Counterfactuals

Current Formal Models of Counterfactuals and Causation

Dean McHugh

Institute of Logic, Language and Computation
University of Amsterdam

NASSLLI 2025, Monday 23 June

Plan

Plan

Counterfactuals

(1) If the lights were off, the room would be dark.

(2) If I hadn't hit the billiard ball, it wouldn't have moved.

(3) If the cause had not occurred, the effect would not have occurred.

Counterfactuals

(1) If the lights were off, the room would be dark.

(2) If I hadn't hit the billiard ball, it wouldn't have moved.

(3) If the cause had not occurred, the effect would not have occurred.

Notation: $A > C$ and $A \mathrel{\Box\!\!\rightarrow} C$ are used to denote "if A were the case, C would be the case".

Plan

Counterfactual analyses of causation

we may define a cause to be an object, followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or in other words, where, if the first object had not been, the second never had existed.

Hume,
Enquiry Concerning Human Understanding
Part VII (1748)

Counterfactual analyses of causation

we may define a cause to be an object, followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or in other words, where, if the first object had not been, the second never had existed.

Hume,
Enquiry Concerning Human Understanding
Part VII (1748)

556 THE JOURNAL OF PHILOSOPHY

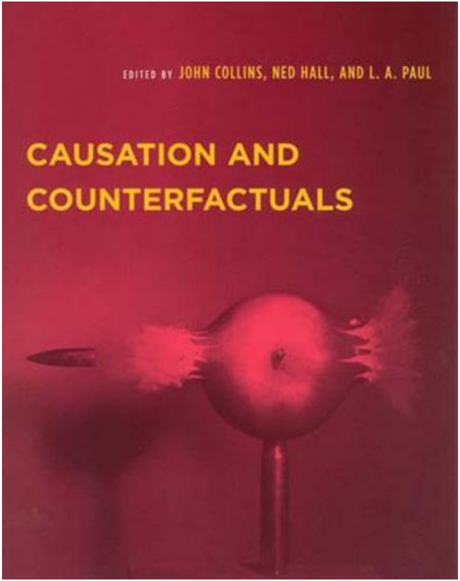
CAUSATION *

HUME defined causation twice over. He wrote "we may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*. Or, in other words, *where, if the first object had not been, the second never had existed*."[1]

Descendants of Hume's first definition still dominate the philosophy of causation: a causal succession is supposed to be a succession that instantiates a regularity. To be sure, there have been improvements. Nowadays we try to distinguish the regularities that count—the "causal laws"—from mere accidental regularities of succession. We subsume causes and effects under regularities by means of descriptions they satisfy, not by over-all similarity. And we allow a cause to be only one indispensable part, not the whole, of the total situation that is followed by the effect in accordance with a law. In present-day regularity analyses, a cause is defined (roughly) as any member of any minimal set of actual conditions that are jointly sufficient, given the laws, for the existence of the effect.

More precisely, let C be the proposition that c exists (or occurs) and let E be the proposition that e exists. Then c causes e, according to a typical regularity analysis,[2] iff (1) C and E are true; and (2) for some nonempty set $\mathfrak{L}$ of true law-propositions and some set $\mathfrak{F}$ of true propositions of particular fact, $\mathfrak{L}$ and $\mathfrak{F}$ jointly imply $C \supset E$, although $\mathfrak{L}$ and $\mathfrak{F}$ jointly do not imply E and $\mathfrak{F}$ alone does not imply $C \supset E$.[3]

David Lewis, 'Causation' (1973)



"But for the cause, the effect would not have occurred"

Cornell Law School

LII Legal Information Institute

About LII ▸ Get the law ▸ Lawyer directory

LII > Wex > **but-for test**

but-for test

The but-for test is a test commonly used in both tort law and criminal law to determine actual causation . The test asks, "but for the existence of X, would Y have occurred?"

Discrimination Law

DISCRIMINATION BECAUSE OF RACE, COLOR, RELIGION, SEX, OR NATIONAL ORIGIN

SEC. 703. (a) It shall be an unlawful employment practice for an employer—

(1) to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin; or

Civil Rights Act of 1964, Section 703(a)(1), p. 255

Discrimination Law

because of such individual's race, color, religion, sex, or national origin

Civil Rights Act of 1964, Section 703(a)(1), p. 255

What you do when you look to see whether there is [sex] discrimination under Title VII is, you say, “Would the same thing have happened to you if you were of a different sex?”

Justice Kagan during oral argument for
Bostock v. Clayton County (2020)

Plan

Plan

“Counterfactual” as contrary to fact?

When we prepare for a crucial experiment, we review the situation and consider what would happen if our hypothesis were true and what would happen if it were false. The subjunctive conditional is essential to the expression of these deliberations. In defending a hypothesis, I may employ a subjunctive conditional even though I believe the antecedent to be true; I may say, “If this were so, that would be so; but, as you see, this is so....”. It is said that detectives talk in this manner.

Chisholm (1946, p. 291)

“Counterfactual” as contrary to fact?

A patient is brought to the hospital in a coma. The doctor says, “I think he must have taken arsenic. He has these symptoms. And these are exactly the symptoms he would have if he had taken arsenic.”

(based on Anderson 1951)

“These are exactly the symptoms he would have **if he had taken arsenic**” does not imply that he didn’t take arsenic.

"These are exactly the symptoms he would have **if he had taken arsenic**" does not imply that he didn't take arsenic.

Other terminology

- Subjunctive conditional
- Past tense conditional (Ippolito 2013, Schulz 2014, Khoo 2015)
- *would*-conditional (Kratzer 1989)
- X-marked conditional (von Fintel and Iatridou 2023); same construction seen outside conditional antecedents:
  "I wish you were here." "I wish it would get warmer."

Plan

The indicative–subjunctive distinction

Adams' pair (Adams 1970)

What we know: someone shot Kennedy.

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

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

Observation: Indicatives, but not subjunctives, are required to be evaluated at worlds compatible with the current information (Stalnaker 1975, Veltman 1986)

Indicatives, but not subjunctives, are required to be evaluated at worlds compatible with the current information.

