

LATD 2010

Semantics of counterfactuals in higher-order fuzzy logic

Libor Běhounek

Institute of Computer Science
Academy of Sciences of the Czech Republic

Ondrej Majer

Institute of Philosophy
Academy of Sciences of the Czech Republic

Counterfactual conditionals

Counterfactuals are conditionals with false antecedents. Their logical analysis is problematic:

If Oswald didn't shoot Kennedy, someone else did.

If Oswald had not shot Kennedy, someone else would have.

The first sentence is true, the second perhaps not. If we read the conditionals as material implications, both sentences turn out to be true—a simple logical analysis does not work.

There have been several attempts to propose an adequate semantics for counterfactuals (most notably by Nelson Goodman), but the most widely accepted semantics was proposed independently by David Lewis and Robert Stalnaker.

Lewis' semantics

Lewis' approach is based on a relation which orders possible worlds wrt their *similarity* to the actual world.

Alternatively he uses the notion of *closeness* $v \leq_w v'$ —the world v is at least as close to the actual world w as the world v' , s.t.:

- $w <_w v$ for any $v \neq w$ (actual world strictly minimal)
- for any v, v' , either $v \leq_w v'$ or $v' \leq_w v$ (linearity)
- if $v \leq_w v'$ and $v' \leq_w u$ then $v \leq_w u$ (transitivity)

Lewis' truth conditions for counterfactuals

The counterfactual conditional $A \square \rightarrow C$ is true at a world w with respect to a similarity ordering \leq_w iff either

- for all v , $v \not\models A$
(ie, there are no A -worlds), or
- there is a v' such that $v' \models A \wedge C$ and
for every v if $v \models A \wedge \neg C$ then $v' < v$
(there is an AC -world which is closer to the actual world
than any $A\neg C$ -world)

Lewis' axiomatic system

Axioms (The logic of counterfactuals VC):

- axioms of the classical propositional calculus
- $A \Box \rightarrow A$
- $(\neg A \Box \rightarrow A) \rightarrow (A \Box \rightarrow B)$
- $(A \Box \rightarrow \neg B) \vee (((A \wedge B) \Box \rightarrow C) \leftrightarrow (A \Box \rightarrow (B \rightarrow C)))$
- $(A \Box \rightarrow B) \rightarrow (A \rightarrow B)$
- $(A \wedge B) \rightarrow (A \Box \rightarrow B)$

Rules: modus ponens, substitution,
deduction within conditionals:

$$\frac{(B_1 \wedge \dots \wedge B_n) \rightarrow C}{(A \Box \rightarrow B_1) \wedge \dots \wedge (A \Box \rightarrow B_n) \rightarrow (A \Box \rightarrow C)}$$

Properties of Lewis' counterfactuals

Counterfactual conditionals do not obey some standard properties of material conditional:

(intuitive counterexamples are easy to find)

Weakening:
$$\frac{A \Box \rightarrow C}{A \wedge B \Box \rightarrow C}$$

Contraposition:
$$\frac{A \Box \rightarrow C}{\neg C \Box \rightarrow \neg A}$$

Transitivity:
$$\frac{A \Box \rightarrow B, B \Box \rightarrow C}{A \Box \rightarrow C}$$

Why a fuzzy semantics for counterfactuals?

Lewis' semantics is based on the notion of **similarity**
of possible worlds

Similarity relations are prominently studied in **fuzzy mathematics**

⇒ Let's see if fuzzy mathematics can provide a viable semantics
for counterfactuals

We shall work in higher-order MTL
(or any other higher-order t-norm logic)

Higher-order fuzzy logic

- Connectives and quantifiers as in first-order fuzzy logics
- Higher-order predicates x, P, Q, \dots
- Primitive (fuzzy) membership $x \in A, P \in Q, \dots$
- Crisp identity $=$, tuples, comprehension terms $\{x \mid \varphi\}$
- Comprehension axioms: $y \in \{x \mid \varphi(x)\} \leftrightarrow \varphi(y)$
- Extensionality axioms: $(\forall x) \Delta(x \in A \leftrightarrow x \in B) \rightarrow A = B$

Models:

First-order predicates = fuzzy sets of individuals

Higher-order predicates = fuzzy sets of fuzzy sets etc.

