Independence in counterfactuals:
Premise semantics, causality, and lumping

Stefan Kaufmann

Department of Linguistics
Northwestern University

Universität Bielefeld
(1) If that match had been scratched, it would have lighted.

“When we say (1), we mean that conditions are such—i.e. the match is well made, is dry enough, oxygen enough is present, etc.—that “The match lights” can be inferred from “The match is scratched.”
Goodman’s puzzle

(1) If that match had been scratched, it would have lighted.

“When we say (1), we mean that conditions are such—i.e. the match is well made, is dry enough, oxygen enough is present, etc.—that “The match lights” can be inferred from “The match is scratched.”

“Thus the connection we affirm may be regarded as joining the consequent with the conjunction of the antecedent and other statements that truly describe relevant conditions.”
Goodman’s puzzle

(1) If that match had been scratched, it would have lighted.

Q: What sentences are to be taken in conjunction with the antecedent as a basis for inferring the consequent?
Goodman’s puzzle

(1) If that match had been scratched, it would have lighted.

Q: What sentences are to be taken in conjunction with the antecedent as a basis for inferring the consequent?

A: True sentences with which the antecedent is cotenable.
Goodman’s puzzle

(1) If that match had been scratched, it would have lighted.

Q: What sentences are to be taken in conjunction with the antecedent as a basis for inferring the consequent?

A: True sentences with which the antecedent is cotenable.

A is cotenable with with S . . . if it is not the case that S would not be true if A were true.

i.e., if A were true, S would (still) be true
Two ingredients of a Goodmanian theory:
This talk

Two ingredients of a Goodmanian theory:

- Premise semantics for adding sentences to the antecedent
- Causal networks for choosing which sentences to add
Two ingredients of a Goodmanian theory:

- Premise semantics for adding sentences to the antecedent
- Causal networks for choosing which sentences to add

One way to combine the two.
Adding sentences to the antecedent
Models

Model $\mathcal{M} = \langle W, V \rangle$, where

- $W$ is a non-empty set of possible worlds;
- $V$ maps propositional variables to subsets of $W$
Models

- Model $\mathcal{M} = \langle W, V \rangle$, where
  - $W$ is a non-empty set of possible worlds;
  - $V$ maps propositional variables to subsets of $W$
- Interpretation function $\llbracket \cdot \rrbracket$ maps sentences to $\{0, 1\}$:
  
  $$\llbracket p \rrbracket_{w}^{\mathcal{M}} = 1 \iff w \in V(p)$$
Models

- Model $\mathcal{M} = \langle W, V \rangle$, where
  - $W$ is a non-empty set of possible worlds;
  - $V$ maps propositional variables to subsets of $W$
- Interpretation function $\llbracket \cdot \rrbracket$ maps sentences to $\{0, 1\}$:

$$\llbracket \rho \rrbracket^\mathcal{M}_w = 1 \iff w \in V(\rho)$$
$$\llbracket \varphi \land \psi \rrbracket^\mathcal{M}_w = 1 \iff \llbracket \varphi \rrbracket^\mathcal{M}_w = \llbracket \psi \rrbracket^\mathcal{M}_w = 1$$
$$\llbracket \neg \varphi \rrbracket^\mathcal{M}_w = 1 \iff \llbracket \varphi \rrbracket^\mathcal{M}_w = 0$$
Counterfactuals

- Premise set: set of propositions
  - representing one way of adding true sentences to the antecedent consistently
Counterfactuals

- Premise set: set of propositions
- Prem$_w$(φ): set of premise sets
  - representing all relevant ways of adding true$_w$ sentences to φ consistently
Counterfactuals

- Premise set: set of propositions
- $\text{Prem}_w(\varphi)$: set of premise sets
- *Would*-counterfactual: $\square \rightarrow$

$\varphi \square \rightarrow \psi$ is true at $w$ if and only if every set in $\text{Prem}_w(\varphi)$ has a superset in $\text{Prem}_w(\varphi)$ which entails $\psi$. 
Counterfactuals

- Premise set: set of propositions
- \( \text{Prem}_w(\varphi) \): set of premise sets

- **Would-counterfactual:** \( \Box \rightarrow \)
  \[ \varphi \Box \rightarrow \psi \text{ is true at } w \text{ if and only if every set in } \text{Prem}_w(\varphi) \text{ has a superset in } \text{Prem}_w(\varphi) \text{ which entails } \psi. \]

