A Semantics for Counterfactuals Based on Fuzzy Logic

In M. Peliš, V. Punčochář (eds.): *The Logica Yearbook 2010,* pp. 25–41, College Publications, 2011.

Libor Běhounek* Ondrej Majer†

Abstract

Lewis–Stalnaker's semantics for counterfactuals is based on the notion of similarity of possible worlds. Since the general notion of similarity is prominently studied in fuzzy mathematics, where it is modeled by fuzzy equivalence relations, it is natural to attempt at reconstructing Lewis' and Stalnaker's ideas in terms of fuzzy similarities. This paper sketches such a reconstruction; full details will be presented in an upcoming paper. We demonstrate that the approach is viable, adequate with respect to the expected properties of counterfactuals, and provides meaningful generalizations of the classical account.

1 Introduction

Counterfactual conditionals, or simply counterfactuals, are conditionals with false antecedents; i.e., conditionals of the form "if it were the case that A, then it would be the case that C", written $A \mathrel{\Box\!\!\rightarrow} C$. Their logical analysis is notoriously problematic: if analyzed straightforwardly by material implication, they would always come out true; nevertheless, intuitively some counterfactuals seem to be true while others false. Consider, e.g., Goodman's (1947) example of the pair of

*Supported by grants No. IAA900090703 of GA AV ČR and P202/10/1826 of GA ČR.

†Supported by grant No. ICC/08/E018 of GA ČR (a part of ESF EUROCORES–LogICCC project FP006).

counterfactuals regarding a piece of butter that has been consumed and was never heated:

(a) If that piece of butter had been heated to 150 F, it would have melted.

(b) If that piece of butter had been heated to 150 F, it would not have melted.

Clearly one of these two sentences (presumably the latter) should be considered false, even though both have false antecedents.

The logical analysis of counterfactual conditionals has for a long time been an important issue in philosophical logic. A notable early attempt to propose an adequate semantics for counterfactuals was done by Nelson Goodman (1947), who also stressed its importance, not only for logic, but as well for the philosophy of science. The most influential analysis was provided independently by Robert Stalnaker (1968) and David Lewis (1973). Their solution is based on the notion of similarity which compares possible worlds according to their (subjective) closeness to the actual world. In the present paper, we shall mostly follow Lewis' formulation of the semantics. According to Lewis, a counterfactual $A \mathrel{\Box\!\!\rightarrow} C$ is true in the actual world iff either A does not hold in any world (then $A \mathrel{\Box\!\!\rightarrow} C$ is trivially true), or all worlds in which both A and C hold are closer (in a given ordering of worlds by similarity to the actual world) to the actual world than some world(s) in which A and $\neg C$ hold.

The general notion of similarity (not just of worlds) is prominently studied in fuzzy mathematics, where it is modeled by fuzzy equivalence relations (Zadeh, 1971). It is, therefore, natural to attempt a reconstruction of Lewis' and Stalnaker's ideas in terms of fuzzy similarities. This paper offers such a reconstruction in the framework of t-norm fuzzy logics (see esp. Hájek, 1998). We demonstrate that the approach is viable and that the resulting semantics is adequate with respect to the expected properties of counterfactuals (i.e., it validates their intuitive and invalidates their counter-intuitive properties). The merits of the fuzzy approach to counterfactuals are briefly discussed in the concluding section. The paper provides a sketch of the approach; more details will be found in an upcoming full paper.

2 Lewis' semantics of counterfactuals

Lewis' approach is based on the notion of (subjective) *similarity* of possible worlds.[1] The intuitive notion of similarity of worlds can be formally rendered in a number of ways, and the rendering impacts the resulting properties of the logic of counterfactuals. Lewis himself (1973) provides a variety of counterfactual logics differing in the properties of the similarity ordering, and also several reformulations that replace the similarity order by alternative notions. Here we shall present his formulation in terms of the relation of *closeness* between possible worlds, with notation $v \leq_w v'$ understood as "the world v is at least as close to the actual world w as the world v'".

