# Counterfactuals

Marco Degano

Philosophical Logic 2024
25 November 2024

# Readings

**Required:**

- Frank Veltman lecture notes on counterfactuals. `https://staff.fnwi.uva.nl/f.j.m.m.veltman/papers/Notes_Counterfactuals.pdf`
- Lecture notes: chapter 6

# Plan

1. Basic Data and Desiderata
2. Similarity Analyses
3. Open Issues

# Outline

# Conditionals

The **material** conditional: $p \supset q$
The material implication in classical logic $p \supset q$ iff $\neg p \vee q$

The **indicative** conditional: $p \rightarrow q$
*If it rains, the sky is gray*

The **counterfactual/subjunctive** conditional: $p \rightsquigarrow q$
*If it had rained, the sky would have been gray*

Hence counterfactuals will be sentences of the form

*If it had been the case that $\phi$, it would have been the case that $\psi$*

# The role of counterfactuals

- **Moral Philosophy:** moral reasoning and understanding responsibility (e.g., 'would the harm have occurred if the defendant had acted differently?')
- **Reasoning:** thinking about alternatives and learning from hypothetical scenarios
- **Causation:** causation as counterfactual dependence: event A causes B if, had A not occurred, B would not have occurred.
- **Metaphysics:** Relationship between the actual world and close vs remote possible worlds.

# Counterfactuals

First approximation: counterfactuals are conditional statements that discuss what would have been the case if the world were different or *contrary to fact*.

Note, however, that we can utter counterfactuals even if they are not contrary to fact. Can you think of some example?

(1) If the city had been built on lower ground, it would have experienced flooding. (And that's what we know from records.)

(2) If we were living in a simulation, we would see some anomalies.

(3) If he were to play the piano tomorrow, he would be overjoyed.

# Counterfactual and Material Implication

Counterfactuals $p \rightsquigarrow q$ differ with material implication $p \supset q$ in many respects.

# No triviality

$p \supset q$ trivially holds when the antecedent is false. But then all counterfactuals would be true.

(4) If the moon had been red, I would not exist.

# No truth-functionality

The antecedent of (5) and (6) is false, but (5) and (6) have distinct truth-conditions.

(5) If I had put the heating on, the room would have been warm.

(6) If I had put the heating on, the room would have exploded.

# No monotonicity / strengthening the antecedent

Counterfactuals are not monotonic. (8) does not follow from (7).

(7) a. If I had put sugar in my coffee, it would have tasted better.
b. $\phi \rightsquigarrow \psi$

(8) a. If I had put sugar and diesel oil in my coffee, it would have tasted better.
b. $(\phi \land \chi) \rightsquigarrow \psi$

# No contraposition

Contraposition does not seem to hold for counterfactuals:

Suppose there are two wolves: wolf A and wolf B. Yesterday, they attacked a sheep.

(9) a. If wolf A had not been around, then the sheep would have (still) been killed.
b. $\phi \rightsquigarrow \psi$

(10) a. If the sheep had not been killed, then wolf A would have been around.
b. $\neg\psi \rightsquigarrow \neg\phi$

# No transitivity

Examples where transitivity fails are discussed in the literature. (11-c) does not follow from (11-a-b).

(11) a. If I hadn't been born, Mike would have been my parent's oldest child.
$\phi \rightsquigarrow \psi$

b. If my parents had never met, I wouldn't have been born.
$\chi \rightsquigarrow \phi$

c. If my parents had never met, Mike would have been my parent's oldest child.
$\chi \rightsquigarrow \psi$

Notice: the order of the sentences seems important.

# Context-dependency

(12) If Amsterdam had been Rome, the weather would have been better.

‘Amsterdam had been Rome’ as ‘Amsterdam located where Rome is’

‘Amsterdam had been Rome’ as ‘Amsterdam (in the Netherlands) given the name Rome’

‘Amsterdam had been Rome’ as ‘Amsterdam being the capital of the Roman Empire’

…

# The First Lewis

C.I. Lewis (1912) [the first Lewis] proposed to analyze counterfactuals as strict conditionals.