- **Might-counterfactual:** \( \Diamond \rightarrow \)
  \[ \varphi \Diamond \rightarrow \psi \text{ is true at } w \text{ if and only if there is a set in } \text{Prem}_w(\varphi) \text{ all of whose supersets in } \text{Prem}_w(\varphi) \text{ are consistent with } \psi. \]
Counterfactuals

- Premise set: set of propositions
- \( \text{Prem}_w(\varphi) \): set of premise sets
- **Would-counterfactual:** \( \square \rightarrow \)
  \[
  \varphi \square \rightarrow \psi \text{ is true at } w \text{ if and only if every set in } \text{Prem}_w(\varphi) \text{ has a superset in } \text{Prem}_w(\varphi) \text{ which entails } \psi.
  \]
- **Might-counterfactual:** \( \Diamond \rightarrow \)
  \[
  \varphi \Diamond \rightarrow \psi \text{ is true at } w \text{ if and only if there is a set in } \text{Prem}_w(\varphi) \text{ all of whose supersets in } \text{Prem}_w(\varphi) \text{ are consistent with } \psi.
  \]
- **Fact:** \( \varphi \square \rightarrow \psi \text{ iff } \neg(\varphi \Diamond \rightarrow \neg \psi) \)
Flavors of premise semantics

- The same definitions of $\square \rightarrow$ and $\Diamond \rightarrow$ give different results depending on the premise sets in $\text{Prem}_w(\varphi)$.
- The premise sets are the place to tweak the theory.
The same definitions of $\square \rightarrow$ and $\diamond \rightarrow$ give different results depending on the premise sets in $\text{Prem}_w(\varphi)$.

The premise sets are the place to tweak the theory.

- **Invariably:** For all $X$ in $\text{Prem}_w(\varphi)$,
  - $\varphi$ is in $X$;
  - $X$ is consistent;
  - all propositions in $X$ other than $\varphi$ are true at $w$. 
Flavors of premise semantics

- The same definitions of □→ and ◊→ give different results depending on the premise sets in $\text{Prem}_w(\varphi)$.
- The premise sets are the place to tweak the theory.
  - Invariably: For all $X$ in $\text{Prem}_w(\varphi)$,
    - $\varphi$ is in $X$;
    - $X$ is consistent;
    - all propositions in $X$ other than $\varphi$ are true at $w$.
  - But not all true propositions are relevant.
    - A function $f$ maps worlds to sets of propositions.
    - All $X$ in $\text{Prem}_w(\varphi)$ must be subsets of $f(w)$. 
Flavors of premise semantics

- The same definitions of $\square \rightarrow$ and $\Diamond \rightarrow$ give different results depending on the premise sets in $\text{Prem}_w(\varphi)$.

- The premise sets are the place to tweak the theory.
  - Invariably: For all $X$ in $\text{Prem}_w(\varphi)$,
    - $\varphi$ is in $X$;
    - $X$ is consistent;
    - all propositions in $X$ other than $\varphi$ are true at $w$.
  - But not all true propositions are *relevant*.
    - A function $f$ maps worlds to sets of propositions.
    - All $X$ in $\text{Prem}_w(\varphi)$ must be subsets of $f(w)$.

- Different versions, characterized by $f$:
  - $\square^n \rightarrow$, $\Diamond^n \rightarrow$: *Naïve* premise semantics
  - $\square^p \rightarrow$, $\Diamond^p \rightarrow$: *Partition* semantics
  - $\square^l \rightarrow$, $\Diamond^l \rightarrow$: *Lumping* semantics
Naïve premise semantics

First try: All true propositions are relevant.
Naïve premise semantics

First try: All true propositions are relevant.

- \( f(w) \): all propositions that are true at \( w \).

\[
f(w) = \{ p \subseteq W | w \in p \}
\]
Naïve premise semantics

First try: All true propositions are relevant.

- $f(w)$: all propositions that are true at $w$.
- $\text{Prem}_w^n(\varphi)$: all consistent subsets of $f(w) \cup \{\varphi\}$ containing $\varphi$.

$$\text{Prem}_w^n(\varphi) = \{ X \subseteq f(w) \mid \bigcap X \neq \emptyset \land \varphi \in X \}$$
Naïve premise semantics

First try: All true propositions are relevant.

- \( f(w) \): all propositions that are true at \( w \).
- \( \text{Prem}^n_w(\varphi) \): all consistent subsets of \( f(w) \cup \{\varphi\} \) containing \( \varphi \).