Various sets of natural properties of the closeness relation can be assumed, generating different logics of counterfactuals. For instance, the following assumptions yield Lewis' counterfactual logic VC:

- Strict minimality of the actual world: $w <_w v$ for any $v \neq w$
- Linearity: for any v, v', either $v \leq_w v'$ or $v' \leq_w v$
- Transitivity: if $v \leq_w v'$ and $v' \leq_w u$, then $v \leq_w u$

The counterfactual conditional $A \mathrel{\Box\!\!\to} C$ is defined to be true in a world w with respect to a closeness relation $\leq_w$, iff:

- Either for all v, $v \not\models A$ (i.e., 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 $v' <_w v$ (i.e., there is an AC-world which is closer to the actual world then any $A\neg C$-world)

Lewis (1973) showed that the logic VC can be axiomatized by the

[1] The similarity relation is a primitive parameter of the semantic model. How the relation is actually obtained is of no concern for the logic of counterfactuals: it is simply assumed to be given, similarly as subjective probabilities are assumed to be given in probability theory.

following axioms:

$$A \mathrel{\Box\!\!\rightarrow} A \tag{1}$$
$$(\neg A \mathrel{\Box\!\!\rightarrow} A) \rightarrow (A \mathrel{\Box\!\!\rightarrow} B) \tag{2}$$
$$(A \mathrel{\Box\!\!\rightarrow} \neg B) \vee (((A \wedge B) \mathrel{\Box\!\!\rightarrow} C) \leftrightarrow (A \mathrel{\Box\!\!\rightarrow} (B \rightarrow C))) \tag{3}$$
$$(A \mathrel{\Box\!\!\rightarrow} B) \rightarrow (A \rightarrow B) \tag{4}$$
$$(A \wedge B) \rightarrow (A \mathrel{\Box\!\!\rightarrow} B) \tag{5}$$

plus the axioms of classical propositional calculus and the rules of modus ponens, substitution, and the rule of deduction within conditionals,

$$\frac{(B_1 \wedge \ldots \wedge B_n) \rightarrow C}{(A \mathrel{\Box\!\!\rightarrow} B_1) \wedge \ldots \wedge (A \mathrel{\Box\!\!\rightarrow} B_n) \rightarrow (A \mathrel{\Box\!\!\rightarrow} C)} \tag{6}$$

The modality $\Box$ of necessity, with the usual meaning of truth in all accessible worlds, is definable in Lewis's system VC in terms of counterfactual implication:

$$\Box A \equiv \neg A \mathrel{\Box\!\!\rightarrow} \bot \tag{7}$$

Various intuitively plausible properties of counterfactuals are derivable in the logic VC, for instance the following ones:

$$\Box(A \rightarrow B) \rightarrow (A \mathrel{\Box\!\!\rightarrow} B) \tag{8}$$
$$\Box\neg A \rightarrow (A \mathrel{\Box\!\!\rightarrow} B) \tag{9}$$
$$\Box B \rightarrow (A \mathrel{\Box\!\!\rightarrow} B) \tag{10}$$
$$(A \mathrel{\Box\!\!\rightarrow} B) \mathbin{\&} \Box(B \rightarrow B') \rightarrow (A \mathrel{\Box\!\!\rightarrow} B') \tag{11}$$
$$(A \mathrel{\Box\!\!\rightarrow} B) \mathbin{\&} \Box(A \leftrightarrow A') \rightarrow (A' \mathrel{\Box\!\!\rightarrow} B) \tag{12}$$

By (4) and (8), the logical strength of counterfactual implication is intermediate between those of material and strict implication. Unlike material conditionals, counterfactual conditionals do *not* obey, i.a., the following laws:

- Weakening: $\dfrac{A \mathrel{\Box\!\!\rightarrow} C}{(A \mathbin{\&} B) \mathrel{\Box\!\!\rightarrow} C}$
- Contraposition: $\dfrac{A \mathrel{\Box\!\!\rightarrow} C}{\neg C \mathrel{\Box\!\!\rightarrow} \neg A}$

- Transitivity: $\dfrac{A \mathrel{\Box\!\!\rightarrow} B, B \mathrel{\Box\!\!\rightarrow} C}{A \mathrel{\Box\!\!\rightarrow} C}$

Informal counterexamples to these rules as well as formal counterexamples in Lewis' semantics can easily be constructed (or see Lewis, 1973).

3 Formal fuzzy logic

Fuzzy logics are many-valued logics suitable mainly for *gradual* predicates, such as *tall, old, warm,* etc., whose underlying quantities can be measured by real numbers (e.g., height in cm's, age in years, temperature in Fahrenheits, etc.). In order to be measured on a common scale, the underlying quantities are conventionally transformed into the unit interval $[0, 1]$ (or another suitable algebra); the transformed values are called the *degrees* of the gradual predicates (e.g., the degrees of tallness, warmness, etc.). Certain algebraic operations are defined on $[0, 1]$ that represent logical connectives and quantifiers, and a degree-based consequence relation between $[0, 1]$-valued gradual propositions is studied.