Consequences:

1. Indicatives, but not subjunctives, presuppose that their antecedent is compatible with our current information.

(6) #If I am wearing a hat, ...

(7) If I were wearing a hat, ...

Indicatives, but not subjunctives, are required to be evaluated at worlds compatible with the current information.

Consequences:

1. Indicatives, but not subjunctives, presuppose that their antecedent is compatible with our current information.

   (6) #If I am wearing a hat, ...

   (7) If I were wearing a hat, ...

2. Since indicatives impose a stronger requirement than subjunctives. Choosing a subjunctive signals that we have to give up some of our current information.
   Counterfactuality is an implicature (Stalnaker 1975, von Fintel 1997, Iatridou 2000, Ippolito 2013, Leahy 2011, 2018)

Plan

Plan

1. Introduction
   - Motivations
2. Preliminaries
   - Counterfactuality
   - Subjunctives
3. The truth conditions of counterfactuals
   - The material conditional
   - The strict conditional
   - Goodman's cotenability theory
   - Stalnaker's semantics
   - Lewis's Semantics
   - Kratzer's semantics
4. Challenges
   - Strengthening with a possibility
   - Reciprocity
   - Substitution

Material conditional

$A \supset C$ is true just in case A is false or C is true: $\neg A \vee C$

Some observations:

- Counterfactuals usually have false antecedents
- If counterfactuals expressed the material conditional, then all counterfactauls with a false antecedent would be...

Material conditional

$A \supset C$ is true just in case A is false or C is true: $\neg A \vee C$

Some observations:

- Counterfactuals usually have false antecedents
- If counterfactuals expressed the material conditional, then all counterfactauls with a false antecedent would be...
  trivially true

(8) If the lights were on, the room would be dark.

(9) If the lights were on, the moon would explode.

Plan

Strict conditional

Going back to C. I. Lewis *A Survey of Symbolic Logic* (1918)

$A \strictif C$ is true just in case necessarily, A is false or C is true: $\Box(A \supset C)$

Strict conditional

Going back to C. I. Lewis *A Survey of Symbolic Logic* (1918)

$A \strictif C$ is true just in case necessarily, A is false or C is true: $\Box(A \supset C)$

Kripke semantics: $\Box A$ is true at a world w just in case A is true at every world accessible from w.

Strict conditional

Going back to C. I. Lewis *A Survey of Symbolic Logic* (1918)

$A \strictif C$ is true just in case necessarily, A is false or C is true: $\Box(A \supset C)$

Kripke semantics: $\Box A$ is true at a world w just in case A is true at every world accessible from w.

Taking $A > C$ as $\Box(A \supset C)$ validates:

1. **Antecedent strengthening.** $A > C$ implies $(A \wedge B) > C$
2. **Contraposition.** $A > C$ implies $\neg C > \neg A$
3. **Transitivity.** $A > B$ and $B > C$ imply $A > C$

Counterexamples to Antecedent Strengthening

Antecedent strengthening. $A > C$ implies $(A \wedge B) > C$

Goodman's match (Goodman 1947)

(10) If I struck this match, it would light. $S > L$

(11) If I dipped this match in water and struck it, it would light. $(W \wedge S) > L$

Counterexamples to Antecedent Strengthening

From Lewis (1973a)

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

Counterexamples to Antecedent Strengthening

From Lewis (1973a)

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

(13) If kangaroos had no tails but used crutches, they would not topple over.

Counterexamples to Contraposition

Contraposition. $A > C$ implies $\neg C > \neg A$

Counterexamples to Contraposition

Contraposition. $A > C$ implies $\neg C > \neg A$

(14) a. If it rained, it didn't rain hard.
b. If it rained hard, it didn't rain.

(von Fintel and Heim 2011)

Counterexamples to Contraposition

Contraposition. $A > C$ implies $\neg C > \neg A$

(14) a. If it rained, it didn't rain hard.
b. If it rained hard, it didn't rain.

(von Fintel and Heim 2011)

(15) a. (Even) if Goethe hadn't died in 1832, he would still be dead now.
b. If Goethe were alive now, he would have died in 1832.

(due to Angelika Kratzer)

Counterexamples to Transitivity

Transitivity. $A > B$ and $B > C$ imply $A > C$

(16) a. If Hoover had been a Communist, he would have been a traitor.
b. If Hoover had been born in Russia, he would have been a Communist.
c. If Hoover had been born in Russia, he would have been a traitor.

(Stalnaker 1968, p. 38)

Plan

Goodman's cotenability theory (Goodman 1947)

> *"A is cotenable with D, ... if it is not the case that D would not be true if A were":* $\neg(A > \neg D)$

$A > B$ is true iff

(i) there is a finite set of true statements D cotenable with A such that A and D together with some laws imply B, and

(ii) there is no set of true statements H cotenable with A such that A and H together with some laws imply $\neg B$.

Goodman's cotenability theory (Goodman 1947)

"A is cotenable with D, ... if it is not the case that D would not be true if A were": $\neg(A > \neg D)$

$A > B$ is true iff

(i) there is a finite set of true statements D cotenable with A such that A and D together with some laws imply B, and

(ii) there is no set of true statements H cotenable with A such that A and H together with some laws imply $\neg B$.

"[T]o establish any counterfactual, it seems that we first have to determine the truth of another. If so, we can never explain a counterfactual except in terms of others, so that the problem of counterfactuals must remain unsolved. Though unwilling to accept this conclusion, I do not at present see any way of meeting the difficulty."

Plan

Stalnaker's semantics

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 1968, pp. 33–34)

Stalnaker's semantics

Let W be the set of possible worlds, R an accessibility relation, and λ the 'absurd world' where every proposition is true.

Let f be a conditional selection function, taking a sentence and a world and outputting a world, $f : \mathcal{L} \times W \to W$.

Stalnaker's semantics

Let W be the set of possible worlds, R an accessibility relation, and λ the 'absurd world' where every proposition is true.

Let f be a conditional selection function, taking a sentence and a world and outputting a world, $f : \mathcal{L} \times W \to W$.