**Problem**: This can’t be right.

If \( \varphi \) is false at \( w \), then

- \( \varphi \nrightarrow^n \psi \) comes down to \( \Box(\varphi \rightarrow \psi) \).
  (strict implication)
- \( \varphi \nrightarrow^n \psi \) comes down to \( \Diamond(\varphi \land \psi) \)
  (logically consistency)
Naïve premise semantics

First try: All true propositions are relevant.

- \( f(w): \) all propositions that are true at \( w \).
- \( \text{Prem}^n_w(\varphi): \) all consistent subsets of \( f(w) \cup \{\varphi\} \) containing \( \varphi \).
- Problem: This can’t be right.
  
  If \( \varphi \) is false at \( w \), then
  
  - \( \varphi \nrightarrow \psi \) comes down to \( \square(\varphi \rightarrow \psi) \).
    (strict implication)
  
  - \( \varphi \nleftrightarrow \psi \) comes down to \( \Diamond(\varphi \land \psi) \)
    (logically consistency)
  
  More generally:
  
  \[
  \varphi \nrightarrow \psi \Leftrightarrow (\varphi \rightarrow \psi) \land (\neg \varphi \rightarrow \square(\varphi \rightarrow \psi))
  \]
  
  \[
  \varphi \nleftrightarrow \psi \Leftrightarrow (\varphi \land \psi) \lor (\neg \varphi \land \Diamond(\varphi \land \psi))
  \]
Partition semantics

Second try: Not *all* true propositions are relevant.

- Speakers have a more coarse-grained view of the facts.
Second try: Not *all* true propositions are relevant.

- Speakers have a more coarse-grained view of the facts.
- $f(w)$ subject only to the condition that it uniquely identify $\{w\}$.

$$\bigcap f(w) = \{w\}$$
Partition semantics

Second try: Not *all* true propositions are relevant.

- Speakers have a more coarse-grained view of the facts.
- $f(w)$ subject only to the condition that it uniquely identify $\{w\}$.
- $\text{Prem}^p_w(\varphi)$: all consistent subsets of $f(w) \cup \{\varphi\}$ containing $\varphi$ (as before).

$$\text{Prem}^p_w(\varphi) = \{X \subseteq f(w) | \bigcap X \neq \emptyset \land \varphi \in X\}$$
Second try: Not *all* true propositions are relevant.

- **Speakers** have a more coarse-grained view of the facts.
- $f(w)$ subject only to the condition that it uniquely identify $\{w\}$.
- $\text{Prem}_w^p(\varphi)$: all consistent subsets of $f(w) \cup \{\varphi\}$ containing $\varphi$ (as before).

**Questions:**
- Which propositions do/don’t belong in $\text{Prem}_w^p(\varphi)$?
- Which piece of the logical machinery regulates membership in $\text{Prem}_w^p(\varphi)$?
Partition semantics

Second try: Not all true propositions are relevant.

- Speakers have a more coarse-grained view of the facts.
- $f(w)$ subject only to the condition that it uniquely identify $\{w\}$.
- $\text{Prem}_w^p(\varphi)$: all consistent subsets of $f(w) \cup \{\varphi\}$ containing $\varphi$ (as before).
- Questions:
  - Which propositions do/don’t belong in $\text{Prem}_w^p(\varphi)$?
  - Which piece of the logical machinery regulates membership in $\text{Prem}_w^p(\varphi)$?
- Kratzer (1989): Closure conditions on sets in $\text{Prem}_w^p(\varphi)$
  - Closure under logical consequence and lumping.
  - Doesn’t quite work as expected; see Kanazawa, Kaufmann and Peters (2005).
Interim summary

Premise semantics:

- Closely related to Stalnaker/Lewis ordering semantics (Lewis, 1981)
Premise semantics:

- Closely related to Stalnaker/Lewis ordering semantics (Lewis, 1981)
- Dominant approach in linguistics
Premise semantics:

- Closely related to Stalnaker/Lewis ordering semantics (Lewis, 1981)
- Dominant approach in linguistics
- A framework, not a theory
- **Question:** How to define premise sets
Premise semantics:

- Closely related to Stalnaker/Lewis ordering semantics (Lewis, 1981)
- Dominant approach in linguistics
- A framework, not a theory
- Question: How to define premise sets

**Next section:** Some ideas from AI and psychology
Recall: ‘\( \varphi \square \rightarrow \psi \)’ is true iff \( \psi \) follows from \( \varphi \) together with true sentences which

- are cotenable with \( \varphi \)
- would (still) be true if \( \varphi \) were true

(Goodman)
Recall: ‘\( \varphi \implies \psi \)’ is true iff \( \psi \) follows from \( \varphi \) together with true sentences which are counterfactually independent of \( \varphi \) (Goodman)
Counterfactual independence