(13) a. If I had put the heating on, the room would be warm.
b. $\Box(\phi \supset \psi)$
c. $M, w \models \Box(\phi \supset \psi)$ iff $\forall w' \in R(w, w') : M, w' \models \phi \supset \psi$

This does not meet the desiderata we just discussed.

# Strict Conditionals

Monotonicity holds for strict conditionals in any normal modal logic.

$\Box(\phi \supset \psi) \models \Box((\phi \wedge \chi) \supset \psi)$

Note: we are not claiming that monotonicity is always ruled out, as the example below seems fine.

(14) a. If I had put the heating on, the room would be warm.
b. If I had put the heating on, and I had been stayed inside, the room would be warm.

But we have already seen examples that violate monotonicity.

# Outline

# The Second Lewis and Stalnaker

Stalnaker (1968) and D. Lewis (1973) [the second Lewis] are the main proponents of the similarity analysis of counterfactuals.

*If it had been the case that $\phi$, it would have been the case that $\psi$*

A counterfactual is true in the actual world $w$ iff

for all possible worlds $w'$ where $\psi$ is true, we have that

1. $\phi$ is true, AND
2. $w'$ is **similar**/differs minimally from $w$.

# Language

We add to the language a binary operator $\rightsquigarrow$ for the counterfactual.

(We do not have explicit modal operators in the language, though they can be added.)

# Frames

We will interpret the languages in frames $\mathfrak{F} = \langle W, \prec \rangle$, where

- $W \neq \emptyset$ a non-empty set of worlds
- $\prec$ is a function which assigns to every $w \in W$ a strict partial ordering $\prec_w$ on some subset $W_w$ of $W$.

'$u \prec_w v$' means '$u$ is **more similar** to $w$ than $v$'.

# Similarity relation

$\prec_w$ is a strict partial order (transitive, irreflexive, asymmetric)[1]

The field $W_w$ of this relation $\prec_w$ is the set of worlds that are accessible from $w$.[2]

[1] We can also define the similarity relation based on $x \preceq_w y := x \prec_w y$ or $x = y$, yielding thus a partial order (transitive, reflexive, antisymmetric).

[2] $W_w = \{x \in W : \exists y \in W : x \prec_w y\} \cup \{y \in W : \exists x \in W : x \prec_w y\}$

# Models

A model is a triple $M = \langle W, \prec, V \rangle$, where $\langle W, \prec \rangle$ is a frame and $V$ is a function which assigns a truth value to every atomic sentence in every world.

$M, w \models \phi$ means that $\phi$ is true in the world $w$ (of the model $M$).

$[\![\phi]\!]_M$ denotes $\{w \in W \mid M, w = \phi\}$, the proposition expressed by $\phi$ (in $M$).

# Semantic Clauses (standard)

$$
\begin{array}{lll}
M, w \models p & \text{iff} & V(p, w) = 1 \\
M, w \models \neg\phi & \text{iff} & M, w \not\models \phi \\
M, w \models \phi \wedge \psi & \text{iff} & M, w \models \phi \text{ and } M, w \models \psi \\
M, w \models \phi \vee \psi & \text{iff} & M, w \models \phi \text{ or } M, w \models \psi \\
M, w \models \phi \supset \psi & \text{iff} & M, w \not\models \phi \text{ or } M, w \models \psi
\end{array}
$$

We use $\subset\supset$ as an abbreviation of the material biconditional.

$\Gamma \models \psi$ iff for any model $M$, world $w$, $M, w \models \gamma$ for every $\gamma \in \Gamma$, then $M, w \models \phi$

# Semantic Clauses ($\rightsquigarrow$)

$M, w \models \phi \rightsquigarrow \psi$ iff for every $u \in W_w \cap [\![\phi]\!]$

the following holds:

there is some $u' \in [\![\phi]\!]$ such that $u' \preceq_w u$ and $M, u'' \models \psi$ for every $u'' \in [\![\phi]\!]$ such that $u'' \preceq_w u'$.