Different fuzzy logics arise by different admissible choices of algebraic operations that realize logical connectives on $[0, 1]$. A prototypical (and probably the best known) example of fuzzy logic is infinite-valued Łukasiewicz logic; other members of the family are, for instance, Gödel–Dummett logic, product fuzzy logic, fuzzy logics BL, MTL, etc. For the sake of simplicity, we shall in this paper restrict our attention to Łukasiewicz logic, even though all our considerations, derivations, and semantic examples actually work for any well-behaved formal fuzzy logic.[2] For basic information on formal fuzzy logic see, e.g., (Hájek, 2010); more technical details can be found, e.g., in (Hájek, 1998). In this section, we shall briefly recall basic definitions of Łukasiewicz logic.

In the standard $[0, 1]$-valued semantics of Łukasiewicz logic, propo-

[2]Namely, for any extension of the fuzzy logic MTL, which is arguably the weakest t-norm–based fuzzy logic suitable for fuzzy mathematics.

sitional connectives are realized by the following operations on $[0,1]$:

$$
\begin{aligned}
x \to y &= \max(1-x+y, 1)\\
x \mathbin{\&} y &= \min(0, x+y-1)\\
x \wedge y &= \min(x, y)\\
x \vee y &= \max(x, y)\\
x \leftrightarrow y &= 1-|x-y|\\
\neg x &= 1-x\\
\Delta x &= 1 \text{ iff } x = 1, \text{ otherwise } \Delta x = 0
\end{aligned}
$$

Notice that Łukasiewicz logic possesses two conjunctive connectives: idempotent $\wedge$ and non-idempotent $\&$. The presence of these two conjunctions is not surprising, considering the fact that Łukasiewicz logic belongs to contraction-free substructural logics (Ono, 2003) all of which possess two conjunctions.[3] The connective Δ, expressing the full degree of a gradual proposition, enables interpretation of classical logic within Łukasiewicz logic (as its connectives and quantifiers behave classically on the values 0 and 1).

Tautologies of propositional Łukasiewicz logic are those formulae that always evaluate to 1. The set of propositional tautologies of Łukasiewicz logic is finitely axiomatizable by a Hilbert-style calculus (see Hájek, 1998).

First-order Łukasiewicz logic extends the propositional syntax in the usual way. Semantically, n-ary first-order predicates are interpreted by n-ary functions from a fixed set of individuals (the universe of discourse) to $[0,1]$. The quantifiers $\forall$ and $\exists$ are interpreted, respectively, as the infimum and supremum of the degrees of all instances of the quantified formula. First-order Łukasiewicz logic can be axiomatically approximated by adding Rasiowa's (1974) axioms for quantifiers to propositional Łukasiewicz logic.[4]

[3] Łukasiewicz logic, similarly as other contraction-free substructural logics, also has another, non-idempotent disjunction $x \oplus y = \min(1, x+y)$, which, however, will not be needed in this paper. (In (29) below, which is the only place where disjunction occurs in this paper, the variants with $\oplus$ and $\vee$ happen to be equivalent, since one of the disjuncts is bivalent.)

[4] Rasiowa's axioms are only complete with respect to a generalized semantics of first-order Łukasiewicz logic, which evaluates formulae in more general algebras than $[0,1]$. The standard $[0,1]$-valued first-order semantics is not finitarily axiomatizable, though it can be axiomatized by an infinitary rule—see (Hay, 1963).

4 Higher-order fuzzy logic

First-order Łukasiewicz logic is sufficiently strong for supporting non-trivial axiomatic mathematical theories. For smooth mathematical work it is expedient to define a Russell-style simple type theory over Łukasiewicz logic (or Henkin-style higher-order Łukasiewicz logic, see Běhounek & Cintula, 2005). The latter is an axiomatic theory over multi-sorted first-order Łukasiewicz logic, with sorts of variables for individuals from a fixed universe of discourse (zeroth-order variables) and for predicates of all orders $n \geq 1$ and arities $k \geq 0$ (n-th-order variables). The language of the theory comprises:

- The primitive memberships predicates $\in$ between successive sorts,
- The bivalent identity predicate $=$ on each sort,
- The functions $\langle \dots \rangle$ for tuples of each order and arity, and
- The comprehension terms $\{x \mid \varphi\}$ of order $n+1$ for each well-typed formula φ and the variable x of each order n.

The theory is axiomatized by the following axiom schemata:

- The identity axioms $x = x$ and $x = y \to (\varphi(x) \to \varphi(y))$, for each well-typed formula φ and each sort of variables;
- The usual technical axioms for handling tuples of each arity and order;
- Comprehension axioms $y \in \{x \mid \varphi(x)\} \leftrightarrow \varphi(y)$ for each well-typed formula φ and each sort of variables;
- Extensionality axioms $(\forall x)\Delta(x \in A \leftrightarrow x \in B) \to A = B$ for all orders.