Recall: ‘\( \varphi \square \rightarrow \psi \)’ is true iff \( \psi \) follows from \( \varphi \) together with true sentences which

\[ \text{are counterfactually independent of } \varphi \]

(Hume) . . . [w]e may define a cause to be

\[ \text{an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second.} \]

\[ \text{Or in other words where, if the first object had not been, the second never had existed.} \]
Counterfactual independence

Recall: ‘$\varphi \quad \rightarrow \quad \psi$’ is true iff $\psi$ follows from $\varphi$ together with true sentences which

- are counterfactually independent of $\varphi$ 
- are causally independent of $\varphi$ 

(Goodman) 
(Hume)
Counterfactual independence

Recall: ‘\( \varphi \Box \rightarrow \psi \)’ is true iff \( \psi \) follows from \( \varphi \) together with true sentences which

- are counterfactually independent of \( \varphi \) (Goodman)
- are causally independent of \( \varphi \) (Hume)

Lewis: Counterfactual analysis of causality

Counterfactuals interpreted in terms of *overall comparative similarity* between possible worlds

Counterfactuals provide *evidence* about causal relations

(see also Collins, Hall and Paul, 2004)
Recall: ‘$\varphi \square \rightarrow \psi$’ is true iff $\psi$ follows from $\varphi$ together with true sentences which

- are counterfactually independent of $\varphi$ (Goodman)
- are causally independent of $\varphi$ (Hume)

Lewis: Counterfactual analysis of causality

Why not take causality as basic?
- (not that all counterfactuals assert causal relationships)
Causal independence

Pearl (2000):

In the last decade, owing partly to advances in graphical models, causality has undergone a major transformation: from a concept shrouded in mystery into a mathematical object with well-defined semantics and well-founded logic . . . Put simply, causality has been mathematized.
Bayesian Network:

- Directed Acyclic Graph (DAG) $\langle U, E \rangle$
  - $U$: set of random variables
  - $E$: relation over $U$ whose transitive closure is asymmetric
Bayesian networks

Bayesian Network:

- Directed Acyclic Graph (DAG) $\langle U, E \rangle$
- Probability distribution over the values assignments

Notation: $P(x_1, \ldots, x_n)$ for $P(X_1 = x_1, \ldots, X_n = x_n)$

[Diagram of a Bayesian network with nodes for Summer?, Rain?, Wet?, Slippery?, and Sprinkler (on/off).]
Markov Assumption: The probability of a variable is completely determined by the value(s) of its parent(s) in the graph.

\[ P(x_5|x_1, x_2, x_3, x_4) = P(x_5|x_4) \]
Markov Assumption: Decomposability

\[
P(x_1, x_2, x_3, x_4, x_5) \\
= P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)P(x_4|x_1, x_2, x_3)P(x_5|x_1, x_2, x_3, x_4) \\
= P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_4)
\]
Bayesian networks

Practical advantages:
- compact representation
- learning from limited data
- efficient inference
Observing that the sprinkler is on:

- Set $X_3$ to ‘on’
- Re-calibrate the probabilities of all other variables.
- Affects the probabilities of the seasons
Causal Bayesian Network:

- Bayesian Network under a special interpretation
- All arrows indicate causal influence
Causal Bayesian Network:

- Bayesian Network under a special interpretation
- All arrows indicate causal influence
- Two modes of inference: Observation and Intervention

Diagram:

- $X_1$: Summer?
- $X_2$: Rain?
- $X_3$: Sprinkler (on/off)
- $X_4$: Wet?
- $X_5$: Slippery?
Turning the sprinkler on

Intervention I: Manipulation

- Manipulate the network structure: Cut all arrows into $X_3$
Intervention I: Manipulation

- Manipulate the network structure: Cut all arrows into $X_3$
- Update as before (now on the modified network)
- Only affects the descendants of $X_3$
  (provided that $X_4, X_5$ are not observed)
Turning the sprinkler on

**Intervention II:** Intervention variable

- A special variable with values
  \{idle, do(X_3 = on), do(X_3 = off)\}

![Diagram of causal relationships involving sprinkler, summer, rain, wetness, and slipperiness.]
**Intervention II: Intervention variable**

- A special variable with values
  - \(idle, do(X_3 = on), do(X_3 = off)\)
  - \(idle\): The value of \(X_3\) is observed
Intervention II: Intervention variable

- A special variable with values 
  \{idle, do(X_3 = on), do(X_3 = off)\}
  - \textit{idle}: The value of \(X_3\) is \textit{observed}
  - \textit{do}(X_3 = \ldots): The value of \(X_3\) is \textit{manipulated}
Turning the sprinkler on

Intervention (either way):

- Prevents backtracking (abductive) inferences
- Similar to Lewisian “miracles” *(Lewis 1973, 1979)*
- Simple rule: All *non-descendants* of the manipulated variable remain unaffected
Observation vs. intervention

Two ways of asking ‘What if $X_i = x_i$?’
Observation vs. intervention

Two ways of asking ‘What if $X_i = x_i$?’

