I. Counterfactuals

I.I. Indicative vs Counterfactual (LfP 8.1)

The difference between 'indicative' and 'counterfactual' conditionals comes out in pairs like the following:

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

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

(1) is indicative, (2) is counterfactual.

Question. What is the difference between (1) and (2)?

Answers.

- Lewis considers assigning them different semantics: indicative conditionals are truth functional, counterfactuals are not.[1]
- Stalnaker: they have the same semantics; the differences between indicative and counterfactual conditionals are merely pragmatic.[2]

We set aside indicative conditionals and focus purely on the sort of semantics proposed for counterfactuals by Lewis and Stalnaker.

I.II. Inferences with counterfactuals (LfP 8.1)

The language of MPL supplies two candidates for formalizing counterfactual conditionals, 'if it had been that ϕ, it would have been that ψ'.

- 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 $\psi \vDash_K \phi \to \psi$.[3]

[1]See Probabilities of Conditionals and Conditional Probabilities in his *Philosophical Papers*, vol. 2

[2]Indicative Conditionals, *Philosophia 5* (1975).

[3]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.

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

False antecedent:

(P1) It's not raining.

(C1) So[?], if it had rained, May would have ordered a nuclear strike on Brussels.

True consequent:

(P2) The lecture is on Wednesday

(C2) So[?], if it had been rescheduled for Thursday, it would have been on Wednesday.

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

I.II.2. Augmentation

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

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

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 $\strictif$ obey contraposition:

Contraposition: $\dfrac{\phi \rightarrow \psi}{\sim\psi \rightarrow \sim\phi}$ $\qquad$ $\dfrac{\phi \strictif \psi}{\sim\psi \strictif \sim\phi}$

But contraposition seems questionable for the English conditional:

Augmentation:

(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

In light of the problems with attempting to symbolize counterfactuals in the language of MPL with the connectives we already have, let's introduce a further primitive connective.

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? Frank Ramsey proposes the following:

> If two people are arguing "If p will q?" and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q. (Ramsey, General Propositions and Causality, in his *Philosophical Papers* 155)

Similarly, Stalnaker suggests that to evaluate a conditional one should:

> First, add the antecedent (hypothetically) to your stock of beliefs; second make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is then true. (Stalnaker Indicative Conditionals, in Jackson (ed) *Conditionals*, 33)

His proposed truth-conditions are analogous:

> 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, ibid., 33–34)

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

To capture 'the closest ϕ-world', we order worlds according to their similarity with $w_@$.

I.III.3. Order-relations

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. We write $a \preceq b$ as short for $\langle a, b\rangle \in \preceq$.[4]

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

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.

Examples.

- $\{\langle a, b\rangle$: a is no taller than $b\}$ is a preorder on the set of persons—but not a partial order if there are ties in height between distinct people.
- $\{\langle A, B\rangle : |A| \leqslant |B|, A, B \subseteq \mathbb{N}\}$ is a preorder on the set of sets of natural numbers.
- $\{\langle A, B\rangle : A \subseteq B \subseteq \mathbb{N}\}$ is a partial order on the set of sets of natural numbers.
- $\{\langle a, b\rangle : a \leqslant b, a, b \in \mathbb{N}\}$ is a linear-order on the set of natural numbers.

Remark. Orders also come in strict varieties: a relation $\prec$ over A is a

- *strict partial order* if $\prec$ is irreflexive and transitive
- *strict linear order* if $\prec$ is irreflexive, transitive and connected on A

Terminology. Unless stated otherwise, (pre-, etc.) orders are henceforth non-strict

Definition of minimal element: When $\preceq$ is a preorder on A, and $B \subseteq A$, b_0 is said to be *a minimal element of* B (with respect to $\preceq$) iff:

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

Remark. When b is a minimal element of B with respect to a *partial* order, then b is unique and called the least element of B.

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

We order the worlds according to their comparative similarity with the actual world.

- We write $w \preceq_{w_@} w'$ to mean that w is at least as similar to the actual world as w' is. (We call $\preceq$ so interpreted, a 'comparative similarity relation')
- Then we can define 'the closest ϕ world' w_ϕ as the ϕ-world most similar to $w_@$—i.e., the least element of the set of ϕ-worlds with respect $\preceq_{w_@}$.

Truth conditions for counterfactuals (second pass): $\phi \mathrel{\Box\!\!\rightarrow} \psi$ is true (at the actual world) iff ψ is true in the closest ϕ-world w_ϕ—i.e w_ϕ is a ϕ-world such that for any ϕ-world u, $w_\phi \preceq_{w_@} u$—(or there is no ϕ-world).

I.III.5. SC-models

Stalnaker's semantics generalize our second pass in a natural way:

- To give conditions for $\phi \,\Box\!\!\rightarrow \psi$ to be true—i.e. true at $w_@$—we consider the closest ϕ-world—i.e. the ϕ-world most similar to $w_@$—where this is defined in terms of the comparative similarity relation for $w_@$, $\preceq_{w_@}$.
- To give conditions for $\phi \,\Box\!\!\rightarrow \psi$ to be true at arbitrary world w, we consider the ϕ-world most similar to w—where this is defined in terms of the comparative similarity relation for w: $u \preceq_w v$ is glossed as u is at least as similar to w as v is.

