Counterfactuals

(1) If kangaroos had no tails, they would topple over.

(2) If Oswald hadn't shot Kennedy, someone else would have.

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

How should logic treat counterfactuals?

MPL has two salient formalizations:

- The material conditional: $\phi \to \psi$
- The strict conditional: $\Box(\phi \to \psi)$ (or $\phi \strictif \psi$)

The ‘paradoxes of material implication’

False antecedent/true consequent:

$$\frac{\sim\phi}{\phi \to \psi} \qquad \frac{\psi}{\phi \to \psi}$$

Augmentation: $\dfrac{\phi \to \psi}{\phi \wedge \chi \to \psi}$ $\qquad$ $\dfrac{\phi \strictif \psi}{\phi \wedge \chi \strictif \psi}$

$$\textbf{Contraposition:} \quad \frac{\phi \rightarrow \psi}{\sim\psi \rightarrow \sim\phi} \qquad \frac{\phi \mathrel{-\!\!3} \psi}{\sim\psi \mathrel{-\!\!3} \sim\phi}$$

Stalnaker's Conditional, SC

Definition of wff (LfP 204):

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

How do we evaluate counterfactuals?

Consider again:

(1) If kangaroos had no tails, they would topple over.

Lewis suggests:

> [(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*, 1)

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

> 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', 33–4)

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 *order* worlds.

Order-relations

Preorder on A: A binary relation $\preceq$ such that

- $\preceq$ 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 $b \preceq c$.

Partial order on A: $\preceq$ is a pre-order, and:

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

Linear order: $\preceq$ is a pre or partial-order, and:

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

Worked Example

Categorize the foxllowing 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 $\preceq$ iff:

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

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. $\subseteq$)

SC-semantics

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* $\preceq_w$:

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

SC-models

Definition of an SC-model: 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 linear order on $\mathscr{W}$: $\preceq_w$
  - $w \preceq_w u$ 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”)

SC-valuations

Definition of SC-valuation:

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

Truth-conditions for $\Box\!\!\rightarrow$ (final version for SC)

$\mathrm{V}_{\mathscr{M}}(\phi \mathrel{\Box\!\!\rightarrow} \psi, w) = 1$ iff ψ is true in the closest ϕ-world to w (or $\mathrm{V}_{\mathscr{M}}(\phi, u) = 0$ for every $u \in \mathscr{W}$).

Paradoxes of Material Implication: Revisited

False antecedent: $\sim\phi \not\models_{\text{SC}} \phi \mathrel{\Box\!\!\rightarrow} \psi$

True consequent: $\psi \not\models_{\text{SC}} \phi \mathrel{\Box\!\!\rightarrow} \psi$

Contraposition: $\phi \mathrel{\Box\!\!\rightarrow} \psi \not\models_{\text{SC}} \sim\psi \mathrel{\Box\!\!\rightarrow} \sim\phi$

Augmentation: $\phi \mathrel{\Box\!\!\rightarrow} \psi \not\models_{\text{SC}} (\phi \wedge \chi) \mathrel{\Box\!\!\rightarrow} \psi$

Stalnaker's semantics—Lewis's criticisms

Definition of an SC-model: 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 **linear order** on $\mathscr{W}$: $\preceq_w$
  - $w \preceq_w u$ 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**")

LC-models

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

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

LC-valuations

Definition of LC-valuation:

- The same clauses as SC-semantics for sentence letters, connectives and $\Box$.
- $\mathrm{LV}_{\mathscr{M}}(\phi \mathrel{\Box\!\!\rightarrow} \psi, w) = 1$ iff there is some world u such that $\mathrm{LV}_{\mathscr{M}}(\phi, u) = 1$ and for every $v \prec_w u$, $\mathrm{LV}_{\mathscr{M}}(\phi \rightarrow \psi, v) = 1$ (or there is no world u such that $\mathrm{LV}_{\mathscr{M}}(\phi, u) = 1$).

Further reading

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