- **Observation**: Conditioning on ‘$X_i = x_i$’
  [Non-descendants of $X_i$ affected]

- **Intervention**: Conditioning on ‘$do(X_i = x_i)$’
  [Non-descendants of $X_i$ not affected]
Two ways of asking ‘What if $X_i = x_i$?’

- **Observation**: Conditioning on ‘$X_i = x_i$’
  [Non-descendants of $X_i$ affected]

- **Intervention**: Conditioning on ‘$do(X_i = x_i)$’
  [Non-descendants of $X_i$ not affected]

**Two hypotheses:**

**H1**: Indicatives involve observation.

**H2**: Counterfactuals involve intervention.
Observation vs. intervention

Two ways of asking ‘What if $X_i = x_i$?’

- Observation: Conditioning on ‘$X_i = x_i$’
  [Non-descendants of $X_i$ affected]
- Intervention: Conditioning on ‘$do(X_i = x_i)$’
  [Non-descendants of $X_i$ not affected]

Two hypotheses:

- **H1:** Indicatives involve observation.
- **H2:** Counterfactuals involve intervention.

Two problems:

- Some indicative conditionals involve intervention. (Kaufmann 2004, 2005b, 2006)
- Not all counterfactual inference involves intervention.
(abstract) **causal** condition:

When A happens, it causes B most of the time.
When B happens, it causes C most of the time.
A happened.
C happened.
Intervention: “Someone intervened directly on B, preventing it from happening. What is the probability that A/C would have happened?”

Observation: “What is the probability that A/C would have happened if we observed that B did not happen?”

Unspecified: “What is the probability that A/C would have happened if B had not happened?”
Intervention: “Someone intervened directly on B, preventing it from happening. What is the probability that A/C would have happened?”

A: 3.9  C: 2.3

Observation: “What is the probability that A/C would have happened if we observed that B did not happen?”

A: 2.7  C: 2.3

Unspecified: “What is the probability that A/C would have happened if B had not happened?”

A: 3.2  C: 2.4

(Scale: 1=very low, 2=low, 3=medium, 4=high, 5=high)
All rocket ships have two components, A and B. Component A causes Component B to operate. In other words, if A, then B.
Sloman and Lagnado (2005), Exp. 5, 6

- **Counterfactual**: “Suppose Component B/A were not operating, would Component A/B still operate?”

- **Explicit prevention**: “Suppose Component B/A were prevented from operating, would Component A/B still operate?”

- **Explicit prevention**: “Suppose Component B/A were prevented from moving, would Component A/B still be moving?”

- **Explicit observation**: “Suppose Component B/A were observed to not be moving, would Component A/B still be moving?”
Sloman and Lagnado (2005), Exp. 5,6

- Counterfactual: “Suppose Component B/A were not operating, would Component A/B still operate?”
  
  if not B, A: 68  
  if not A, B: 2.6

- Explicit prevention: “Suppose Component B/A were prevented from operating, would Component A/B still operate?”
  
  if not B, A: 89  
  if not A, B: 5.3

- Explicit prevention: “Suppose Component B/A were prevented from moving, would Component A/B still be moving?”
  
  if not B, A: 85  
  if not A, B: 19

- Explicit observation: “Suppose Component B/A were observed to not be moving, would Component A/B still be moving?”
  
  if not B, A: 22  
  if not A, B: 30
No simple relationship between counterfactuals and intervention.