Constraints on the selection function:

1. A is true at $f(A, w)$.
2. $f(A, w) = \lambda$ is the absurd world λ only if there is no possible world with respect to w in which A is true.
3. If A is true in w then $f(A, w) = w$.
4. If A is true in $f(B, w)$ and B is true in $f(A, w)$, then $f(A, w) = f(B, w)$.

These conditions on the selection function are necessary in order that this account be recognizable as an explication of the conditional

(Stalnaker 1968, p. 36)

Stalnaker's semantics

> *The informal truth conditions that were suggested above required that the world selected differ minimally from the actual world. [...] the selection is based on an ordering of possible worlds with respect to their resemblance to the based world.*

(Stalnaker 1968, pp. 35–36)

Plan

Lewis's sphere semantics (Lewis 1973b)

A model $M = (W, V, \mathcal{S})$ where

- W is a non-empty set of worlds
- V : $Atomics \to \mathcal{P}(W)$ is a valuation, telling us which atomic sentnces are true in which worlds
- $\mathcal{S}$: $W \to \mathcal{P}(\mathcal{P}(W))$ is a system of spheres, assigning to each world w a set of sets of worlds S_w

Lewis's sphere semantics (Lewis 1973b)

A model $M = (W, V, \mathcal{S})$ where

- W is a non-empty set of worlds
- V : $Atomics \to \mathcal{P}(W)$ is a valuation, telling us which atomic sentnces are true in which worlds
- $\mathcal{S}$: $W \to \mathcal{P}(\mathcal{P}(W))$ is a system of spheres, assigning to each world w a set of sets of worlds S_w

Constraints on the spheres:

1. Each $S \in S_w$ is nonempty
2. S_w is centered on w: $\{w\} \in S_w$
3. S_w is nested: for any $S, S' \in S_w$, $S \subseteq S'$ or $S' \subseteq S$

Lewis's sphere semantics (Lewis 1973b)

A model $M = (W, V, \mathcal{S})$ where

- W is a non-empty set of worlds
- V : $Atomics \to \mathcal{P}(W)$ is a valuation, telling us which atomic sentnces are true in which worlds
- $\mathcal{S}$: $W \to \mathcal{P}(\mathcal{P}(W))$ is a system of spheres, assigning to each world w a set of sets of worlds S_w

Constraints on the spheres:

1. Each $S \in S_w$ is nonempty
2. S_w is centered on w: $\{w\} \in S_w$
3. S_w is nested: for any $S, S' \in S_w$, $S \subseteq S'$ or $S' \subseteq S$

$A > C$ is true at w iff

(i) no sphere in S_w contains an A-world, or

(ii) some sphere $S \in S_w$ contains an A-world, and every A-world in S is a C-world

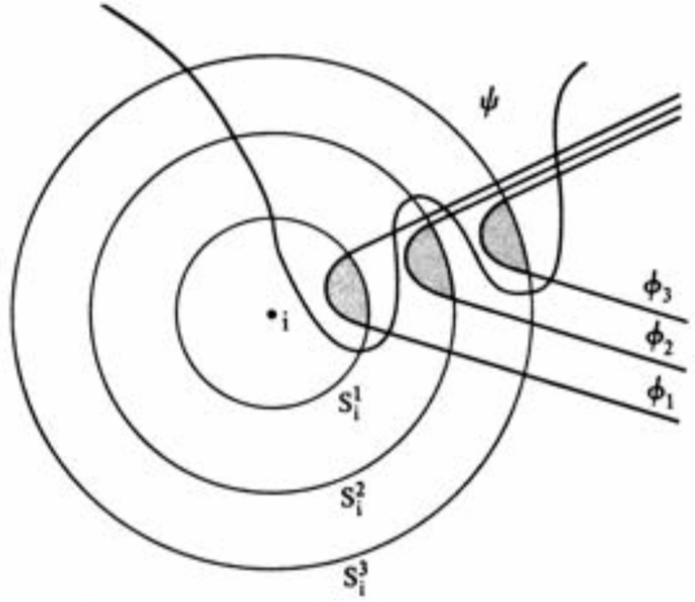


Figure: Lewis's sphere diagrams

Lewis's ordering semantics (Lewis 1981)

Let W be a set. For any $w \in W$ let $\leq_w$ be a reflexive and transitive binary relation over W.

For any sentences A and C and $w \in W$ define that a conditional $A > C$ is true at w (denoted $w \models A > C$) just in case

$$\forall x \models A \; \exists y \models A : y \leq_w x \land \forall z \leq_w y, z \models A \supset C,$$

where $A \supset C$ is the material conditional (that is, equivalent to $\neg A \lor C$).

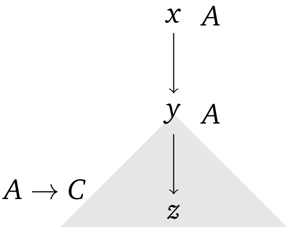
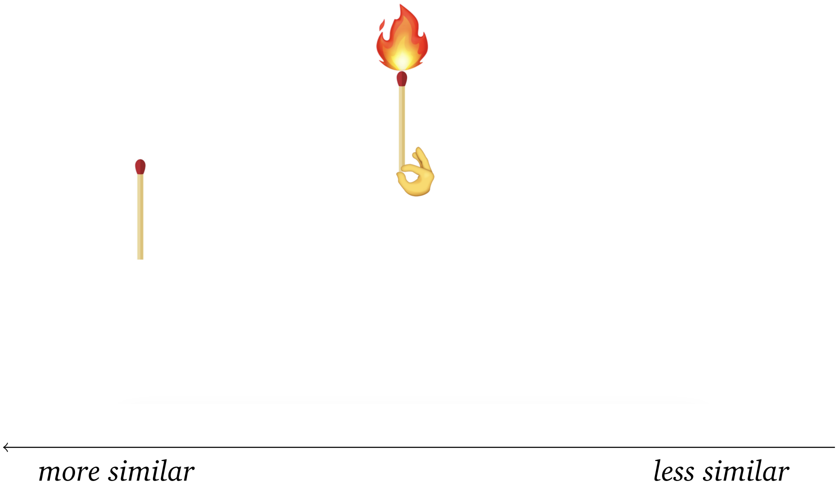


Figure: Illustrating the truth conditions of $A > C$.