In 'words': for all $\phi$-worlds $u$ which are accessible from $w$, we can find a $\phi$-world $u'$ such that $u'$ is more or equally similar to $w$ than $u$ and for all $\phi$-worlds $u''$ which are more or equally similar to $w$ than $u'$, $\psi$ must be supported in $u''$.

No matter how close to $w$ you go, you can never find a $\phi$ world where $\psi$ is false.

We can make more simple with an additional assumption.

# The Limit Assumption

Limit Assumption: For every $w \in W$, the relation $\prec_w$ is well-founded.

Two views:

The relation $\prec_w$ is well-founded iff every subset $V$ of $W_w$ has a minimal (closest) element (i.e, there is some $u \in V$ such that for no $v \in V, v \prec_w u$).

The relation $\prec_w$ is well-founded iff there is no infinitely descending chain in $W_w$ (i.e there is no sequence of worlds $u_1, \ldots, u_n, \ldots$ in $W_w$ such that for every $n, u_{n+1} \prec_w u_n$).

# Semantic clauses ($\rightsquigarrow$ + limit assumption)

Under the limit assumption, the new semantic clause for the counterfactual

$M, w \models (\phi \rightsquigarrow \psi)$ iff $M, u \models \psi$ for every closest $[\![\phi]\!]$-world $u$ to $w$.

Lewis does not accept the limit assumption. Is it reasonable to make such an assumption?

# Lewis example

*If the line had been longer than one inch, it would have been one hundred miles long*

What is the **closest** world which is more similar to the actual world where the line is longer?

For any $1 + k$, we can always find a closer world (e.g., $1 + k/2$)

# Two examples

Assuming the limit assumption, show that:

$$(p \rightsquigarrow q) \wedge (p \rightsquigarrow r) \models (p \wedge q) \rightsquigarrow r$$

$$\not\models (p \rightsquigarrow r) \supset ((p \wedge q) \rightsquigarrow r)$$

The latter exemplifies the failure of monotonicity.

# Axiomatization

The logic generated by the semantics sketched above is given by the following axioms and rules:

Taut: If $\phi$ has the form of a classical tautology, then $\vdash \phi$
CI: $\vdash \phi \rightsquigarrow \phi$
CC: $\vdash ((\phi \rightsquigarrow \psi) \wedge (\phi \rightsquigarrow \chi)) \rightarrow (\phi \rightsquigarrow (\psi \wedge \chi))$
CW: $\vdash (\phi \rightsquigarrow \psi) \rightarrow (\phi \rightsquigarrow (\psi \vee \chi))$
ASC: $\vdash ((\phi \rightsquigarrow \psi) \wedge (\phi \rightsquigarrow \chi)) \rightarrow ((\phi \wedge \psi) \rightsquigarrow \chi)$
AD: $\vdash ((\phi \rightsquigarrow \chi) \wedge (\psi \rightsquigarrow \chi)) \rightarrow ((\phi \vee \psi) \rightsquigarrow \chi)$
MP: $\phi \rightarrow \psi, \phi \vdash \psi$
REA: If $\vdash \phi \leftrightarrow \psi$, then $\phi \rightsquigarrow \chi \vdash \psi \rightsquigarrow \chi$

We call this system $\mathbf{P}$.

# Correspondence Theory

We impose further constraints on the similarity relation $\prec$.

And we look how certain formulas ‘correspond’ to, or characterize, specific structural properties of similarity frames $\langle W, \prec \rangle$.

# Weak Centering

Weak Centering: $w \in W_w$ for every $w \in W$, and for no $v \in W_w$ it holds that $v \prec_w w$.

This corresponds to:

Modus Ponens for $\rightsquigarrow$ $(\mathrm{MP}^{\rightsquigarrow})$ : $\phi \rightsquigarrow \psi, \phi \models \psi$

No world can be closer to a world $w$ than $w$ itself.

We are conflating the counterfactual conditional with the material conditional.

# Strong Centering

Strong Centering: $w \in W_w$ for every $w \in W$, and for every $v \in W_w$ such that $v \neq w, w \prec_w v$.