“Representing intervention is not always as easy as forcing a variable to some value and cutting the variable off from its causes. Indeed, most of the data reported here show some variability in people’s responses. People are not generally satisfied to simply implement a \texttt{do} operation. People often want to know precisely how an intervention is taking place.”
Causal networks:

- empirically testable
- mathematically elegant
- computationally tractable
- precise statement and testing of hypotheses about causal inference
Interim Summary

Causal networks:
- empirically testable
- mathematically elegant
- computationally tractable
- precise statement and testing of hypotheses about causal inference

**Question:** What about a “causal premise semantics”? And how is all this related to possible, worlds, anyway?
Causal networks and possible worlds

**Networks**
- Event

**Worlds**
- Proposition

\[ X = x \]
Causal networks and possible worlds

**Networks**
- Event
- Variable

**Worlds**
- Proposition
- Partition
Causal networks and possible worlds

**Networks**
- Event
- Variable
- Network of variables

**Worlds**
- Proposition
- Partition
- Network of partitions

\[ \begin{array}{|c|c|}
\hline
x & \bar{x} \\
\hline
\end{array} \]

\[ \begin{array}{|c|c|}
\hline
x y & \bar{x} y \\
\hline
x \bar{y} & \bar{x} \bar{y} \\
\hline
\end{array} \]

X cuts across Y, but not vice versa.
Counterfactual alternatives

Suppose $x$ and $y$ are both true at world $w$. 
Only $x$ and $x \cdot y$ are relevant for the truth of counterfactuals.

- $y \notin f(w)$
- $f(w) = \{x, xy\}$
If \( y \) were false, \( x \) would still be true.
Counterfactual alternatives

If \( x \) were false, \( y \) might also be false.
Downstream inference

Sprinkler (on/off)  Rain (yes/no)

Wet (yes/no)

Sprinkler and Rain are independent.
Downstream inference

Sprinkler on

no Rain

Wet

on

off

rain

no rain

wet

wet

wet

dry

True propositions (at the world of evaluation)
(3)  a. If the sprinkler were off . . .
(3)  

a. ✓ If the sprinkler were off, it would be dry.  
b. ✗ If the sprinkler were off, it would be raining.
Goodman’s match

(3)  a. ✔If the match had been struck, it would have lighted.
b. ✗If the match had been struck, there would have been no oxygen.
Summary on downstream inference

Premise semantics and causality:

- Causal structure affects the set $f(w)$ of propositions relevant for the truth of counterfactuals.
- Whenever $X \rightarrow Y$, $f(w)$ contains $X_w$ and $X_w Y_w \rightarrow not Y_w$.
- Counterfactual reasoning about causes involves “undoing” their effects; but not vice versa.

介入.
Summary on downstream inference

Premise semantics and causality:

- Causal structure affects the set $f(w)$ of propositions relevant for the truth of counterfactuals.
- Whenever $X \rightarrow Y$, $f(w)$ contains $X_w$ and $X_w Y_w \leftarrow not Y_w$.
- Counterfactual reasoning about causes involves “undoing” their effects; but not vice versa.

**Intervention.**

But what about non-intervention counterfactuals?
Observation vs. intervention

Kratzer (1989):

King Ludwig of Bavaria likes to spend his weekends at Leoni Castle. Whenever the Royal Bavarian flag is up and the lights are on, the King is in the Castle. At the moment, the lights are on, the flag is down, and the King is away. Suppose now counterfactually that the flag were up.

(4) a. If the flag were up, the King would be in the Castle.
Kratzer (1989):

King Ludwig of Bavaria likes to spend his weekends at Leoni Castle. Whenever the Royal Bavarian flag is up and the lights are on, the King is in the Castle. At the moment, the lights are on, the flag is down, and the King is away. Suppose now counterfactually that the flag were up. Well, then the King would be in the Castle and the lights would still be on. But why wouldn’t the lights be out and the King still be away?

(4)  a. ✓ If the flag were up, the King would be in the Castle.
Kratzer (1989):

\[ \text{Let us change the scenario just a little bit \ldots I say to you: \text{"Suppose I hoisted the flag\ldots\"} \ldots \text{Would my hoisting the flag bring the King back into the Castle?} \]

(4) a. ✓ If the flag were up, the King would be in the Castle.
b. If I hoisted the flag, the King would appear in the Castle.
Let us change the scenario just a little bit . . . I say to you: “Suppose I hoisted the flag . . .” . . . Would my hoisting the flag bring the King back into the Castle? No. The counterfactual expressed by [4b] is false.