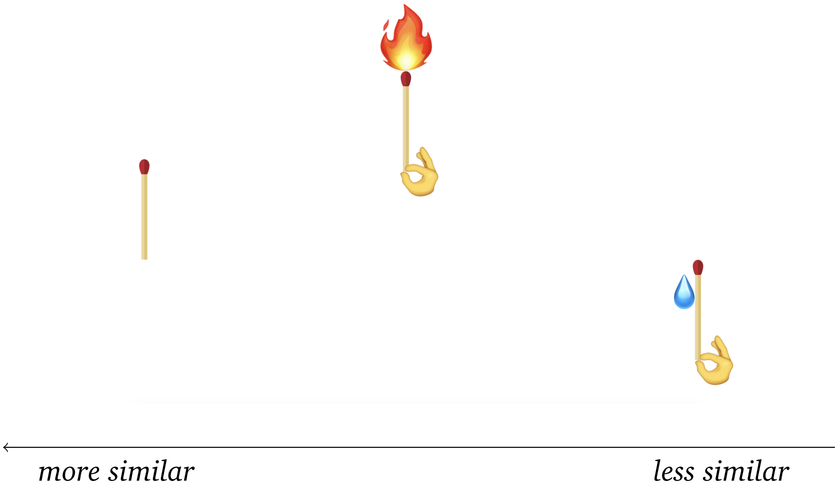
(17) a. If this match had been scratched, it would have lit.
b. If this match had been wet and scratched, it would have lit.

more similar *less similar*

(17) a. If this match had been scratched, it would have lit.
b. If this match had been wet and scratched, it would have lit.



(17) a. If this match had been scratched, it would have lit.
b. If this match had been wet and scratched, it would have lit.



Plan

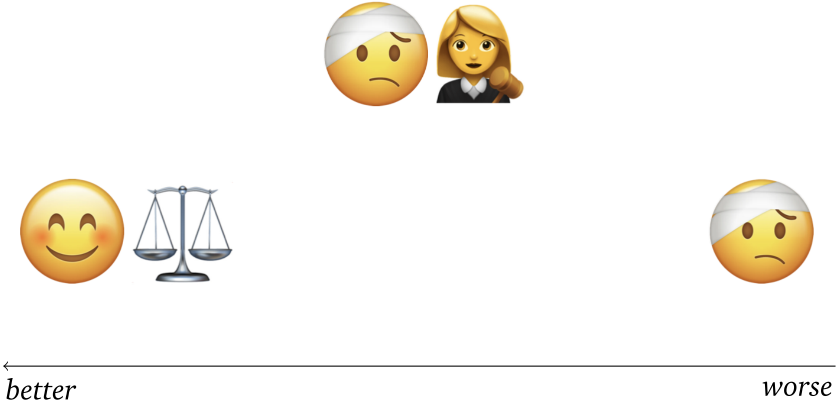
Deontic paradoxes (Kratzer 1981)

(18) a. Justice must be given to everyone.
b. If someone was unjustly treated, the injustice must be amended.
c. If someone was unjustly treated, the injustice must be rewarded.

Assumptions

1. $\Box A$ is true at w iff A is true at every v accessible from w (Kripke 1959)
2. v is accessible from w iff the rules of w are followed in v
3. *If A, must B* expresses $\Box(A \to B)$

Problem
$\Box A$ entails $\Box(\neg A \to B)$ and $\Box(\neg A \to \neg B)$ (Kratzer 1981, p. 70)


better
worse

Kratzer's semantics of *would*-conditionals

There is an intuitive and appealing way of thinking about the truthconditions for counterfactuals. It is an analysis that, in my heart of hearts, I have always believed to be correct...

A "would"-counterfactual is true in a world w iff every way of adding propositions that are true in w to the antecedent while preserving consistency reaches a point where the resulting set of propositions logically implies the consequent.

— Angelika Kratzer (2012, p. 127)

Kratzer's premise semantics (Kratzer 1981)

A proposition is a set of worlds. Let f and g be functions, each taking a world and returning a set of propositions.

- f is the *modal base*. It determines the accessibility relation.

  A world w' is accessible from a world w just in case w' makes true every proposition in the modal base at w.

$$wRw' \quad \equiv \quad w' \in \bigcap f(w)$$

Kratzer's premise semantics (Kratzer 1981)

A proposition is a set of worlds. Let f and g be functions, each taking a world and returning a set of propositions.

f is the *modal base*. It determines the accessibility relation.

A world w' is accessible from a world w just in case w' makes true every proposition in the modal base at w.

$$wRw' \quad \equiv \quad w' \in \bigcap f(w)$$

- For example, if every proposition in $f(w)$ is true at w, accessibility is reflexive: every world accesses itself.

Kratzer's premise semantics (Kratzer 1981)

A proposition is a set of worlds. Let f and g be functions, each taking a world and returning a set of propositions.

f is the *modal base*. It determines the accessibility relation.

A world w' is accessible from a world w just in case w' makes true every proposition in the modal base at w.

$$wRw' \quad \equiv \quad w' \in \bigcap f(w)$$

- For example, if every proposition in $f(w)$ is true at w, accessibility is reflexive: every world accesses itself.
- If $f(w)$ is empty, every world is accessible.

Kratzer's premise semantics (Kratzer 1981)

A proposition is a set of worlds. Let f and g be functions, each taking a world and returning a set of propositions.

f is the *modal base*. It determines the accessibility relation.

A world w' is accessible from a world w just in case w' makes true every proposition in the modal base at w.

$$wRw' \quad \equiv \quad w' \in \bigcap f(w)$$

- For example, if every proposition in $f(w)$ is true at w, accessibility is reflexive: every world accesses itself.
- If $f(w)$ is empty, every world is accessible.
- Kratzer (1981) proposes that counterfactuals have an empty modal base.

Kratzer's premise semantics (Kratzer 1981)

A proposition is a set of worlds. Let f and g be functions, each taking a world and returning a set of propositions.

f is the *modal base*. It determines the accessibility relation.

