Conditionals: between language and reasoning

Class 12: background semantics for counterfactuals

February 2, 2018

Two loose ends

1. Causal modeling semantics and minimal change semantics

- We saw that law-based approaches like Veltman's and Pearl's allow us to make concrete predictions where minimal change semantics *per se* is silent.
- This is compatible with m.-c. semantics being correct as far as it goes: perhaps it just needs to be supplemented with a theory of similarity.
- Maybe we can actually use causal structures to provide such a theory: i.e., from a causal model we can distill a similarity ordering which yields the right results.
- This indeed succeeds under certain assumptions (Marti & Pinosio 2014)

Two loose ends

1. Causal modeling semantics and minimal change semantics

   - Today we'll look at some data that contradict minimal change semantics.
   - How can we do that?
   - We can't just argue that these data are not predicted by a certain similarity ordering: for one might just reply that we are using the wrong ordering.

Two loose ends

1. Causal modeling semantics and minimal change semantics

   - Today we'll look at some data that contradict minimal change semantics.
   - How can we do that?
   - We can't just argue that these data are not predicted by a certain similarity ordering: for one might just reply that we are using the wrong ordering.
   - We need to show data that cannot be predicted by any similarity ordering.

Two loose ends

1. Causal modeling semantics and minimal change semantics

- Today we'll look at some data that contradict minimal change semantics.
- How can we do that?
- We can't just argue that these data are not predicted by a certain similarity ordering: for one might just reply that we are using the wrong ordering.
- We need to show data that cannot be predicted by any similarity ordering.
- In other words: we must challenge the logic of minimal change semantics.
- The logic of a theory doesn't depend on any specific choice of contextual parameters, but rather on the mathematical workings of the semantics.

Two loose ends

1. Causal modeling semantics and minimal change semantics

- Today we'll look at some data that contradict minimal change semantics.
- How can we do that?
- We can't just argue that these data are not predicted by a certain similarity ordering: for one might just reply that we are using the wrong ordering.
- We need to show data that cannot be predicted by any similarity ordering.
- In other words: we must challenge the logic of minimal change semantics.
- The logic of a theory doesn't depend on any specific choice of contextual parameters, but rather on the mathematical workings of the semantics.
- Moreover, we'll look at how we can dispense with the notion of similarity, replacing it with the simpler and less mysterious notion of a background.

Two loose ends

2. Causal modeling semantics and arbitrary antecedents

- We saw that Pearl's causal modeling semantics gives a simple and powerful analysis of counterfactuals in terms of intervention.
- On the other hand, the analysis is not fully general, but is restricted to a limited class of counterfactuals.
- We can only interpret antecedents which are conjunctions of atoms:

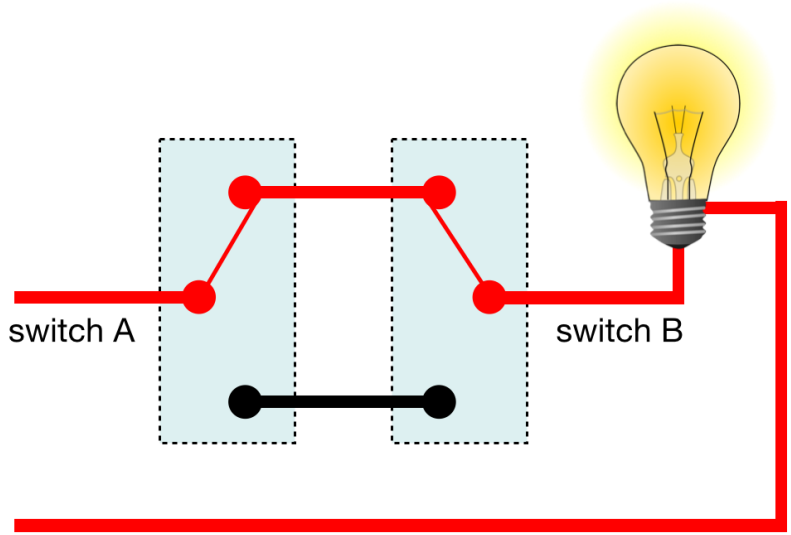
$$(X = a_1) \wedge \cdots \wedge (X_n = a_n)$$

