I. Counterfactuals

I.I. Counterfactual conditionals

Our subject is *counterfactual conditionals*:

(1) If kangaroos had no tails, they would topple over.
(2) If Oswald hadn't shot Kennedy, someone else would have.

In general, conditionals of roughly the following form:

(C) If it had been that ϕ, it would have been that ψ.

How should logic treat such conditionals?

I.II. Counterfactuals in MPL? (LfP 8.1)

The language of MPL supplies two candidates for formalizing counterfactual conditionals:
- The material conditional: $\phi \to \psi$
- The strict conditional: $\Box(\phi \to \psi)$. We'll abbreviate this as $\phi \strictif \psi$

But these symbolizations face some well known difficulties.

I.II.1. False antecedent; true consequent

The material conditional faces what are sometimes called 'paradoxes of material implication'. As a first example, the material conditional 'obeys' the following inference patterns:

False antecedent/true consequent: $\dfrac{\sim\phi}{\phi \to \psi}$ $\qquad \dfrac{\psi}{\phi \to \psi}$

Remark. When we say that it 'obeys' the patterns we simply mean that the conclusion is a semantic consequence of the premiss in K: that is, $\sim\phi \vDash_K \phi \to \psi$ and $\psi \vDash_K \phi \to \psi$.[1]

But it is not obvious that 'if' in English sustains the analogous inferences:

False antecedent:

(P1) It's not snowing.
(C1) So[?], if it had snowed, Johnson would have ordered a nuclear strike on Brussels.

True consequent:

(P2) The lecture is on Monday.
(C2) So[?], if it had been rescheduled for Thursday, it would have been on Monday.

It seems that (P1)/(P2) may be true when (C1)/(C2) are false.

[1] This also holds good for PL. But since we're talking about the language of MPL we focus on semantic consequence relations suitable for this language.

I.II.2. Augmentation

Other prima facie problem cases also cause trouble for the strict conditional. Both $\rightarrow$ and $⥽$ obey augmentation:

$$\textbf{Augmentation:}\quad \frac{\phi \rightarrow \psi}{\phi \wedge \chi \rightarrow \psi} \qquad \frac{\phi ⥽ \psi}{\phi \wedge \chi ⥽ \psi}$$

Remark. This inference is also known as strengthening the antecedent.

But the following English inferences are questionable:

Augmentation:

(P3) If you had flicked the light switch, the light would be on

(C3) So[?], if you had flicked the switch, having removed the bulb, it would be on.

I.II.3. Contraposition

Similarly, both $\rightarrow$ and $⥽$ obey contraposition:

$$\textbf{Contraposition:}\quad \frac{\phi \rightarrow \psi}{\sim\psi \rightarrow \sim\phi} \qquad \frac{\phi ⥽ \psi}{\sim\psi ⥽ \sim\phi}$$

But contraposition seems questionable for the English conditional:

Contraposition:

(P4) [Speaker looks at light clouds above] If it rained, it wouldn't rain heavily.

(C4) So[?], if it rained heavily, it wouldn't rain.

I.III. Stalnaker's conditional, SC (LfP 8.3)

I.III.1. Syntax of SC

To avoid these problems, we might take a different approach (following Robert Stalnaker and David Lewis).

We enrich the language of MPL with a further binary connective: $\Box\!\!\rightarrow$.

Syntactically, $\Box\!\!\rightarrow$ functions just like $\rightarrow$.

Definition of wff (LfP 204):

- Every sentence letter α is a wff.
- If ϕ and ψ are wffs, then $(\phi \rightarrow \psi)$, $\sim\phi$ and $\Box\phi$ are wffs, and so is $(\phi \Box\!\!\rightarrow \psi)$.

I.III.2. Stalnaker's semantics: first pass

How do we evaluate conditionals? Lewis suggests the following:

> [(1)] seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over. (Counterfactuals (Blackwell), 1)

Stalnaker proposes similar truth-conditions for $A \Box\!\!\rightarrow B$:

> 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. (Indicative Conditionals, in Jackson (ed) *Conditionals*, 33–4)

More generally:[2]

Truth-conditions for counterfactuals (first pass): $\phi \Box\!\!\rightarrow \psi$ is true in w iff ψ is true in the closest ϕ-world to w (or there are no ϕ-worlds).

To capture 'the closest ϕ-world to w', we deploy an *ordering* of worlds.

[2] A ϕ-world is simply a world where ϕ is true.

