuals.” Philosophical studies, 160(1): pp. 139–166.

Burgess, John P (1981). “Quick Completeness Proofs for Some Logics of Conditionals.” Notre Dame Journal of Formal Logic, 22(1): pp. 76–84.

Cariani, Fabrizio, Magdalena Kaufmann, and Stefan Kaufmann (2013). “Deliberative Modality under Epistemic Uncertainty.” Linguistics and Philosophy, 36(3): pp. 225–259.

Chisholm, Roderick M (1946). “The Contrary-to-Fact Conditional.” Mind, 55(219): pp. 289–307.

Condoravdi, Cleo (2002). “Temporal Interpretation of Modals: Modals for the Present and for the Past.” In D. Beaver, S. Kaufmann, B. Clark, and L. Casillas (eds.) The Construction of Meaning, Palo Alto, CA: CSLI Publications.

Dehghani, Morteza, Rumen Iliev, and Stefan Kaufmann (2012). “Causal explanation and fact mutability in counterfactual reasoning.” Mind & Language, 27(1): pp. 55–85.

Fine, Kit (1975). “Review of Lewis’ Counterfactuals.” Mind, 84: pp. 451–458.
Fine, Kit (2012a). “Counterfactuals Without Possible Worlds.” Journal of Philosophy, 109(3): pp. 221–246.
Fine, Kit (2012b). “A Difficulty for the Possible Worlds Analysis of Counterfactuals.” Synthese, 189(1): pp. 29–57.
von Fintel, Kai (2001). “Counterfactuals in a Dynamic Context.” In M. Kenstowicz (ed.) Ken Hale: a life in language, Cambridge, MA: MIT Press, pp. 123–152.
Galles, David, and Judea Pearl (1998). “An axiomatic characterization of causal counterfactuals.” Foundations of Science, 3(1): pp. 151–182.
Gillies, Anthony S (2007). “Counterfactual scorekeeping.” Linguistics and Philosophy, 30(3): pp. 329–360.
Goodman, Nelson (1947). “The Problem of Counterfactual Condition- als.” The Journal of Philosophy, 44(5): pp. 113

Thank You !