(4) a. ✓ If the flag were up, the King would be in the Castle.
   b. ✗ If I hoisted the flag, the King would appear in the Castle.
Observation vs. intervention

- Observation vs. intervention
  - expressed in the linguistic form of the antecedent
  - results in truth-conditional difference

(4)  
a. ✓ If the flag were up, the King would be in the Castle.
b. ✗ If I hoisted the flag, the King would appear in the Castle.
Goodman vs. Kratzer

Kratzer called her King of Bavaria example a “simplified variant” of Goodman’s match example. Is it?
Goodman vs. Kratzer

Kratzer:

King Ludwig of Bavaria likes to spend his weekends at Leoni Castle. Whenever the Royal Bavarian flag is up and the lights are on, the King is in the Castle. At the moment, the lights are on, the flag is down, and the King is away. Suppose now counterfactually that the flag were up. Well, then the King would be in the Castle and the lights would still be on. But why wouldn’t the lights be out and the King still be away?
Goodman vs. Kratzer

Kratzer:

King Ludwig of Bavaria likes to spend his weekends at Leoni Castle. *Whenever the Royal Bavarian flag is up and the lights are on, the King is in the Castle.* At the moment, the lights are on, the flag is down, and the King is away. Suppose now counterfactually that the flag were up. Well, then the King would be in the Castle and the lights would still be on. But why wouldn’t the lights be out and the King still be away?

Goodman’s match:

*Whenever a match is struck and oxygen is present, the match lights.*
**Goodman vs. Kratzer**

Struck?  Oxygen?

King?

Lights?  Flag?  Lights?

**Match:** Striking affects lighting

**King:** Flag may or may not affect the king
Goodman vs. Kratzer

**Struck?**

**Oxygen?**

**Lights?**

**King?**

**Flag?**

**Lights?**

**Match:** Striking affects lighting

**King:** Flag may or may not affect the king

▶ Observation vs. intervention again
Observation vs. intervention

Without intervention, the status of the sprinkler is determined by the weather.

With intervention, they are independent.
Observation vs. intervention

Intervention?  Rain?

Sprinkler?  Wet?

True propositions (at the world of evaluation).
Observation vs. intervention

Without intervention:

(5) a. If the sprinkler were off...
Observation vs. intervention

Without intervention:

(5)  
  a. If the sprinkler were off, it would be raining.
  b. If the sprinkler were off, it would be wet.
Observation vs. intervention

With intervention:

(5)  a. If the sprinkler were turned off. . .
Observation vs. intervention

With intervention:

(5)  a. ✓ If the sprinkler were turned off, it would be dry.
    b. ✗ If the sprinkler were turned off, it would be wet.
(6)  a. If I turned the sprinkler off . . .
    b. If the sprinkler were off . . .

- The two antecedents license different inferences about non-effects
Semantic agenda

(6)  

a. If I turned the sprinkler off . . .

b. If the sprinkler were off . . .

- The two antecedents license different inferences about non-effects
- How the difference between intervention and observation expressed linguistically?
  
  Aspectual properties? Thematic roles?

Dowty (1979, 1981)
The two antecedents license different inferences about non-effects.

How the difference between intervention and observation expressed linguistically?

What should a model-theoretic analysis look like?

Intertia worlds? Stereotypical ordering sources?
Conclusion

- Premise semantics: A useful general-purpose framework (not a theory itself)
Premise semantics: A useful general-purpose framework (not a theory itself)
Causal networks: Elegant, well-understood, empirically testable (but simplistic claims about counterfactual inference)
Conclusion

- Premise semantics: A useful general-purpose framework (not a theory itself)
- Causal networks: Elegant, well-understood, empirically testable (but simplistic claims about counterfactual inference)
- Linguistic questions still largely unexplored
Conclusion

- Premise semantics: A useful general-purpose framework (not a theory itself)
- Causal networks: Elegant, well-understood, empirically testable (but simplistic claims about counterfactual inference)
- Linguistic questions still largely unexplored
- Lots of work ahead.
The End.
Elements of causal premise semantics

How to interpret $\varphi \implies \psi$ at world $w$:

- $\langle U, \rightarrow \rangle$: Causal network
Elements of causal premise semantics

How to interpret $\varphi \square \rightarrow \psi$ at world $w$:

- $\langle U, \rightarrow \rangle$: Causal network
- $\rightarrow\rightarrow$: Transitive closure of $\rightarrow$
How to interpret $\varphi \Box \rightarrow \psi$ at world $w$:

- $\langle U, \rightarrow \rangle$: Causal network
- $\rightarrow$: Transitive closure of $\rightarrow$
- $[X_w]$: Proposition that variable $X$ has value $X_w$

$$[X_w] =_{df} \{ v \in W | X_v = X_w \}$$
Elements of causal premise semantics

How to interpret $\varphi \square \to \psi$ at world $w$:

- $\langle U, \to \rangle$: Causal network
- $\to$: Transitive closure of $\to$
- $[X_w]$: Proposition that variable $X$ has value $X_w$
- $f(w)$: Set of relevant propositions restricted by non-descendants