A world w' is accessible from a world w just in case w' makes true every proposition in the modal base at w.

$$wRw' \quad \equiv \quad w' \in \bigcap f(w)$$

- For example, if every proposition in $f(w)$ is true at w, accessibility is reflexive: every world accesses itself.
- If $f(w)$ is empty, every world is accessible.
- Kratzer (1981) proposes that counterfactuals have an empty modal base.
- For bouletic modals (*want*, *desire*), the modal base is doxastic: it only contains worlds compatible with the agent's beliefs at w.

Kratzer's premise semantics (Kratzer 1981)

g is the *ordering source*.

It determines an order over worlds.

One world w' is as good as another w'', relative to $g(w)$, just in case every proposition in $g(w)$ that is true at w'' is true at w'.

$$w' \leq_{g(w)} w'' \quad \equiv \quad \forall p \in g(w) : w'' \in p \rightarrow w' \in p$$

Kratzer's premise semantics (Kratzer 1981)

- g is realistic just in case w makes every proposition in $g(w)$ true: $w \in g(w)$

  For counterfactuals this corresponds to modus ponens:
  A and $A > C$ imply C

Kratzer's premise semantics (Kratzer 1981)

- g is realistic just in case w makes every proposition in $g(w)$ true: $w \in g(w)$

  For counterfactuals this corresponds to modus ponens: A and $A > C$ imply C

- g is totally realistic just in case only w makes every proposition in $g(w)$ true: $\bigcap g(w) = \{w\}$

  For counterfactuals, this corresponds to strong centering: A and C imply $A > C$

Kratzer's premise semantics (Kratzer 1981)

- g is realistic just in case w makes every proposition in $g(w)$ true: $w \in g(w)$

  For counterfactuals this corresponds to modus ponens: A and $A > C$ imply C

- g is totally realistic just in case only w makes every proposition in $g(w)$ true: $\bigcap g(w) = \{w\}$

  For counterfactuals, this corresponds to strong centering: A and C imply $A > C$

Kratzer's premise semantics (Kratzer 1981)

- g is realistic just in case w makes every proposition in $g(w)$ true: $w \in g(w)$

  For counterfactuals this corresponds to modus ponens: A and $A > C$ imply C

- g is totally realistic just in case only w makes every proposition in $g(w)$ true: $\bigcap g(w) = \{w\}$

  For counterfactuals, this corresponds to strong centering: A and C imply $A > C$

Kratzer (1981): counterfactuals have a totally realistic ordering source

Lewis (1975) on adverbs of quantification

always, usually, mostly, rarely, sometimes, never...

(19)
a. Sometimes, if a cat is happy, it purrs.
b. Never, if a cat is happy, does it purr.
c. Usually, if a cat is happy, does it purr.

Lewis (1975) on adverbs of quantification

always, usually, mostly, rarely, sometimes, never...

(19) a. Sometimes, if a cat is happy, it purrs.
b. Never, if a cat is happy, does it purr.
c. Usually, if a cat is happy, does it purr.

Lewis argues that these cannot be parsed as $Q(\textit{if } A, C)$, where Q is a unary operator

Lewis (1975) on adverbs of quantification

always, usually, mostly, rarely, sometimes, never...

(19) a. Sometimes, if a cat is happy, it purrs.
b. Never, if a cat is happy, does it purr.
c. Usually, if a cat is happy, does it purr.

Lewis argues that these cannot be parsed as $Q(\textit{if } A, C)$, where Q is a unary operator

Sometimes/never/usually is this conditional sentence true: "if a cat is happy, it purrs".

Lewis (1975) on adverbs of quantification

always, usually, mostly, rarely, sometimes, never...

(19) a. Sometimes, if a cat is happy, it purrs.
b. Never, if a cat is happy, does it purr.
c. Usually, if a cat is happy, does it purr.

Lewis argues that these cannot be parsed as $Q(\textit{if } A, C)$, where Q is a unary operator

Sometimes/never/usually is this conditional sentence true: “if a cat is happy, it purrs”.

But they *can* be parsed as $Q(A)(C)$, where Q is a binary operator

- $\textit{sometimes}(A)(C) = A \cap C \neq \emptyset$
- $\textit{never}(A)(C) = A \subseteq \neg C$
- $\textit{usually}(A)(C) = \#(A \wedge C)$ is greater than $\#(A \wedge \neg C)$

The restrictor view

> *The history of the conditional is the story of a syntactic mistake. There is no two-place if ... then connective in the logical forms for natural languages. If-clauses are devices for restricting the domains of operators. Whenever there is no explicit operator, we have to posit one.*

(Kratzer 1986, p. 656)

The restrictor view

> *The history of the conditional is the story of a syntactic mistake. There is no two-place if … then connective in the logical forms for natural languages. If-clauses are devices for restricting the domains of operators. Whenever there is no explicit operator, we have to posit one.*

(Kratzer 1986, p. 656)

Kratzer's implementation of the restrictor view

Conditional antecedents add the proposition they express to the modal base.

if A, C is true with respect to w, f, g just in case

C is true with respect to $w, f + A, g$,

where $f + A(w) = f(w) \cup \{|A|\}$ for any world w, and $|A|$ is the set of worlds where A is true

Lewis (1975) on adverbs of quantification

quantificational operators, including adverbs of quantification, probability operators, and other modal operators, do not operate over "conditional propositions." The persistent belief that there could be such "conditional propositions" is based on a simple syntactic mistake. If-clauses need to be parsed as adverbial modifiers that restrict operators that might be silent and a distance away. This is what we might call "the restrictor view" of if-clauses.