In the intended models of the theory,[5] individual variables range over a fixed set X (the universe of discourse). First-order k-ary predicates are interpreted as $[0,1]$-valued fuzzy sets (for $k = 1$) or k-ary

Nevertheless, Rasiowa's axiomatic approximation is sufficient for almost all practical purposes.

[5] Just like in classical higher-order logic, the theory of intended models is not recursively axiomatizable (as it interprets true arithmetic). The above axiomatization is sound with respect to intended models, but complete only with respect to a more general class of models. Nevertheless, the axiomatic approximation is again sufficient for almost all practical purposes.

fuzzy relations (for $k > 1$) on X; i.e., as functions $X^k \to [0,1]$. Second-order k-ary predicates are interpreted as fuzzy sets (or relations) of fuzzy sets (or relations) of individuals; i.e., as functions $(X_1)^k \to [0,1]$, where X_1 is the range of 1st-order predicate variables. In general, $(n+1)$-st-order k-ary predicates are interpreted as fuzzy sets (or relations) of order $n+1$; i.e., as functions $(X_n)^k \to [0,1]$, where X_n is the range of n-th-order predicate variables. The membership predications $x \in A$ are assigned the value (in $[0,1]$) that the function interpreting A assigns to the interpretation of x.

Comprehension terms $\{x \mid \varphi(x)\}$ (of any order) denote fuzzy sets to which each element x belongs to the degree of $\varphi(x)$. Classical (bivalent) sets can be represented by fuzzy sets whose membership functions only take values in the two-element set of degrees $\{0,1\}$. The condition of being bivalent is expressible as $(\forall x)\Delta(x \in A \vee x \notin A)$, where $x \notin A$ abbreviates $\neg(x \in A)$.

Various fuzzy mathematical notions can be defined over Łukasiewicz logic by reinterpreting the formulae of classical mathematics in higher-order Łukasiewicz logic. In what follows, we shall need the following elementary defined notions:

$$\emptyset = \{x \mid 0\} \tag{13}$$

$$A \subseteq B \equiv (\forall x)(x \in A \to x \in B) \tag{14}$$

Moreover we shall use the following defined connectives that compare the degrees of formulae:

$$\varphi \leq \psi \equiv \Delta(\varphi \to \psi) \tag{15}$$

$$\varphi = \psi \equiv \Delta(\varphi \leftrightarrow \psi) \tag{16}$$

The definition is justified by the fact that in Łukasiewicz logic (as well as in other fuzzy logics), the degree of $\varphi \to \psi$ is 1 iff the degree of ψ is at least as large as the degree of φ, and the degree of $\varphi \leftrightarrow \psi$ is 1 iff the degrees of φ and ψ are equal.[6]

[6] The defined connective = of (16), expressing the identity of degrees, should not be confused with the identity predicates on each sort of variables: though denoted here by the same sign, they are always disambiguated by the context, as the arguments of the latter are terms, while the arguments of the former are formulae.

5 Similarity relations

Similarity relations are in fuzzy mathematics standardly modeled as fuzzy equivalence relations (Zadeh, 1971), i.e., binary fuzzy relations S that satisfy (to degree 1) the axioms of fuzzy reflexivity, symmetry, and transitivity in Łukasiewicz (or another fuzzy) logic:

$$\Delta(\forall x)Sxx \tag{17}$$

$$\Delta(\forall xy)(Sxy \to Syx) \tag{18}$$

$$\Delta(\forall xyz)(Sxy \,\&\, Syz \to Sxz) \tag{19}$$

A stronger notion of *separated* (or *unimodal*) fuzzy equivalence (sometimes also called *fuzzy equality*) strengthens the reflexivity condition (17) by the additional requirement that only identical elements are fully similar:

$$(\forall xy)(\Delta Sxy \leftrightarrow x = y) \tag{20}$$

Let a fuzzy relation $S^M\colon X^2 \to [0,1]$ be the interpretation of the predicate S in a given standard $[0,1]$-valued semantic model M (with the universe X) for Łukasiewicz logic. Then the function $d(x,y) = 1 - S^M xy$ is a (bounded) pseudometric iff S is a fuzzy equivalence relation (i.e., satisfies (17)–(19) to degree 1) in M, and is a bounded metric iff the fuzzy equivalence is separated (i.e., satisfies (18)–(20) to degree 1 in M).[7]