Trick: we can encode all of these binary relations, $\preceq_w$, as a single ternary relation $\preceq$.

> **Induced two place relation:** When R is a three-place relation, write R_z for the induced two-place relation obtained by fixing z as the third argument: $R_z xy$ iff $Rxyz$.

One last preliminary definition:

> **Definition of closest ϕ-world to w.** Given a similarity relation $\preceq_w$, a world $u \in \mathscr{W}$ is said to be a *ϕ-world maximally close to w* iff:
> - $V_{\mathscr{M}}(\phi, u) = 1$ and
> - $u \preceq_w v$ for any $v \in \mathscr{W}$ with $V_{\mathscr{M}}(\phi, v) = 1$

Remark. Given anti-symmetry, there is exactly one such world, if there are any—consequently we may speak of 'the closest ϕ-world to w' whenever there is at least one ϕ-world.

SC-models add a comparative similarity relation for each world to a SMPL-model.

> **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: ("nearness")
>   - The binary relation $\preceq_w$ is a linear order on $\mathscr{W}$, for each $w \in \mathscr{W}$
>   - $w \preceq_w u$ for each $u, w \in \mathscr{W}$ ("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 there is some $v \in \mathscr{W}$ is such that $V_{\mathscr{M}}(\phi, v) = 1$ [i.e. v is a ϕ-world], then there is $u \in \mathscr{W}$ such that u is a ϕ-world maximally close to w. ("limit")

Remarks.

- $\mathscr{W}$ and $\mathscr{I}$ perform the same role as in SMPL-models—note no $\mathscr{R}$.
- 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).
- Base intuitively tells us that w is at least as similar to itself as any other world is.
- Sider spells out the definition of 'closest ϕ-world' in the 'limit' clause.
- Given base, w is the closest ϕ-world when w is itself a ϕ-world.

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* $\mathscr{M}$, $\mathrm{V}_{\mathscr{M}}$, 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:

- $\mathrm{V}_{\mathscr{M}}(\alpha, w) = \mathscr{I}(\alpha, w)$, for each sentence letter α
- $\mathrm{V}_{\mathscr{M}}(\phi \to \psi, w) = 1$ 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 $v \in \mathscr{W}$
- $\mathrm{V}_{\mathscr{M}}(\phi \mathbin{\Box\!\!\rightarrow} \psi, w) = 1$ iff $\mathrm{V}_{\mathscr{M}}(\psi, u) = 1$ 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 $\Box\!\!\rightarrow$ is equivalent to the following:

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

I.III.7. Validity

Validity is defined in the standard way:

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

Worked Example. Show that $\phi \mathbin{-\!\!3} \psi \vDash_{\mathrm{SC}} \phi \mathbin{\Box\!\!\rightarrow} \psi$

I.III.9. Inferences with SC counterfactuals

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

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

See Exercise Sheet 8.

I.IV. Lewis's criticism of Stalnaker

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?

- Reflexivity on $\mathscr{W}$ requires that each world u is at least as similar to w as itself—plausible.
- Transitivity requires that when u_1 is at least as similar to w as u_2, and u_2 at least as similar to w as u_3, then u_1 is at least as similar to w as u_3—also plausible (so $\preceq_w$ is plausibly a preorder.)
- 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—this is less obvious. Lewis objects: why couldn't two worlds be equally similar to w?
- Linearity requires that any two worlds u_1 and u_2 are comparable with respect to their similarity to w: either u_1 is at least as similar to w as u_2 is, or vice versa—this is not uncontroversial, but Lewis points out that it's possible to make complex, multi-dimensional similarity comparisons (e.g. between cities).

I.IV.2. Against the limit-assumption

Stalnaker's also assumes that $\preceq_w$ conforms to base and limit. How plausible are these assumptions?

- Base requires that w is at least as similar to w as any world is—plausible.
- Limit requires that there be a unique ϕ-world most similar to w (if there are any ϕ-worlds)—less obvious. Lewis objects: why couldn't there be two or more equally close ϕ-worlds? Why couldn't there be an infinite sequence of ever-closer ϕ worlds, with no closest?

Remark. Does this criticism dubiously rely on taking informal glosses of $\preceq_w$, etc. seriously? These assumptions also generate validities that may be thought questionable—see Task E.

I.V. Lewis's conditional (LfP 8.4)

I.V.1. LC-models

Given the problems they generate, Lewis's semantics drop the assumptions that $\preceq_w$ is anti-symmetric and the limit assumption, and modify 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: ("nearness")
  - The binary relation $\preceq_w$ 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 $w \in \mathscr{W}$: ("interpretation function")
  - $\mathscr{I}(\alpha, w) = 0$ or 1 for each sentence letter α.

Remark. Base tells us 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 Mendehlsohn, *First-order Modal Logic* (Kluwer, 1998)