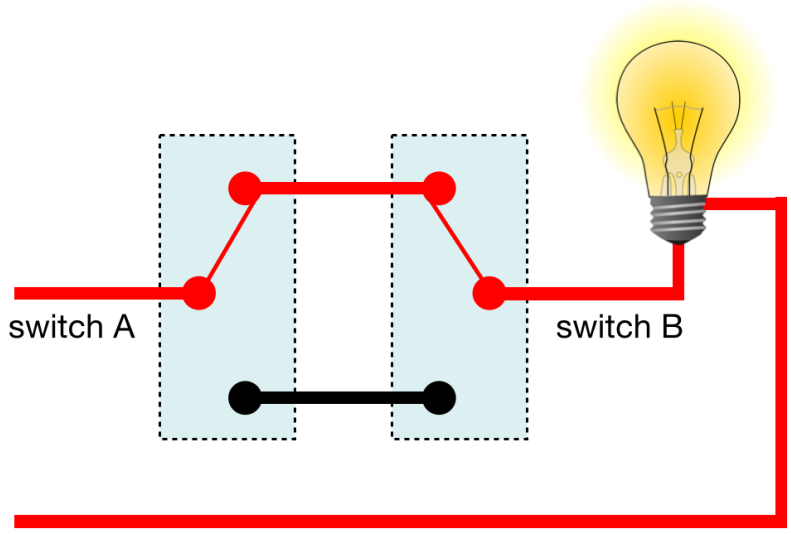
- How can we get rid of this limitation?
- The proposal we'll look at today allows us to use a causal model to interpret counterfactuals with arbitrary antecedents and consequents.

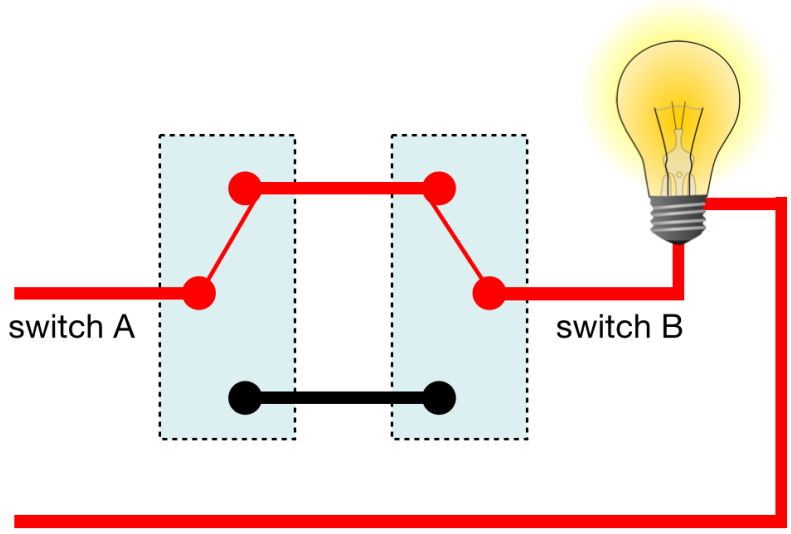
Evidence against minimal change semantics

Two switches again

Imagine a long hallway with a light in the middle and with two switches, one at each end. One switch is called switch A and the other one is called switch B. As the following wiring diagram shows, the light is on whenever both switches are in the same position (both up or both down); otherwise, the light is off. Right now, switch A and switch B are both up, and the light is on. But things could be different. . .




switch A
switch B



(1) If switch A was down, the light would be off.

(2) If switch B was down, the light would be off.

(3) If switch A or switch B was down, the light would be off.

(4) If switch A and switch B were not both up, the light would be off.

(5) If switch A and switch B were not both up, the light would be on.