The correspondence to (pseudo)metrics and the fact that they can be conveniently handled by means of formal fuzzy logic (cf. Běhounek, Bodenhofer, & Cintula, 2008) makes fuzzy equivalence relations a suitable representation of gradual relations of closeness or similarity. Consequently, they have become the standard model of similarity relations in fuzzy mathematics, and we shall call fuzzy equivalence relations simply (fuzzy) similarities further on. It should, however, be stressed that formal fuzzy logics admit other algebras of degrees besides the $[0,1]$ interval (see, e.g., Hájek, 1998). Fuzzy equivalence relations thus generalize real-valued (pseudo)metrics by admitting various scales of

[7] In a similar manner, fuzzy equivalence relations correspond to (generally unbounded) pseudometrics $d(x,y) = -\log S^M xy$ in product logic, pseudoultrametrics in Gödel logic, and various generalizations of pseudometrics in other fuzzy logics (differing in the operation used in the triangle inequality), and to corresponding kinds of metrics if the equivalence relation is separated.

abstract degrees of closeness (or similarity); consequently, modeling similarity or closeness by equivalence relations over fuzzy logic does *not* enforce measuring them by real numbers.

6 Similarity of worlds and degrees

Applying the apparatus of fuzzy similarities to Lewis' concept of the similarity of possible worlds, we assume that there is a fuzzy relation Σ on the (classical, bivalent) set W of possible worlds, which satisfies the axioms of fuzzy similarity (17)–(19). The formula Σxy is informally interpreted as "the world x is similar to the world y", and the degree (from $[0,1]$ or another algebra of degrees admissible for Łukasiewicz logic) that is assigned to it in a particular model of (higher-order) Łukasiewicz logic is interpreted as the degree of similarity between these worlds.

The first idea how to model Lewis' ternary relation "the world x is at least as close to the actual world w as the world y" in terms of the similarity relation Σ is to define it straightforwardly as $\Sigma yw \leq \Sigma xw$, where $\leq$ is the degree-comparing connective (15). However, for reasons both methodological and technical, it turns out to be more appropriate to use a *fuzzy* comparison $\lesssim$ of degrees, rather than the bivalent comparison connective $\leq$, defining:

$$x \preccurlyeq_w y \;\equiv\; \Sigma yw \lesssim \Sigma xw. \tag{21}$$

The intended definition could be informally interpreted as "x is more or *roughly as* similar to w as y". This form follows the fuzzy paradigm to employ, whenever reasonable, a (fuzzy) indistinguishability (or similarity) relation rather than bivalent equality. The particular motivation for using $\lesssim$ instead of $\leq$ is the natural assumption that worlds indistinguishable from x (as regards their closeness to the actual world) should in the evaluation of counterfactuals play a rôle similar to that of x.

In order to define the fuzzy ordering $\lesssim$ of degrees (of similarity to the actual world) informally interpretable as "more or roughly as", we need to define another similarity, $\sim$, this time on degrees. Besides the usual axioms of similarity, we shall require $\sim$ to satisfy two additional conditions that specify the kind of similarity relations suitable for our purposes. Omitting the initial Δ's (for satisfaction to degree 1) and quantification over all degrees, the conditions on $\sim$ can be listed as follows:

$$(\alpha \sim \alpha) \tag{22}$$
$$(\alpha \sim \beta) \to (\beta \sim \alpha) \tag{23}$$
$$(\alpha \sim \beta) \mathbin{\&} (\beta \sim \gamma) \to (\alpha \sim \gamma) \tag{24}$$
$$(\alpha \sim \beta) \to (\alpha \leftrightarrow \beta) \tag{25}$$
$$(\exists \beta \neq \alpha)(\beta \sim \alpha) \tag{26}$$
$$(\alpha \leq \beta \leq \gamma) \to ((\alpha \sim \gamma) \to (\alpha \sim \beta)) \tag{27}$$
$$(\gamma \leq \beta \leq \alpha) \to ((\alpha \sim \gamma) \to (\alpha \sim \beta)) \tag{28}$$

The conditions (22)–(24) are just the axioms of fuzzy similarity. The condition (25) of congruence with respect to the equivalence connective expresses the substitutivity of similar degrees in gradual inference: its equivalent formulation is $\alpha \mathbin{\&} (\alpha \sim \beta) \to \beta$. It furthermore ensures that the similarity $\sim$ is separated, since $\Delta(\alpha \sim \beta) \leftrightarrow (\alpha = \beta)$ is a corollary of (25); in other words, each degree is *fully* similar only to itself. The condition (26) ensures that $\sim$ differs from the bivalent identity $=$, but still to each degree there are arbitrarily similar degrees.[8] The latter two conditions also ensure that $\sim$ is indeed fuzzy, as by (25) and (26) the relation $\sim$ has to be infinite-valued. Finally, the conditions (27)–(28) express the compatibility of $\sim$ with the ordering $\leq$ of degrees: it ensures that closer (in the sense of $\leq$) degrees are (non-strictly) more similar to each other.