This corresponds to:

Conjunctive Sufficiency: $(\phi \wedge \psi) \rightarrow (\phi \rightsquigarrow \psi)$

**No world** different from $w$ can be **as close as** $w$ than $w$ itself.

# Uniqueness

Connectedness: for any $u, v \in W_w$, either $u = w$, or $u \prec_w v$, or $v \prec_w u$.

No two distinct worlds remain incomparable.

Under the limit assumption, we get the following correspondence.

*Uniqueness:* for any $u, v \in W_w$, either $u = w$, or $u \prec_w v$, or $v \prec_w u$.

$$(\phi \rightsquigarrow \psi) \vee (\phi \rightsquigarrow \neg\psi)$$

For any antecedent $\phi$, there is **at most one** $[\![\phi]\!]$ world closest to the actual world.

# Bizet and Verdi

(15) If Bizet and Verdi had been compatriots, Bizet would have been Italian.

(16) If Bizet and Verdi had been compatriots, Verdi would have been French.

Can we accept (17) without accepting (16) or (15)?

(17) If Bizet and Verdi had been compatriots, either Verdi would have been French or Bizet would have been Italian.

If yes, then uniqueness must be rejected.

# Almost-Connectedness

Almost-Connectedness: for any $u, v, w \in W_z$, if $u \prec_z w$, then either $u \prec_z v$ or $v \prec_z w$.

Note that this does not require that all worlds are comparable as in connectedness.

The corresponding axiom scheme is this:

Strengthening with a Possibility (ASP):

$$(\neg(\phi \rightsquigarrow \neg\psi) \wedge (\phi \rightsquigarrow \chi)) \rightarrow ((\phi \wedge \psi) \rightsquigarrow \chi)$$

Note: Read $\neg(\phi \rightsquigarrow \neg\psi)$ as 'If it had been the case that $\phi$, it might have been the case that $\psi$.'.

Hence, we can strengthen $\phi \rightsquigarrow \chi$ with $\psi$ as long as $\phi$ is compatible with $\psi$.

# Possible Counterexample to ASP

(18) a. It's not the case that if Verdi and Satie had been compatriots, Satie and Bizet would not have been compatriots.
b. If Verdi and Satie had been compatriots, Bizet would have been French.

(19) If both Verdi and Satie, and Satie and Bizet had been compatriots, Bizet would have been French.

# Outline

1. Basic Data and Desiderata

2. Similarity Analyses

3. Open Issues

# Famous Tichý example

*‘Consider a man, call him Jones, who is possessed of the following dispositions as regards wearing his hat. Bad weather invariably induces him to wear a hat. Fine weather, on the other hand, affects him neither way: on fine days he puts his hat on or leaves it on the peg, completely at random. Suppose moreover that actually the weather is bad, so Jones is wearing his hat.’*

(20) If the weather had been fine, Jones would have been wearing his hat.

We do not accept (20), but we should under a similarity analysis.

# System of weights

More in general, when can we say that a world is **more similar** to another one with respect to the actual world?

Lewis proposes a system of weights:

- Avoid big, widespread, diverse violations of law. ("big miracles")
- Maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
- Avoid even small, localized, simple violations of law. ("little miracles")
- **It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.**

The last principle can account for Tichý examples, taking 'wearing the hat' as 'a particular fact'.

# Particular Facts

However, consider the following:

I decline to bet on the toss of a coin. It is tossed anyway. It lands heads.

(21) If it had bet on heads, I would have won.

In this case, particular facts are kept constant, against Lewis system of weights.

Notice: if the betting might have caused to toss the coin differently, the counterfactual doesn't seem to hold.

# Disjunctive Antecedents

(22) a. If it had rained or snowed, the streets would have been wet.
$\phi \vee \psi \rightsquigarrow \chi$

b. If it had rained, the streets would have been wet.
$\phi \rightsquigarrow \chi$

c. If it had snowed, the streets would have been wet.
$\psi \rightsquigarrow \chi$

But consider the following often discussed example:

(23) a. If Spain had fought for the Axis or the Allies, she would have fought for the Axis.

b. If Spain had fought for the Allies, she would have fought for the Axis.

(23-a) does not entail (23-b). These examples are still discussed in the current literature.