I.III.3. Order-relations—a quick primer

First, a few preliminaries about order relations.

Definition of preorder: A binary relation $\preceq$ is a (non-strict) *preorder* of a set A iff:

- $\preceq$ is reflexive on A: $a \preceq a$ for each $a \in A$ and
- $\preceq$ is transitive: $a \preceq c$ for any a, b, c such that $a \preceq b$ and $b \preceq c$.

Notation. As usual, we write $a \preceq b$ as short for $\langle a, b\rangle \in \preceq$.[3]

Definition of a partial order: A (non-strict) preorder $\preceq$ of A is said to be a (non-strict) *partial order* of a set A iff, moreover:

- $\preceq$ is anti-symmetric: $a = b$ for any a and b such that $a \preceq b$ and $b \preceq a$.

Definition of linear order: A (non-strict) pre- or partial order $\preceq$ on A is said to be *linear* iff, moreover:

- $\preceq$ is connected on A: $a \preceq b$ or $b \preceq a$ or $a = b$, for any $a, b \in A$.

Terminology. Partial orders meeting this condition are called *linear orders.*

Worked Example. Categorize the following orders:

- $A \preceq_1 B$ iff $|A| \leqslant |B|$ (for $A, B \subseteq \mathbb{N}$)
- $A \preceq_2 B$ iff $A \subseteq B$ (for $A, B \subseteq \mathbb{N}$)
- $a \preceq_3 b$ iff $a \leqslant b$ (for $a, b \in \mathbb{N}$)

Definition of least element: When $\preceq$ is a partial order on A, and $B \subseteq A$, b_0 is said to be *the least element of* B wrt $\preceq$ (i.e. with respect to $\preceq$) iff:

- $b_0 \in B$ and every $b \in B$ is such that $b_0 \preceq b$.

Remark. the least element, if it exists, is unique.

Worked Example. Do the following sets have least elements?

- $\{2, 4, 6, \ldots\}$ wrt $\preceq_3$ (i.e. $\leqslant$)
- $\{A : A \subseteq \mathbb{N}\}$ wrt $\preceq_2$ (i.e. $\subseteq$)
- $\{A : A \subseteq \mathbb{N}, A \neq \varnothing\}$ wrt $\preceq_2$ (i.e. $\subseteq$)

[3] One might prefer $a \precsim b$ for preorders to dispel any hint of antisymmetry.

I.III.4. Stalnaker's semantics: second pass

Recall our first-pass truth-conditions for $\phi \mathrel{\Box\!\!\rightarrow} \psi$:

> **Truth-conditions for counterfactuals (first pass):** $\phi \mathrel{\Box\!\!\rightarrow} \psi$ is true in w iff ψ is true in the closest ϕ-world to w (or there are no ϕ-worlds).

We can spell out 'the closest ϕ-world to w' using a *comparative similarity relation* $\preceq_w$:

- Informally: $u \preceq_w v$ tells us that u is at least as similar to w as v is.

> **Definition of closest ϕ-world:** A world u is a *ϕ-world maximally close to w* iff:
> - u is a ϕ-world
> - $u \preceq_w v$ for any ϕ-world v

Remarks.

- Assuming $\preceq_w$ is a partial order, there is exactly one maximally close ϕ-world to w if there are any—consequently we may speak of 'the closest ϕ-world to w'.
- The closest ϕ-world to w is then the least element of the set of ϕ-worlds (wrt $\preceq_w$).

I.III.5. SC-models

An SC-model enriches an SMPL-model with a linear order $\preceq_w$ for each world:

> **Definition of an SC-model:** An SC-model is a triple $\langle \mathscr{W}, \preceq, \mathscr{I} \rangle$ where:
> - $\mathscr{W}$ is a non-empty set ("the set of possible worlds")
> - $\preceq$ is a three-place relation over $\mathscr{W}$, such that, for each $w \in \mathscr{W}$: ("nearness")
>   - $\preceq$ encodes a binary relation $\preceq_w$ that is a linear order on $\mathscr{W}$
>   - $w \preceq_w u$ for each $u \in \mathscr{W}$ ("base")
> - $\mathscr{I}$ is a two-place function such that, for each $w \in \mathscr{W}$: ("interpretation function")
>   - $\mathscr{I}(\alpha, w) = 0$ or 1 for each sentence letter α
>   - If some $v \in \mathscr{W}$ is a ϕ-world, then some $u \in \mathscr{W}$ is a ϕ-world maximally close to w. ("limit")

Remarks.