By means of the relation $\sim$, the fuzzy comparison "more or roughly as" $\lesssim$ of degrees can be defined as follows:

$$\alpha \lesssim \beta \ \equiv \ (\alpha < \beta) \vee (\alpha \sim \beta). \tag{29}$$

The conditions (22)–(26) then entail the following corresponding prop-

[8]By the semantics of $\exists$ in Łukasiewicz logic, (26) ensures that for each α, the supremum of the *degrees of similarity* of other degrees to α is 1.

erties of $\lesssim$:

$$(\alpha \sim \beta) \to (\alpha \lesssim \beta) \tag{30}$$
$$(\alpha \lesssim \beta) \,\&\, (\beta \lesssim \alpha) \to (\beta \sim \alpha) \tag{31}$$
$$(\alpha \lesssim \beta) \,\&\, (\beta \lesssim \gamma) \to (\alpha \lesssim \gamma) \tag{32}$$
$$(\alpha \lesssim \beta) \to (\alpha \to \beta) \tag{33}$$
$$(\alpha \leq \beta) \to (\alpha \lesssim \beta) \tag{34}$$
$$(\exists \beta \neq \alpha)(\beta \lesssim \alpha) \tag{35}$$
$$(\alpha \leq \beta \leq \gamma) \to ((\gamma \lesssim \alpha) \to (\beta \lesssim \alpha)) \tag{36}$$
$$(\gamma \leq \beta \leq \alpha) \to ((\alpha \lesssim \gamma) \to (\alpha \lesssim \beta)) \tag{37}$$

The properties (30)–(32) state that $\lesssim$ is a *similarity-based fuzzy ordering* (for which see Bodenhofer, 2000) with respect to the similarity $\sim$. The conditions (33)–(35) entail corollaries analogous to those of (25) and (26), namely:

$$\begin{gathered} \alpha \,\&\, (\alpha \lesssim \beta) \to \beta \\ \Delta(\alpha \lesssim \beta) \leftrightarrow (\alpha \leq \beta), \end{gathered} \tag{38}$$

and also enforce $\lesssim$ to be a fuzzy (rather than bivalent) ordering of degrees, strictly weaker than $\leq$. The sets of axioms (22)–(28) and (30)–(37) are in fact equivalent, as the former can be obtained from the latter if $\sim$ is defined as the symmetrization of $\lesssim$, i.e., $\alpha \sim \beta \equiv (\alpha \lesssim \beta) \wedge (\beta \lesssim \alpha)$, or equivalently, $(\alpha \lesssim \beta) \,\&\, (\beta \lesssim \alpha)$.

By means of the fuzzy ordering $\lesssim$ on degrees (of similarity of worlds), we can now define Lewis' ternary relation $\preccurlyeq$ of closeness of worlds as intended above in (21), namely:

$$x \preccurlyeq_w y \;\equiv\; \Sigma yw \lesssim \Sigma xw,$$

interpreted as "x is more or roughly as similar to w as y".

7 Fuzzy semantics of counterfactuals

Having interpreted the similarity of worlds and Lewis' ternary relation of closeness of worlds in the fuzzy setting, we can now define the semantics of counterfactuals based on these notions. The primitive parameters of the semantics, supposed just to be given (cf. footnote 1 above), are the fuzzy similarity relation Σ on the (bivalent) set W of possible worlds and the fuzzy similarity relation $\sim$ on degrees (of similarity of possible worlds), assumed to satisfy the axioms given in Section 6. As usual in intensional semantics, the meanings of propositions are identified with (here in general fuzzy) sets of possible worlds.