For each variable $X$, $f(w)$ contains the proposition that

- $X$ has value $X_w$ and
- $X$’s non-descendants $Y, Y', \ldots$ have values $Y_w, Y'_w, \ldots$

\[
f(w) = \left\{ [X_w] \cap [Y_w] \cap [Y'_w] \cap \ldots | X \Rightarrow Y^i, X \in U \right\}
\]
Elements of causal premise semantics

How to interpret $\varphi \square \rightarrow \psi$ at world $w$:

- $\langle U, \rightarrow \rangle$: Causal network
- $\rightarrow$: Transitive closure of $\rightarrow$
- $[X_w]$: Proposition that variable $X$ has value $X_w$
- $f(w)$: Set of relevant propositions restricted by non-descendants
- $\text{Prem}^c_w(\varphi)$: all consistent subsets of $f(w) \cup \{\varphi\}$ containing $\varphi$ and closed under logical consequence (relative to $f(w)$)

For all $X \in \text{Prem}^c_w(\varphi), p \in f(w)$:
If $p$ logically follows from $X \cap f(w)$, then $p \in X$
‘$\square \rightarrow$’ and strict implication

- If $\varphi$ is false at $w$, then $\varphi \overset{n}{\rightarrow} \psi$ comes down to $\square(\varphi \rightarrow \psi)$ (strict implication).
'□→' and strict implication

- If \( \varphi \) is false at \( w \), then \( \varphi \; \square^n \; \psi \) comes down to \( \square(\varphi \rightarrow \psi) \) (strict implication).

- In particular, if \( \square(\varphi \rightarrow \psi) \) is false, then so is \( \varphi \; \square^n \; \psi \).
If $\varphi$ is false at $w$, then $\varphi \nrightarrow^n \psi$ comes down to $\Box(\varphi \rightarrow \psi)$ (strict implication).

In particular, if $\Box(\varphi \rightarrow \psi)$ is false, then so is $\varphi \nrightarrow^n \psi$.

Suppose $\Box(\varphi \rightarrow \psi)$ is false.
If $\varphi$ is false at $w$, then $\varphi \nrightarrow^n \psi$ comes down to $\square(\varphi \rightarrow \psi)$ (strict implication).

In particular, if $\square(\varphi \rightarrow \psi)$ is false, then so is $\varphi \nrightarrow^n \psi$.

Suppose $\square(\varphi \rightarrow \psi)$ is false.

Then there is a world $w'$ at which $\varphi$ is true and $\psi$ is false.
If $\varphi$ is false at $w$, then $\varphi \nrightarrow^n \psi$ comes down to $\Box(\varphi \to \psi)$ (strict implication).

In particular, if $\Box(\varphi \to \psi)$ is false, then so is $\varphi \nrightarrow^n \psi$.

$\{w, w'\}$ is a proposition true at $w$.
\[\begin{align*}
\text{‘} \square \rightarrow \text{’ and strict implication} \\
\text{If } \varphi \text{ is false at } w, \text{ then } \varphi \overset{n}{\square} \psi \text{ comes down to } \square (\varphi \rightarrow \psi) \text{ (strict implication).} \\
\text{In particular, if } \square (\varphi \rightarrow \psi) \text{ is false, then so is } \varphi \overset{n}{\square} \psi. \\
\{w, w'\} \text{ is a proposition true at } w \\
X = \{\varphi, \{w, w'\}\} \text{ is a premise set} \\
\text{consistent;} \\
\text{contains } \varphi; \\
\text{all propositions except } \varphi \text{ true at } w
\end{align*}\]
‘$\square \rightarrow$’ and strict implication

- If $\varphi$ is false at $w$, then $\varphi \overset{n}{\square \rightarrow} \psi$ comes down to $\square(\varphi \rightarrow \psi)$ (strict implication).
- In particular, if $\square(\varphi \rightarrow \psi)$ is false, then so is $\varphi \overset{n}{\square \rightarrow} \psi$.

- $\{w, w'\}$ is a proposition true at $w$
- $X = \{\varphi, \{w, w'\}\}$ is a premise set
- $X$ and all its supersets entail $\neg \psi$
If $\varphi$ is false at $w$, then $\varphi \Box^n \psi$ comes down to $\Box(\varphi \rightarrow \psi)$ (strict implication).

In particular, if $\Box(\varphi \rightarrow \psi)$ is false, then so is $\varphi \Box^n \psi$.

- $\{w, w'\}$ is a proposition true at $w$
- $X = \{\varphi, \{w, w'\}\}$ is a premise set
- $X$ and all its supersets entail $\neg \psi$
- Hence $\varphi \Box^n \psi$ is false