(Kratzer 2012, p. 107)

ANGELIKA KRATZER

AN INVESTIGATION OF THE LUMPS OF THOUGHT

Contents

Kratzer derives the ordering source from a set of propositions

$$w' \leq_{g(w)} w'' \quad \equiv \quad \forall p \in g(w) : w'' \in p \rightarrow w' \in p$$

Example

Illustration: p and q are true at w.
Consider: "if p were false, q would still be true"

$$g_1(w) = \{p, q\}$$
$$g_2(w) = \{p \cap q\}$$

Kratzer's semantics of *would*-conditionals

Definition (Admissible base set)

Let S be the set of situations (i.e. sets of parts of possible worlds). For any world w, an *admissible Base Set* is a subset F_w of $P(S)$ satisfying:

1. Truth: Every proposition in F_w is true at w: $w \in \bigcap F_w$.
2. Persistence: All proposition in F_w is persistent (for all $p \in F_w$ and situations s, s', if $s \in p$ and s is part of s', then $s' \in p$).
3. Cognitive Viability: All $p \in F_w$ are cognitively viable.
4. Non-Redundancy: F_w is not redundant (it does not contain propositions p and q such that $p \neq q$ and $p \cap W \subseteq q \cap W$).
5. Completeness: $\bigcap F_w$ contains all and only worlds that are indistinguishable from w, given the grain set by Cognitive Viability.

Define that proposition p *lumps* proposition q at world w just in case for any situation that is part of w in which p is true, q is true.

Kratzer's semantics of *would*-conditionals

Given an admissible base set and a proposition p, Kratzer defines the *crucial set* $F_{w,p}$ as follows.

Definition (The crucial set)

For any world w, admissible base set F_w, and proposition p, $F_{w,p}$ is the set of all subsets A of $F_w \cup \{p\}$ satisfying the following conditions:

1. A is consistent
2. $p \in A$
3. A is closed under lumping: for all $q \in A$ and $r \in F_w$: if q lumps r in w, then $r \in A$.

Definition (Truth conditions of "would"-counterfactuals)

Given a world w and an admissible Base Set F_w, a "would"-counterfactual with antecedent p and consequent q is true in w iff for every set in $F_{w,p}$ there is a superset in $F_{w,p}$ that logically implies q.

Angelika and the Zebra

Last year, a zebra escaped from the Hamburg zoo. The escape was made possible by a forgetful keeper who forgot to close the door of a compound containing zebras, giraffes, and gazelles. A zebra felt like escaping and took off. The other animals preferred to stay in captivity. Suppose now counterfactually that some other animal had escaped instead. Would it be another zebra? Not necessarily. I think it might have been a giraffe or a gazelle.

(Kratzer 1989, p. 625)

Angelika and the Zebra

Last year, a zebra escaped from the Hamburg zoo. The escape was made possible by a forgetful keeper who forgot to close the door of a compound containing zebras, giraffes, and gazelles. A zebra felt like escaping and took off. The other animals preferred to stay in captivity. Suppose now counterfactually that some other animal had escaped instead. Would it be another zebra? Not necessarily. I think it might have been a giraffe or a gazelle.

(Kratzer 1989, p. 625)

Angelika and the Zebra

> *Yet if the similarity theory of counterfactuals were correct, we would expect that, everything else being equal, similarity with the animal that actually escaped should play a role in evaluating this particular piece of counterfactual reasoning. Given that all animals in the compound under consideration had an equal chance of escaping, the most similar worlds to our world in which a different animal escaped are likely to be worlds in which another zebra escaped. That is, on the similarity approach, the counterfactual expressed by [(20)] should be false in our world.*
>
> *(20) If a different animal had escaped instead, it might have been a gazelle.*

(Kratzer 1989, p. 625)

How lumping helps the zebra case

We have to build the premise sets.

How lumping helps the zebra case

We have to build the premise sets.

- Let's call the zebra that escaped "John"

How lumping helps the zebra case

We have to build the premise sets.

- Let's call the zebra that escaped "John"
- In the actual world, "a zebra escaped" lumps "John escaped"

How lumping helps the zebra case

We have to build the premise sets.

- Let's call the zebra that escaped "John"
- In the actual world, "a zebra escaped" lumps "John escaped"
- Every situation in the actual world where "a zebra escaped" is true is one where "John escaped" is true

How lumping helps the zebra case

We have to build the premise sets.

- Let's call the zebra that escaped "John"
- In the actual world, "a zebra escaped" lumps "John escaped"
- Every situation in the actual world where "a zebra escaped" is true is one where "John escaped" is true
- We must add the antecedent "a different animal escaped instead"

How lumping helps the zebra case

We have to build the premise sets.

- Let's call the zebra that escaped "John"
- In the actual world, "a zebra escaped" lumps "John escaped"
- Every situation in the actual world where "a zebra escaped" is true is one where "John escaped" is true
- We must add the antecedent "a different animal escaped instead"
- If we added "a zebra escaped" we would have to add "John escaped", by closure under lumping

How lumping helps the zebra case

We have to build the premise sets.

- Let's call the zebra that escaped "John"
- In the actual world, "a zebra escaped" lumps "John escaped"
- Every situation in the actual world where "a zebra escaped" is true is one where "John escaped" is true
- We must add the antecedent "a different animal escaped instead"
- If we added "a zebra escaped" we would have to add "John escaped", by closure under lumping
- But then the set contains both "John escaped" and "if a different animal had escaped instead". But this is inconsistent!

How lumping helps the zebra case

We have to build the premise sets.

- Let's call the zebra that escaped "John"
- In the actual world, "a zebra escaped" lumps "John escaped"
- Every situation in the actual world where "a zebra escaped" is true is one where "John escaped" is true
- We must add the antecedent "a different animal escaped instead"
- If we added "a zebra escaped" we would have to add "John escaped", by closure under lumping
- But then the set contains both "John escaped" and "if a different animal had escaped instead". But this is inconsistent!
- $\therefore$ we cannot add "a zebra escaped" to the premise set

Plan

Plan

Strengthening with a Possibility (aka rational monotonicity):

$$\frac{A > C \qquad A \diamondsuit\!\!\rightarrow B}{(A \wedge B) > C}$$

This is valid on both Stalnaker's semantics and Lewis' (1973) sphere semantics for counterfactuals.

Counterexample to strengthening with a possibility

Ginsberg (1986, p. 50), Boylan and Schultheis (2017, 2021):

> *Alice, Billy, and Carol are playing a simple game of dice. Anyone who gets an odd number wins $10; anyone who gets even loses $10. The die rolls are, of course, independent. What Alice rolls has no effect on what Billy rolls and vice versa. Likewise for Alice and Carol as well as for Billy and Carol. Each player throws their dice. Alice gets odd; Billy gets even; Carol gets odd.*

(21)
a. If Alice and Billy had thrown the same type of number, then at least one person would still have won $10.
b. If Alice and Billy had thrown the same type of number, then Alice, Billy, and Carol could have all thrown the same type of number.
c. If Alice, Billy, and Carol had all thrown the same type of number, then at least one person would still have won $10.