The definition of the semantics of counterfactuals in the fuzzy setting will be based on the straightforward idea that the counterfactual $A \mathrel{\Box\!\!\rightarrow} C$ is true in the world w if the closest (with respect to w) A-worlds are also C-worlds.[9] The A-worlds closest to w can be defined as the fuzzy set of *minima* of a fuzzy set A of worlds in a fuzzy relation (not necessarily an ordering) $\preccurlyeq_w$ (cf. Běhounek et al., 2008):

$$\mathrm{Min}_{\preccurlyeq_w} A \;=\; \{x \mid x \in A \wedge (\forall a)(a \in A \rightarrow x \preccurlyeq_w a)\},$$

i.e., the fuzzy set of elements of A at least as close to w as any element of A (which is the classical definition just reinterpreted in terms of Łukasiewicz logic). Basic properties of minima in fuzzy relations are easily derivable in higher-order fuzzy logic (cf. Běhounek et al., 2008).

Given a model M of higher-order Łukasiewicz logic and the parameters Σ and $\sim$, the semantic value (or *extension*) of the counterfactual $A \mathrel{\Box\!\!\rightarrow} C$ in the possible world w can now be defined as follows:

$$(A \mathrel{\Box\!\!\rightarrow} C)_w \;\equiv\; (\mathrm{Min}_{\preccurlyeq_w} A) \subseteq C, \tag{39}$$

where $\subseteq$ is fuzzy inclusion defined in (14). The definition expresses, in the fuzzy sense, that all $\preccurlyeq_w$-minimal (i.e., closest to w) A-worlds are also C-worlds. Notice that since the right-hand side of (39) is evaluated in Łukasiewicz logic, $(A \mathrel{\Box\!\!\rightarrow} C)_w$ denotes a *degree* in $[0, 1]$ (or another algebra admissible for Łukasiewicz logic).

The meaning (or *intension*) of the counterfactual proposition $A \mathrel{\Box\!\!\rightarrow} C$ is then identified with the fuzzy set of possible worlds to which a possible world w belongs to degree $(A \mathrel{\Box\!\!\rightarrow} C)_w$, i.e.,

$$(A \mathrel{\Box\!\!\rightarrow} C) \;=\; \{w \mid (A \mathrel{\Box\!\!\rightarrow} C)_w\}.$$

With these definitions at hand, a standard account of fuzzy intensional semantics for propositional logic of counterfactuals can already

[9] This actually corresponds more to Stalnaker's approach than to Lewis'. It will be seen in Section 8 (cf. footnote 10) that unlike in Stalnaker's bivalent framework, in our fuzzy setting this *does not* involve the implausible Limit Assumption.

be given. We omit it here for space restrictions; it runs along the usual lines, defining a formula A of propositional logic of counterfactuals to have degree 1 in a model (formed of the bivalent set W of worlds, the parameters Σ and $\sim$, and an evaluation of propositional variables in each world) if its intension in the model is the whole set W of its possible worlds, and to be a *tautology* (written $\models A$) if it has degree 1 in all models. A detailed account of the semantics as well as a thorough discussion of all defined notions will be given in a full paper (in preparation).

8 Properties of fuzzy counterfactuals

Various properties of fuzzy counterfactuals can be derived in higher-order Łukasiewicz logic. We give just a sample of our results here, omitting the proofs (which will be given in a full paper under preparation).

First, the condition (26) ensures that the fuzzy set of minima of A in $\preccurlyeq_w$ is non-empty iff A is non-empty. Consequently, in any world w in any given model, $(A \mathrel{\Box\!\!\rightarrow} C)_w$ has the full degree 1 for all C if and only if $A = \emptyset$. In other words, a counterfactual is trivially true only if its antecedent is impossible.[10]

Second, the undesirable properties of counterfactuals mentioned in Section 2 are indeed refutable in our framework: fuzzy models can be constructed (by adapting Lewis' bivalent counterexamples) that invalidate the rules of weakening, contraposition, and transitivity for counterfactual implication. The counterexamples disprove not only their graded variants, but also the weaker forms with premises satisfied to the full degree:

$$\not\models \ \Delta(A \mathrel{\Box\!\!\rightarrow} C) \rightarrow ((A \,\&\, B) \mathrel{\Box\!\!\rightarrow} C)$$
$$\not\models \ \Delta(A \mathrel{\Box\!\!\rightarrow} C) \rightarrow (\neg C \mathrel{\Box\!\!\rightarrow} \neg A)$$
$$\not\models \ \Delta(A \mathrel{\Box\!\!\rightarrow} B) \,\&\, \Delta(B \mathrel{\Box\!\!\rightarrow} C) \rightarrow (A \mathrel{\Box\!\!\rightarrow} C)$$

[10] Our definition (39), though straightforwardly using the fuzzy *minimum* of A in the closeness relation, thus need not use the implausible Limit Assumption (cf. Lewis, 1973, p. 20) that there always exist the closest among A-worlds (for $A \neq \emptyset$). In the fuzzy setting, the rather natural non-triviality condition (26) on the similarity of degrees automatically entails that the *fuzzy* set of minima of a non-empty fuzzy set in the fuzzy ordering by closeness is non-empty (though the minimal worlds need not exist to the full degree).