- $\mathscr{W}$ and $\mathscr{I}$ perform the same role as in SMPL-models. (SC-models omit $\mathscr{R}$.)
- $\preceq$ encodes $\preceq_w$ as follows: $u \preceq_w v$ iff $\langle u, v, w \rangle \in \preceq$.
- 'ϕ-world maximally close to w' is defined as above, where $u \in \mathscr{W}$ is a ϕ-world if $\mathrm{V}_{\mathscr{M}}(\phi, u) = 1$ (and $\mathrm{V}_{\mathscr{M}}$ is the SC-valuation, defined below).
- The condition of $\preceq_w$ being a linear order is equivalent to saying that $\preceq_w$ is strongly connected on $\mathscr{W}$, transitive and anti-symmetric (which is how Sider words the definition, LfP 205).
- Sider spells out the definition of 'closest ϕ-world' in the limit clause.
- Given base, if w is itself a ϕ-world, w is the closest ϕ-world to w.

I.III.6. SC-valuations

Definition of SC-valuation: Given an SC-model $\mathscr{M} = \langle \mathscr{W}, \preceq, \mathscr{I} \rangle$, the *valuation for* $\mathscr{M}$, $V_{\mathscr{M}}$, is the unique two place function that assigns 0 or 1 to each wff, for each $w \in \mathscr{W}$, that meets the following conditions, for any wffs ϕ and ψ:

- $V_{\mathscr{M}}(\alpha, w) = \mathscr{I}(\alpha, w)$, for each sentence letter α
- $V_{\mathscr{M}}(\phi \to \psi, w) = 1$ iff $V_{\mathscr{M}}(\phi, w) = 0$ or $V_{\mathscr{M}}(\psi, w) = 1$
- $V_{\mathscr{M}}(\sim\phi, w) = 1$ iff $V_{\mathscr{M}}(\phi, w) = 0$
- $V_{\mathscr{M}}(\Box\phi, w) = 1$ iff $V_{\mathscr{M}}(\phi, v) = 1$ for all $v \in \mathscr{W}$
- $V_{\mathscr{M}}(\phi \mathrel{\Box\!\!\to} \psi, w) = 1$ iff $V_{\mathscr{M}}(\psi, u) = 1$ for every ϕ-world u maximally close to w.

Remark. Sider spells out the definition of 'maximally close' in the semantic clause.

The clause for $\Box\!\!\to$ is equivalent to the following:

Truth-conditions for $\Box\!\!\to$ (final version for SC) $V_{\mathscr{M}}(\phi \mathrel{\Box\!\!\to} \psi, w) = 1$ iff ψ is true in the closest ϕ-world to w (or $V_{\mathscr{M}}(\phi, u) = 0$ for every $u \in \mathscr{W}$).

I.III.7. Validity

Validity is defined in the standard way:

Definition of SC-validity: a wff ϕ is *SC-valid*—$\vDash_{SC} \phi$—iff $V_{\mathscr{M}}(\phi, w) = 1$ for every SC-model $\mathscr{M} = \langle W, \preceq, \mathscr{I} \rangle$ and every world w in $\mathscr{W}$.

I.III.8. Semantic consequence

As is semantic consequence:

Definition of SC-semantic consequence: a wff ϕ is a *SC-semantic consequence* of a set of wffs Γ—$\Gamma \vDash_{SC} \phi$—iff $V_{\mathscr{M}}(\phi, w) = 1$ for every SC-model $\mathscr{M} = \langle W, \preceq, \mathscr{I} \rangle$ and every world w in $\mathscr{W}$ such that $V_{\mathscr{M}}(\gamma, w) = 1$ for each $\gamma \in \Gamma$.

I.III.9. Inferences with SC counterfactuals

The questionable inferences are *not* sustained by Stalnaker's conditional:

False antecedent: $\sim\phi \not\models_{\mathrm{SC}} \phi \mathrel{\Box\!\!\rightarrow} \psi$
True consequent: $\psi \not\models_{\mathrm{SC}} \phi \mathrel{\Box\!\!\rightarrow} \psi$
Contraposition: $\phi \mathrel{\Box\!\!\rightarrow} \psi \not\models_{\mathrm{SC}} \sim\psi \mathrel{\Box\!\!\rightarrow} \sim\phi$
Augmentation: $\phi \mathrel{\Box\!\!\rightarrow} \psi \not\models_{\mathrm{SC}} (\phi \wedge \chi) \mathrel{\Box\!\!\rightarrow} \psi$

See Exercise Sheet 8.