Main experiment

- We conducted a survey using Amazon's Mechanical Turk.
- Each participant saw the text and figure above.
- We used (6) as a filler, which is clearly false in the described scenario:

  (6) If switch A and switch B were both down, the light would be off.

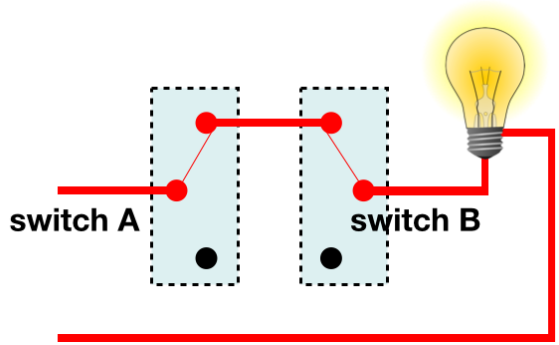
- Each participant saw a target item and the filler, in random order, and was asked to judge the sentences as 'true', 'false', or 'indeterminate'.
- We rejected data from participants whose native language was not English, judged the filler incorrectly or had already taken the survey.

Table 3: Results of the main experiment

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\overline{A}$ > OFF	256	169	66.02%	6	2.34%	81	31.64%
$\overline{B}$ > OFF	235	153	65.11%	7	2.98%	75	31.91%
$\overline{A} \vee \overline{B}$ > OFF	362	251	69.33%	14	3.87%	97	26.80%
$\neg(A \wedge B)$ > OFF	372	82	22.04%	136	36.56%	154	41.40%
$\neg(A \wedge B)$ > ON	200	43	21.50%	63	31.50%	94	47.00%

Excluding confounds

- The data about (5) ensure that participants did not read "not both up" as "both not up": otherwise (5) would be equivalent to the filler, and true.
- We also wanted to make sure that (4) are (5) are not rejected based on context-independent reasons (e.g., excessive processing load).
- For this, we tested our sentences in a different scenario:
  the light is on if and only if the switches are both up.



Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\overline{A}$ > OFF	52	41	78.85%	5	9.61%	6	11.54%
$\overline{B}$ > OFF	68	60	88.24%	5	7.35%	3	4.41%
$\overline{A} \vee \overline{B}$ > OFF	110	104	94.55%	1	0.91%	5	4.54%
$\neg(A \wedge B)$ > OFF	116	99	85.34%	9	7.76%	8	6.90%
$\neg(A \wedge B)$ > ON	103	19	18.45%	79	76.70%	5	4.85%

Post-hoc test II

- We assumed that being down is equivalent to not being up:

$$\overline{A} \equiv \neg A \qquad \overline{B} \equiv \neg B$$

- We ran a post-hoc test to make sure that this assumption is innocent.
- For this, we tested versions of our sentences where *down* is explicitly replaced by *not up*, in the setting of our main experiment.
- The difference between the two blocks remains in place:

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\neg A >$ OFF	36	27	75.00%	1	2.78%	8	22.22%
$\neg B >$ OFF	43	28	65.12%	7	16.28%	8	18.60%
$\neg A \vee \neg B >$ OFF	80	48	60.00%	16	20.00%	16	20.00%
$\neg(A \wedge B) >$ OFF	372	82	22.04%	136	36.56%	154	41.40%
$\neg(A \wedge B) >$ ON	200	43	21.50%	63	31.50%	94	47.00%

We conclude that in our scenario (at least under the most widespread reading), (7) and (8) are true, but (9) is not.

(7) If switch A was not up, the light would be off. ✓

(8) If switch B was not up, the light would be off. ✓

(9) If switch A and switch B were not both up, the light would be off. ✗

In symbols:

$\neg A > \mathit{Off}$	✓
$\neg B > \mathit{Off}$	✓
$\neg(A \wedge B) > \mathit{Off}$	✗

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Lemma: given any relative similarity ordering $\leq$,

$$\min_w(\neg(A \wedge B)) \ \subseteq\ \min_w(\neg A) \cup \min_w(\neg B)$$

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Lemma: given any relative similarity ordering $\leq$,

$$\min_w(\neg(A \wedge B)) \ \subseteq\ \min_w(\neg A) \cup \min_w(\neg B)$$

Proof.