Third, the desirable properties of counterfactuals are valid in our framework. For instance, it is provable in higher-order Łukasiewicz logic that the strength of the counterfactual conditional is intermediate between the material and strict conditionals (cf. (4) and (8) above in Section 2):

$$\models \ \Box(A \to C) \to (A \,\Box\!\!\to C)$$
$$\models \ (A \,\Box\!\!\to C) \to (A \to C),$$

if the (S5-)modality of necessity is defined in the natural way as the proposition

$$(\Box A) \ = \ \{w \mid (\forall v)(v \in A)\}.$$

Many further classical tautologies involving counterfactuals are provable in higher-order Łukasiewicz logic, for instance the properties (1) and (9)–(12) as well as the rule (6) of deduction within conditionals. Some classical tautologies, however, only hold for full degrees: for instance, it is provable that

$$\models \ \Delta(\Box\neg A) \leftrightarrow \Delta(A \,\Box\!\!\to \bot), \tag{40}$$

but counterexamples can be given to this formula without the Δ's (cf. (7)); similarly for Lewis' axiom (2).

9 Conclusions

We have shown that basic ideas of Lewis–Stalnaker semantics of counterfactuals can be reconstructed in higher-order Łukasiewicz logic (in fact, mutatis mutandis, in every t-norm fuzzy logic). The semantics has been shown to be adequate to the intuitive understanding of counterfactual conditionals; i.e., it validates plausible properties of counterfactuals and invalidates implausible ones (at least of those considered in this paper).

The new formalization of Lewis–Stalnaker semantics is based on an application of a well-developed general theory of fuzzy similarity relations to the particular case of the similarity of possible worlds. Fuzzy logic, which can be interpreted as logic of gradual predications, is particularly suitable for the purpose, since the relation of similarity is indisputably gradual (some things are more similar than others).

The ensuing fuzzy semantics moreover automatically accommodates counterfactuals that involve gradual predicates (i.e., such that come in degrees). It also admits the gradualness of counterfactuals themselves, i.e., the fact that some counterfactuals are perceived to be truer than others. The price paid for these advantages is the use of non-classical logic in the semantics; however, since t-norm logics are well-established (being not too different from intuitionistic or linear logic) and mathematics based on these logics is sufficiently developed (see Běhounek & Cintula, 2005; Běhounek et al., 2008), the cost is not too high and seems to be worth the gains.

The present paper only dealt with the fuzzy semantics of the logic of counterfactuals; its axiomatization as well as the study of its syntactic or semantic variants is left for future work.

References

Běhounek, L., Bodenhofer, U., & Cintula, P. (2008). Relations in Fuzzy Class Theory: Initial steps. *Fuzzy Sets and Systems*, *159*, 1729–1772.

Běhounek, L., & Cintula, P. (2005). Fuzzy class theory. *Fuzzy Sets and Systems*, *154*, 34–55.

Bodenhofer, U. (2000). A similarity-based generalization of fuzzy orderings preserving the classical axioms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *8*, 593–610.

Goodman, N. (1947). The problem of counterfactual conditionals. *The Journal of Philosophy*, *44*, 113–128.

Hájek, P. (1998). *Metamathematics of fuzzy logic.* Dordercht: Kluwer.

Hájek, P. (2010). Fuzzy logic. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* World Wide Web. (Fall 2010 edition.)

Hay, L. S. (1963). Axiomatization of the infinite-valued predicate calculus. *Journal of Symbolic Logic*, *28*, 77–86.

Lewis, D. (1973). *Counterfactuals.* Oxford: Blackwell.

Ono, H. (2003). Substructural logics and residuated lattices—an introduction. In V. F. Hendricks & J. Malinowski (Eds.), *50 years of Studia Logica* (pp. 193–228). Dordrecht: Kluwer.

Rasiowa, H. (1974). *An algebraic approach to non-classical logics.* Amsterdam: North-Holland.

Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Oxford: Oxford University Press.

Zadeh, L. A. (1971). Similarity relations and fuzzy orderings. *Information Sciences*, *3*, 177–200.

Libor Běhounek
Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic
behounek@cs.cas.cz
http://www.cs.cas.cz/behounek

Ondrej Majer
Institute of Philosophy
Academy of Sciences of the Czech Republic
Jilská 1, 110 00 Prague 1, Czech Republic
majer@site.cas.cz