I.IV. Lewis's criticism of Stalnaker (LfP 8.7)

I.IV.1. Against anti-symmetry

Stalnaker's semantics take it for granted that the comparative similarity relation $u \preceq_w v$—'u is at least as similar to w as v is'—is a linear order on $\mathscr{W}$. How plausible is this?

- Anti-symmetry requires that there are no ties in similarity between distinct worlds: there are no distinct worlds u_1 and u_2 where u_1 is at least as similar to w as u_2 is, and vice versa.
- Lewis objects: why couldn't two worlds be equally similar to w?

I.IV.2. Against the limit assumption

Stalnaker also assumes that $\preceq_w$ conforms to limit. How plausible is this assumption?

- Limit requires that there be a unique ϕ-world most similar to w (if there are any ϕ-worlds)
- Lewis objects:
  - Why couldn't there be two or more equally close ϕ-worlds?
  - Why couldn't there be a sequence of ever closer ϕ worlds, with no closest?

I.V. Lewis's conditional (LfP 8.8)

I.V.1. LC-models

Given the problems they generate, Lewis's semantics drops the assumptions that $\preceq_w$ is anti-symmetric and the limit assumption, and modifies base.

Definition of an LC-model: An LC-model is a triple $\langle \mathscr{W}, \preceq, \mathscr{I} \rangle$ where:

- $\mathscr{W}$ is a non-empty set ("the set of possible worlds")
- $\preceq$ is a three-place relation over $\mathscr{W}$, such that, for each $w \in \mathscr{W}$: ("nearness")
  - The binary relation $\preceq_w$ is a linear preorder on $\mathscr{W}$
  - if $u \preceq_w w$, then $u = w$, for any $u \in \mathscr{W}$ ("base" [modified])
- $\mathscr{I}$ is a two-place function such that, for each $w \in \mathscr{W}$: ("interpretation function")
  - $\mathscr{I}(\alpha, w) = 0$ or 1 for each sentence letter α.

Remark. Modified base tells us, in effect, that any world distinct from w is not as similar to w as w is. This follows from the SC base assumption given antisymmetry.

I.V.2. LC-valuations

Having rejected limit, Lewis can not give his truth-conditions in terms of 'the closest ϕ-world' for there need not be any such world (even when there are ϕ-worlds).

Definition of LC-valuation: Given an LC-model $\mathscr{M} = \langle \mathscr{W}, \preceq, \mathscr{I} \rangle$, the *valuation for* $\mathscr{M}$, $\mathrm{LV}_{\mathscr{M}}$, is the unique two place function that assigns 0 or 1 to each wff, for each $w \in \mathscr{W}$, that meets the following conditions, for any wffs ϕ and ψ:

- $\mathrm{LV}_{\mathscr{M}}(\alpha, w) = \mathscr{I}(\alpha, w)$, for each sentence letter α
- $\mathrm{LV}_{\mathscr{M}}(\phi \to \psi, w) = 1$ iff $\mathrm{LV}_{\mathscr{M}}(\phi, w) = 0$ or $\mathrm{LV}_{\mathscr{M}}(\psi, w) = 1$
- $\mathrm{LV}_{\mathscr{M}}(\sim\phi, w) = 1$ iff $\mathrm{LV}_{\mathscr{M}}(\phi, w) = 0$
- $\mathrm{LV}_{\mathscr{M}}(\Box\phi, w) = 1$ iff $\mathrm{LV}_{\mathscr{M}}(\phi, v) = 1$ for all $v \in \mathscr{W}$
- $\mathrm{LV}_{\mathscr{M}}(\phi \mathrel{\Box\!\!\to} \psi, w) = 1$ iff there is some world u such that $\mathrm{LV}_{\mathscr{M}}(\phi, u) = 1$ and for every $v \preceq_w u$, $\mathrm{LV}_{\mathscr{M}}(\phi \to \psi, v) = 1$ (or there is no world u such that $\mathrm{LV}_{\mathscr{M}}(\phi, u) = 1$).

J. Further reading

- Burgess, *Philosophical Logic* (Princeton University Press, 2009)
- Fitting and Mendelsohn, *First-order Modal Logic* (Kluwer, 1998)