- Take $v \in \min_w(\neg(A \wedge B))$.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Lemma: given any relative similarity ordering $\leq$,

$$\min_w(\neg(A \wedge B)) \ \subseteq\ \min_w(\neg A) \cup \min_w(\neg B)$$

Proof.

- Take $v \in \min_w(\neg(A \wedge B))$.
- Since v is a $\neg(A \wedge B)$ world, either A or B is false at v.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Lemma: given any relative similarity ordering $\leq$,

$$\min_w(\neg(A \wedge B)) \ \subseteq\ \min_w(\neg A) \cup \min_w(\neg B)$$

Proof.

- Take $v \in \min_w(\neg(A \wedge B))$.
- Since v is a $\neg(A \wedge B)$ world, either A or B is false at v.
- Suppose A is false at v. Then v is a $\neg A$-world.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Lemma: given any relative similarity ordering $\leq$,

$$\min_w(\neg(A \wedge B)) \ \subseteq\ \min_w(\neg A) \cup \min_w(\neg B)$$

Proof.

- Take $v \in \min_w(\neg(A \wedge B))$.
- Since v is a $\neg(A \wedge B)$ world, either A or B is false at v.
- Suppose A is false at v. Then v is a $\neg A$-world.
- Can there be another $\neg A$ world u with $u <_w v$?

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Lemma: given any relative similarity ordering $\leq$,

$$\min_w(\neg(A \wedge B)) \ \subseteq\ \min_w(\neg A) \cup \min_w(\neg B)$$

Proof.

- Take $v \in \min_w(\neg(A \wedge B))$.
- Since v is a $\neg(A \wedge B)$ world, either A or B is false at v.
- Suppose A is false at v. Then v is a $\neg A$-world.
- Can there be another $\neg A$ world u with $u <_w v$?
- If so, u would be a $\neg(A \wedge B)$ world closer than v.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Lemma: given any relative similarity ordering $\leq$,

$$\min_w(\neg(A \wedge B)) \ \subseteq\ \min_w(\neg A) \cup \min_w(\neg B)$$

Proof.

- Take $v \in \min_w(\neg(A \wedge B))$.
- Since v is a $\neg(A \wedge B)$ world, either A or B is false at v.
- Suppose A is false at v. Then v is a $\neg A$-world.
- Can there be another $\neg A$ world u with $u <_w v$?
- If so, u would be a $\neg(A \wedge B)$ world closer than v.
- This contradics $v \in \min_w(\neg(A \wedge B))$.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \models \neg(A \wedge B) > \mathit{Off}$$

Lemma: given any relative similarity ordering $\leq$,

$$\min_w(\neg(A \wedge B)) \subseteq \min_w(\neg A) \cup \min_w(\neg B)$$

Proof.

- Take $v \in \min_w(\neg(A \wedge B))$.
- Since v is a $\neg(A \wedge B)$ world, either A or B is false at v.
- Suppose A is false at v. Then v is a $\neg A$-world.
- Can there be another $\neg A$ world u with $u <_w v$?
- If so, u would be a $\neg(A \wedge B)$ world closer than v.
- This contradics $v \in \min_w(\neg(A \wedge B))$.
- Thus, $v \in \min_w(\neg A)$.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Lemma: given any relative similarity ordering $\leq$,

$$\min_w(\neg(A \wedge B)) \ \subseteq\ \min_w(\neg A) \cup \min_w(\neg B)$$

Proof.