Plan

Proof.

For any world w and sentence A, let $w \models A$ denote that A is true at w. Pick any world w and suppose $A > B$, $B > A$ and $B > C$ are true at w. To show that $A > C$ is true at w, pick any $x \models A$. We have to show that there is a $y \models A$ such that $y \leq_w x$ and for all $z \leq_w y$, $z \models A \to C$, where $\to$ is the material conditional.

Since $w \models A > B$ and $x \models A$, there is a $v \models A$ such that $v \leq_w x$ and (i) for all $v' \leq_w v$, $v' \models A \to B$. Since $\leq_w$ is reflexive, $v \leq_w v$, so $v \models A \to B$. Thus $v \models B$.

Since $w \models B > A$ and $v \models B$, there is a $u \models B$ such that $u \leq_w v$ and (ii) for all $u' \leq_w u$, $u' \models B \to A$. Since $w \models B > C$ and $u \models B$, there is a $y \models B$ such that $y \leq_w u$ and (iii) for all $z \leq_w y$, $z \models B \to C$. Since $y \leq_w u$, by (ii), $y \models B \to A$. Then as $y \models B$, $y \models A$. And as $y \leq_w u \leq_w v \leq_w x$, by transitivity of $\leq_w$, $y \leq_w x$.

We show that $z \models A \to C$ for all $z \leq_w y$. Pick any $z \leq_w y$. Then $z \leq_w y \leq_w u \leq_w v$, so by transitivity of $\leq_w$, $z \leq_w v$. Then by (i), $z \models A \to B$. And since $z \leq_w y$, by (iii), $z \models B \to C$. Hence $z \models A \to C$. □

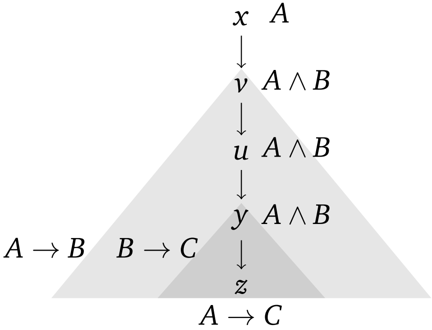


Figure: Illustrating the proof that Reciprocity is valid on Lewis's semantics.

Reciprocity (a.k.a. CSO)

$$\frac{A > B \qquad B > A \qquad B > C}{A > C}\ \text{Reciprocity}$$

Bacon's counterexample (Bacon 2013)

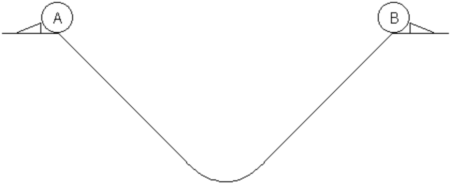


Figure: Bacon's counterexample to reciprocity.

(22)
a. If A fell, B would fall.
b. If B fell, A would fall.
c. If A fell, the light would turn green.
d. If B fell, the light would turn red.

Bacon's counterexample (Bacon 2013)

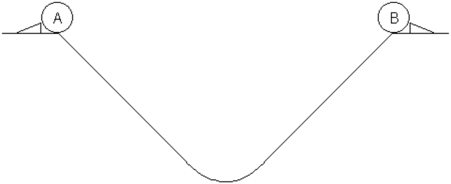


Figure: Bacon's counterexample to reciprocity.

(22)
a. If A fell, B would fall.
b. If B fell, A would fall.
c. If A fell, the light would turn green.
d. If B fell, the light would turn red.

(23)
a. If A had fallen, B would have fallen.
b. If B had fallen, A would have fallen.
c. If A had fallen, the light would have turned green.
d. If B had fallen, the light would have turned red.

Plan

1. Introduction
   - Motivations
2. Preliminaries
   - Counterfactuality
   - Subjunctives
3. The truth conditions of counterfactuals
   - The material conditional
   - The strict conditional
   - Goodman's cotenability theory
   - Stalnaker's semantics
   - Lewis's Semantics
   - Kratzer's semantics
4. Challenges
   - Strengthening with a possibility
   - Reciprocity
   - Substitution

Substitution

$$\frac{A \models B \qquad B \models A}{(A > C) \leftrightarrow (B > C)}\ \text{Substitution}$$

If A and B are logically equivalent, they are substitutable salva veritate in conditional antecedents.

Substitution

$$\frac{A \models B \qquad B \models A}{(A > C) \leftrightarrow (B > C)} \text{ Substitution}$$

If A and B are logically equivalent, they are substitutable salva veritate in conditional antecedents.

$$\frac{A \models B}{A > B} \text{ Entailment}$$

Given Entailment, Reciprocity implies Substitution.

$$\frac{\dfrac{A \models B}{A > B} \qquad \dfrac{B \models A}{B > A}}{(A > C) \leftrightarrow (B > C)} \text{ Reciprocity}$$

Infinitely many: a counterexample to Substitution

From Fine (2014, p. 328): There is one poison apple and infinitely many safe apples.

$\cdots \infty$

Infinitely many: a counterexample to Substitution

From Fine (2014, p. 328): There is one poison apple and infinitely many safe apples.

(24) If Eve ate infinitely many of the green apples,
she would be fine.

Infinitely many: a counterexample to Substitution

From Fine (2014, p. 328): There is one poison apple and infinitely many safe apples.

$\cdots \infty$

(24) If Eve ate infinitely many of the green apples,
she would be fine.

(25) If Eve ate infinitely many of the apples,
she would be fine.

Infinitely many: a counterexample to Substitution

From Fine (2014, p. 328): There is one poison apple and infinitely many safe apples.

 $\cdots \infty$

(24) If Eve ate infinitely many of the green apples,
she would be fine.

(25) If Eve ate infinitely many of the apples,
she would be fine.

The antecedents are logically equivalent.