Soundness wrt intended models

Completeness wrt generalized models

(as in classical higher-order logic)

Defined notions

$$\varphi \leq \psi \equiv \Delta(\varphi \rightarrow \psi)$$

$$\varphi = \psi \equiv \Delta(\varphi \leftrightarrow \psi)$$

$$\emptyset = \{x \mid 0\}$$

$$A \cap B = \{x \mid x \in A \wedge x \in B\}$$

$$A \subseteq B \equiv (\forall x)(x \in A \rightarrow x \in B)$$

Similarity relations = fuzzy equivalence relations (to degree 1)

Axioms: Sxx , $Sxy \rightarrow Syx$, $Sxy \ \& \ Syz \rightarrow Sxz$

Ordering worlds by similarity

Σxy ... the world x is similar to the world y

$$x \preceq_w y \equiv \Sigma wy \lesssim \Sigma wx$$

... x is *more or roughly as* similar to w as y

\lesssim rather than \leq , by the fuzzy paradigm: when reasonable, use indistinguishability (or similarity) rather than strict equality

Worlds indistinguishable from x (to a large degree)

should play a role (to a large degree)

\Rightarrow We also need a similarity relation on degrees

Similarity on degrees (of the similarity of worlds)

Axioms for \sim :

$$(\alpha \sim \alpha)$$

$$(\alpha \sim \beta) \rightarrow (\beta \sim \alpha)$$

$$(\alpha \sim \beta) \& (\beta \sim \gamma) \rightarrow (\alpha \sim \gamma)$$

= similarity

$$(\alpha \sim \beta) \rightarrow (\alpha \leftrightarrow \beta)$$

$$\vdash \alpha \& (\alpha \sim \beta) \rightarrow \beta$$

$$\vdash \Delta(\alpha \sim \beta) \leftrightarrow (\alpha = \beta)$$

$$(\exists \beta \neq \alpha)(\beta \sim \alpha)$$

Def: $\alpha \lesssim \beta \equiv (\alpha < \beta) \vee (\alpha \sim \beta)$

$$(\alpha \lesssim \alpha)$$

$$(\alpha \lesssim \beta) \& (\beta \lesssim \alpha) \rightarrow (\alpha \sim \beta)$$

$$(\alpha \lesssim \beta) \& (\beta \lesssim \gamma) \rightarrow (\alpha \lesssim \gamma)$$

= fuzzy ordering

$$(\alpha \lesssim \beta) \rightarrow (\alpha \rightarrow \beta)$$

$$\vdash \alpha \& (\alpha \lesssim \beta) \rightarrow \beta$$

$$\vdash \Delta(\alpha \lesssim \beta) \leftrightarrow (\alpha \leq \beta)$$

$$(\exists \beta \not\leq \alpha)(\beta \lesssim \alpha)$$

Fuzzy semantics for counterfactuals

Recall: $x \preceq_w y \equiv \Sigma wy \lesssim \Sigma wx$

The closest A -worlds: $\text{Min}_{\preceq_w} A = \{x \mid x \in A \wedge (\forall a \in A)(x \preceq_w a)\}$

The properties of minima in fuzzy orderings are well known

(cf. Běhounek–Bodenhofer–Cintula, FSS 2008)

Define: $\|A \square \rightarrow B\|_w \equiv (\text{Min}_{\preceq_w} A) \subseteq B$

... the closest A -worlds are B -worlds (fuzzily!)

Properties of fuzzy counterfactuals

Non-triviality: $\Delta(A \square \rightarrow B)$ for all B only if $A = \emptyset$

Non-desirable properties are invalid:

$$\not\vdash (A \square \rightarrow B) \ \& \ (B \square \rightarrow C) \rightarrow (A \square \rightarrow C)$$

$$\not\vdash (A \square \rightarrow C) \rightarrow (A \ \& \ B \square \rightarrow C)$$

$$\not\vdash (A \square \rightarrow C) \rightarrow (\neg C \square \rightarrow \neg A)$$

Desirable properties are valid, eg:

$$\vdash \square(A \rightarrow B) \longrightarrow (A \square \rightarrow B) \longrightarrow (A \rightarrow B)$$

Properties of fuzzy counterfactuals

Many theorems on $\Box \rightarrow$ easily derivable in higher-order MTL, eg:

$$\vdash \Box B \rightarrow (A \Box \rightarrow B)$$

$$\vdash \Box \neg A \rightarrow (A \Box \rightarrow B)$$

$$\vdash (A \Box \rightarrow B) \ \& \ \Box(B \rightarrow B') \rightarrow (A \Box \rightarrow B')$$

$$\vdash (A \Box \rightarrow B) \ \& \ \Box(A \leftrightarrow A') \rightarrow (A' \Box \rightarrow B) \quad (\not\vdash \text{ for } \Box(A \rightarrow A'))$$

Some of Lewis' tautologies only hold for full degrees, eg:

$$\vdash \Delta \Box \neg A \leftrightarrow \Delta(A \Box \rightarrow \perp) \quad (\text{but } \not\vdash \text{ without the } \Delta\text{'s})$$

Conclusions

Advantages:

- Automatic accommodation of gradual counterfactuals
“If ants were *large*, they would be *heavy*.”
- Accommodation of graduality of counterfactuals
(some counterfactual conditionals seem to hold
to larger degrees than others)
- Standard fuzzy handling of the similarity of worlds

Disadvantages:

- Needs non-classical logic for semantic reasoning
(but a well-developed one \Rightarrow a low cost for experts)

Future work:

- Axiomatization of the logic of fuzzy counterfactuals
- Fuzzification of Lewis' second system with internalized \approx