- Take $v \in \min_w(\neg(A \wedge B))$.
- Since v is a $\neg(A \wedge B)$ world, either A or B is false at v.
- Suppose A is false at v. Then v is a $\neg A$-world.
- Can there be another $\neg A$ world u with $u <_w v$?
- If so, u would be a $\neg(A \wedge B)$ world closer than v.
- This contradics $v \in \min_w(\neg(A \wedge B))$.
- Thus, $v \in \min_w(\neg A)$.
- Similarly, if B is false at v we conclude $v \in \min_w(\neg B)$.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off}\,,\ \neg B > \mathit{Off}\ \models\ \neg(A \wedge B) > \mathit{Off}$$

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off}, \neg B > \mathit{Off} \models \neg(A \wedge B) > \mathit{Off}$$

Proof:

- Suppose $\neg A > \mathit{Off}$ and $\neg B > \mathit{Off}$ are true at w.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Proof:

- Suppose $\neg A > \mathit{Off}$ and $\neg B > \mathit{Off}$ are true at w.
- Then $\min_w(\neg A) \subseteq |\mathit{Off}|$ and $\min_w(\neg B) \subseteq |\mathit{Off}|$.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Proof:

- Suppose $\neg A > \mathit{Off}$ and $\neg B > \mathit{Off}$ are true at w.
- Then $\min_w(\neg A) \subseteq |\mathit{Off}|$ and $\min_w(\neg B) \subseteq |\mathit{Off}|$.
- By the previous lemma, $\min_w(\neg(A \wedge B)) \subseteq \min_w(\neg A) \cup \min_w(\neg B)$.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Proof:

- Suppose $\neg A > \mathit{Off}$ and $\neg B > \mathit{Off}$ are true at w.
- Then $\min_w(\neg A) \subseteq |\mathit{Off}|$ and $\min_w(\neg B) \subseteq |\mathit{Off}|$.
- By the previous lemma, $\min_w(\neg(A \wedge B)) \subseteq \min_w(\neg A) \cup \min_w(\neg B)$.
- So $\min_w(\neg(A \wedge B)) \subseteq |\mathit{Off}|$.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Proof:

- Suppose $\neg A > \mathit{Off}$ and $\neg B > \mathit{Off}$ are true at w.
- Then $\min_w(\neg A) \subseteq |\mathit{Off}|$ and $\min_w(\neg B) \subseteq |\mathit{Off}|$.
- By the previous lemma, $\min_w(\neg(A \wedge B)) \subseteq \min_w(\neg A) \cup \min_w(\neg B)$.
- So $\min_w(\neg(A \wedge B)) \subseteq |\mathit{Off}|$.
- Thus, $\neg(A \wedge B) > \mathit{Off}$ is true at w.

Theorem. The following is valid in minimal change semantics:

$$\neg A > \mathit{Off},\ \neg B > \mathit{Off} \ \models\ \neg(A \wedge B) > \mathit{Off}$$

Proof:

- Suppose $\neg A > \mathit{Off}$ and $\neg B > \mathit{Off}$ are true at w.
- Then $\min_w(\neg A) \subseteq |\mathit{Off}|$ and $\min_w(\neg B) \subseteq |\mathit{Off}|$.
- By the previous lemma, $\min_w(\neg(A \wedge B)) \subseteq \min_w(\neg A) \cup \min_w(\neg B)$.
- So $\min_w(\neg(A \wedge B)) \subseteq |\mathit{Off}|$.
- Thus, $\neg(A \wedge B) > \mathit{Off}$ is true at w.

Conclusion

No similarity ordering can predict that the first two sentences below are true but the third is not:

$\neg A > \mathit{Off}$	✓
$\neg B > \mathit{Off}$	✓
$\neg(A \wedge B) > \mathit{Off}$	✗

Remark.

It is not a particular conception of similarity that is problematic:
Our observations contradict the logic of ordering semantics.

Background semantics

If switch A was down...

- we hold the position of B fixed.

If switch A and switch B were not both up...

- we consider all positions for the switches,
  not only those most similar to the actual one.

If switch A was down...

- we hold the position of B fixed.

If switch A and switch B were not both up...