Eve eats infinitely many of the green apples just in case she eats infinitely many of the apples.

Infinitely many: a counterexample to substitution

From Goodsell (2022):
There is a 1 Euro coin and infinitely many pennies, all facing tails.
The light is green just in case the 1 Euro coin is facing tails.

(26) If infinitely many of the coins were facing heads,
the light would still be green.

(27) If infinitely many of the pennies were facing heads,
the light would still be green.

References I

Adams, Ernest W. (1970). Subjunctive and indicative conditionals. *Foundations of Language* 6.1, pp. 89–94.

Anderson, Alan Ross (1951). A Note on Subjunctive and Counterfactual Conditionals. *Analysis* 12.2, pp. 35–38. DOI: 10.1093/analys/12.2.35.

Bacon, Andrew (2013). In defence of a naıve conditional epistemology. *Manuscript*.

Boylan, David and Ginger Schultheis (2017). Strengthening principles and counterfactual semantics. *Proceedings of the 21st Amsterdam Colloquium*. Ed. by Floris Roelofsen Alexandre Cremers Thom van Gessel, pp. 155–164. URL: https://semanticsarchive.net/Archive/jZiM2FhZ/AC2017-Proceedings.pdf.

References II

- Boylan, David and Ginger Schultheis (2021). How strong is a counterfactual? *The Journal of Philosophy* 118.7, pp. 373–404. DOI: 10.5840/jphil2021118728.
- Chisholm, R. M. (1946). The contrary-to-fact conditional. *Mind* 55.220, pp. 289–307.
- Fine, Kit (2014). Permission and possible worlds. *dialectica* 68 (3), pp. 317–336. DOI: 10.1111/1746-8361.12068.
- von Fintel, Kai (1997). The presupposition of subjunctive conditionals. *MIT working papers in linguistics* 25, pp. 29–44. URL: http://web.mit.edu/fintel/fintel-1998-subjunctive.pdf.
- von Fintel, Kai and Irene Heim (2011). Intensional semantics. *MIT Textbook*. URL: http://web.mit.edu/fintel/fintel-heim-intensional.pdf.
- von Fintel, Kai and Sabine Iatridou (2023). Prolegomena to a theory of X-marking. *Linguistics and Philosophy* 46.6, pp. 1467–1510.

References III

- Ginsberg, Matthew L. (1986). Counterfactuals. *Artificial Intelligence* 30, pp. 35–79. DOI: 10.1016/0004-3702(86)90067-6.
- Goodman, Nelson (1947). The Problem of Counterfactual Conditionals. *Journal of Philosophy* 44.5, pp. 113–128. DOI: 10.2307/2019988.
- Goodsell, Zachary (2022). Tossing Morgenbesser's Coin. *Analysis* 82.2, pp. 214–221.
- Iatridou, Sabine (2000). The Grammatical Ingredients of Counterfactuality. *Linguistic Inquiry* 31.2, pp. 231–270. DOI: 10.1162/002438900554352.
- Ippolito, Michela (2013). *Subjunctive Conditionals: A Linguistic Analysis*. MIT Press.
- Khoo, Justin (2015). On indicative and subjunctive conditionals. *Philosophers' Imprint* 15.32, pp. 1–40. URL: http://hdl.handle.net/2027/spo.3521354.0015.032.

References IV

- Kratzer, Angelika (1981). The notional category of modality. *Words, Worlds, and Contexts: New Approaches in Word Semantics*. Ed. by Hans-Jürgen Eikmeyer and Hannes Rieser. De Gruyter, pp. 39–74. DOI: 10.1515/9783110842524-004.
- — (1986). Conditionals. *Chicago Linguistics Society* 22.2, pp. 1–15.
- — (1989). An investigation of the lumps of thought. *Linguistics and philosophy*, pp. 607–653. DOI: 10.1007/BF00627775.
- — (2012). *Modals and conditionals*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199234684.001.0001.
- Kripke, Saul (1959). A Completeness Theorem in Modal Logic. *Journal of Symbolic Logic* 24.1, pp. 1–14. DOI: 10.2307/2964568. URL: https://projecteuclid.org:443/euclid.jsl/1183733464.
- Leahy, Brian (2011). Presuppositions and antipresuppositions in conditionals. *Semantics and linguistic theory*. Vol. 21, pp. 257–274. DOI: 10.3765/salt.v21i0.2613.

References V

- Leahy, Brian (2018). Counterfactual antecedent falsity and the epistemic sensitivity of counterfactuals. *Philosophical Studies* 175.1, pp. 45–69. DOI: 10.1007/s11098-017-0855-z.
- Lewis, Clarence Irving (1918). *A survey of symbolic logic*. Berkeley University of California Press.
- Lewis, David (1973a). Causation. *Journal of Philosophy* 70.17, pp. 556–567. DOI: 10.2307/2025310.
- — (1973b). *Counterfactuals*. Wiley-Blackwell.
- — (1975). Adverbs of quantification. *Formal Semantics of Natural Language*. Ed. by Edward Keenan. Cambridge University Press, pp. 3–15. DOI: 10.1017/CBO9780511897696.003.
- — (1981). Counterfactuals and Comparative Possibility. *Ifs: Conditionals, Belief, Decision, Chance and Time*. Ed. by William L. Harper, Robert Stalnaker, and Glenn Pearce. Springer, pp. 57–85. DOI: 10.1007/978-94-009-9117-0_3.

References VI

- Schulz, Katrin (2014). Fake tense in conditional sentences: A modal approach. *Natural language semantics* 22.2, pp. 117–144. DOI: 10.1007/s11050-013-9102-0.
- Stalnaker, Robert (1968). A theory of conditionals. *Ifs*. Ed. by William L. Harper, Robert Stalnaker, and Glenn Pearce. Springer, pp. 41–55. DOI: 10.1007/978-94-009-9117-0_2.
- — (1975). Indicative conditionals. *Philosophia* 5, pp. 269–286. DOI: 10.1007/BF02379021.
- Veltman, Frank (1986). Data Semantics and the Pragmatics of Indicative Conditionals. *On Conditionals*. Ed. by Elizabeth C. Traugott et al. Cambridge University Press.