- we consider all positions for the switches,
  not only those most similar to the actual one.

So...

- There seems to be no general pressure to consider only scenarios that are minimally different from the actual world.

If switch A was down...

- we hold the position of B fixed.

If switch A and switch B were not both up...

- we consider all positions for the switches,
  not only those most similar to the actual one.

So...

- There seems to be no general pressure to consider only scenarios that are minimally different from the actual world.
- But then, why do we hold the position of B fixed when faced with the assumption that A was down?

If switch A was down...

- we hold the position of B fixed.

If switch A and switch B were not both up...

- we consider all positions for the switches,
  not only those most similar to the actual one.

So...

- There seems to be no general pressure to consider only scenarios
  that are minimally different from the actual world.
- But then, why do we hold the position of B fixed
  when faced with the assumption that A was down?
- Proposal: in that case, the position of B is not called into question.
  It can be regarded as part of the background for the assumption.

Background semantics

- In background semantics we abandon the "minimal change principle", proposing instead a distinction between background and foreground facts.
- When faced with an assumption, we determine which facts to manipulate (foreground) and which to leave alone (background).
- Background facts are held fixed; foreground facts are allowed to change, and their change is not subject to any minimality constraint.

Background semantics

- In background semantics we abandon the "minimal change principle", proposing instead a distinction between background and foreground facts.
- When faced with an assumption, we determine which facts to manipulate (foreground) and which to leave alone (background).
- Background facts are held fixed; foreground facts are allowed to change, and their change is not subject to any minimality constraint.
- $\varphi > \psi$ is true iff ψ follows by causal laws from φ + background facts (this can be seen as an implementation of the meta-linguistic theory).

Principles ruling the background/foreground split:

Principles ruling the background/foreground split:

1. Facts responsible for the falsity of the assumption must be foregrounded.
    - Give up at least those facts which are in conflict with the assumption.

Principles ruling the background/foreground split:

1. Facts responsible for the falsity of the assumption must be foregrounded.
   - Give up at least those facts which are in conflict with the assumption.
2. Causal consequences of foregrounded facts must be foregrounded.
   - If we give up a fact, we must also give up things that depend on it.

Principles ruling the background/foreground split:

1. Facts responsible for the falsity of the assumption must be foregrounded.
   - Give up at least those facts which are in conflict with the assumption.
2. Causal consequences of foregrounded facts must be foregrounded.
   - If we give up a fact, we must also give up things that depend on it.
3. By default, facts are backgrounded.
   - Don't give up facts without a reason.

Causal models (drawing on Pearl 00, Kaufmann 13, Santorio 15)
A causal model has a set V of causal variables and a set L of causal laws.

Causal variables
A causal variable X is a partition of the logical space.

- The cells are called settings.
- The true setting of X at w is called the value of X at w, denoted X_w.

In our context, $V = \{?A, ?B, ?On\}$:

- $?A = \{A, \overline{A}\}$
- $?B = \{B, \overline{B}\}$
- $?On = \{On, Off\}$

Causal laws

A causal law I is a triple $\langle C_I, E_I, m_I \rangle$, where:

- C_I is a set of variables (the causes)
- E_I is a variable (the effect)
- m_I is a map from settings of C_I to settings of E_I

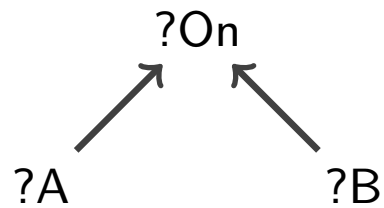
In our context, there is only one law.

- Causes: $?A, ?B$
- Effect: $?On$
- Map:

  $A, B \mapsto On$ $\quad A, \overline{B} \mapsto Off$

  $\overline{A}, B \mapsto Off$ $\quad \overline{A}, \overline{B} \mapsto On$

Causal graph:



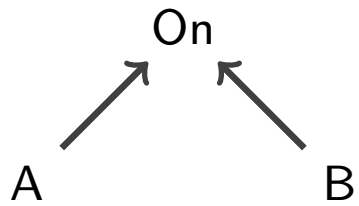
Facts

The facts at w are values of the causal variables:

$$\mathcal{F}_w = \{X_w \mid X \in V\}$$

In our scenario:

$\mathcal{F}_w = \{A, B, On\}$



Fact f contributes to the falsity of assumption a at world w if there exists some set F' of facts such that:

- F' is logically consistent with a
- $F' \cup \{f\}$ is logically inconsistent with a

Fact f contributes to the falsity of assumption a at world w if there exists some set F' of facts such that:

- F' is logically consistent with a
- $F' \cup \{f\}$ is logically inconsistent with a

Assump.	Facts contributing to its falsity
$\overline{A}$	A
$\overline{B}$	B
$\neg(A \wedge B)$	A, B

Background conditions

A set $\mathcal{B}(w, a) \subseteq \mathcal{F}_w$ is a background for assumption a at w if it satisfies:

1. Facts responsible for falsity of the antecedent are foregrounded:
   if f contributes to the falsity of a then $f \notin \mathcal{B}(w, a)$

2. Consequences of foregrounded facts are foregrounded:
   if $f \notin \mathcal{B}(w, a)$ and f' is causally dependent on f, then $f' \notin \mathcal{B}(w, a)$.

Background conditions

A set $\mathcal{B}(w, a) \subseteq \mathcal{F}_w$ is a background for assumption a at w if it satisfies:

1. Facts responsible for falsity of the antecedent are foregrounded:
   if f contributes to the falsity of a then $f \notin \mathcal{B}(w, a)$

2. Consequences of foregrounded facts are foregrounded:
   if $f \notin \mathcal{B}(w, a)$ and f' is causally dependent on f, then $f' \notin \mathcal{B}(w, a)$.

For any assumption there is a greatest background:

- $\mathcal{B}^{max}(w, a) = \mathcal{F}_w - \{f \mid f$ contributes to the falsity of a or
  f is dependent on such a fact$\}$

Background conditions

A set $\mathcal{B}(w, a) \subseteq \mathcal{F}_w$ is a background for assumption a at w if it satisfies:

1. Facts responsible for falsity of the antecedent are foregrounded:
   if f contributes to the falsity of a then $f \notin \mathcal{B}(w, a)$

2. Consequences of foregrounded facts are foregrounded:
   if $f \notin \mathcal{B}(w, a)$ and f' is causally dependent on f, then $f' \notin \mathcal{B}(w, a)$.

For any assumption there is a greatest background:

- $\mathcal{B}^{max}(w, a) = \mathcal{F}_w - \{f \mid f$ contributes to the falsity of a or
  f is dependent on such a fact$\}$

Assump.	$\mathcal{B}^{max}(w, a)$
$\overline{A}$	$\{B\}$
$\overline{B}$	$\{A\}$
$\neg(A \wedge B)$	$\emptyset$

Hypothetical context

Making an assumption a in world w given a background $\mathcal{B}(w, a)$ results in the hypothetical context $f_{\mathcal{B}}(w, a)$ consisting of worlds where:

- a is true
- all facts in $\mathcal{B}(w, a)$ are true
- the causal laws are obeyed (simplification)

Hypothetical context

Making an assumption a in world w given a background $\mathcal{B}(w, a)$ results in the hypothetical context $f_{\mathcal{B}}(w, a)$ consisting of worlds where:

- a is true
- all facts in $\mathcal{B}(w, a)$ are true
- the causal laws are obeyed (simplification)

Assessment of a counterfactual

Given a causal model M, a background map $\mathcal{B}$, and a world w:

$$M, w \models \varphi > \psi \quad \iff \quad f_{\mathcal{B}}(w, |\varphi|) \subseteq |\psi|$$

Hypothetical context

Making an assumption a in world w given a background $\mathcal{B}(w, a)$ results in the hypothetical context $f_{\mathcal{B}}(w, a)$ consisting of worlds where:

- a is true
- all facts in $\mathcal{B}(w, a)$ are true
- the causal laws are obeyed (simplification)

Assessment of a counterfactual

Given a causal model M, a background map $\mathcal{B}$, and a world w:

$$M, w \models \varphi > \psi \quad \iff \quad f_{\mathcal{B}}(w, |\varphi|) \subseteq |\psi|$$

NB

Notice that here φ and ψ are arbitrary: unlike Pearl's, this account is not restricted to special antecedents.

$\overline{A} > \mathit{Off}$

- Maximal background: $\{B\}$

$\overline{A} > \mathit{Off}$

- Maximal background: $\{B\}$
- Hypothetical context: $|\overline{A}| \cap |B| \cap |\mathit{law}| \subseteq |\mathit{Off}|$

$\overline{A} > \mathit{Off}$

- Maximal background: $\{B\}$
- Hypothetical context: $|\overline{A}| \cap |B| \cap |\mathit{law}| \subseteq |\mathit{Off}|$
- True.

$\overline{A} > \mathit{Off}$

- Maximal background: $\{B\}$
- Hypothetical context: $|\overline{A}| \cap |B| \cap |\mathit{law}| \subseteq |\mathit{Off}|$
- True.

$\overline{B} > \mathit{Off}$

- True (analogous).

$\overline{A} > \mathit{Off}$

- Maximal background: $\{B\}$
- Hypothetical context: $|\overline{A}| \cap |B| \cap |\mathit{law}| \subseteq |\mathit{Off}|$
- True.

$\overline{B} > \mathit{Off}$

- True (analogous).

$\neg(A \wedge B) > \mathit{Off}$

$\overline{A} > \mathit{Off}$

- Maximal background: $\{B\}$
- Hypothetical context: $|\overline{A}| \cap |B| \cap |\mathit{law}| \subseteq |\mathit{Off}|$
- True.

$\overline{B} > \mathit{Off}$

- True (analogous).

$\neg(A \wedge B) > \mathit{Off}$

- Maximal background: $\emptyset$

$\overline{A} > \mathit{Off}$

- Maximal background: $\{B\}$
- Hypothetical context: $|\overline{A}| \cap |B| \cap |\mathit{law}| \subseteq |\mathit{Off}|$
- True.

$\overline{B} > \mathit{Off}$

- True (analogous).

$\neg(A \wedge B) > \mathit{Off}$

- Maximal background: $\emptyset$
- Hypothetical context: $|\neg(A \wedge B)| \cap |\mathit{law}| \not\subseteq |\mathit{Off}|$

$\overline{A} > \mathit{Off}$

- Maximal background: $\{B\}$
- Hypothetical context: $|\overline{A}| \cap |B| \cap |\mathit{law}| \subseteq |\mathit{Off}|$
- True.

$\overline{B} > \mathit{Off}$

- True (analogous).

$\neg(A \wedge B) > \mathit{Off}$

- Maximal background: $\emptyset$
- Hypothetical context: $|\neg(A \wedge B)| \cap |\mathit{law}| \not\subseteq |\mathit{Off}|$
- Not true.

$\overline{A} > \mathit{Off}$

- Maximal background: $\{B\}$
- Hypothetical context: $|\overline{A}| \cap |B| \cap |\mathit{law}| \subseteq |\mathit{Off}|$
- True.

$\overline{B} > \mathit{Off}$

- True (analogous).

$\neg(A \wedge B) > \mathit{Off}$

- Maximal background: $\emptyset$
- Hypothetical context: $|\neg(A \wedge B)| \cap |\mathit{law}| \not\subseteq |\mathit{Off}|$
- Not true.

$\neg(A \wedge B) > \mathit{On}$

- Not true (analogous).

The background theory naturally explains the majority judgments we found:

(10) If switch A was down, the light would be off. ✓

(11) If switch B was down, the light would be off. ✓

(12) If switch A and switch B were not both up, the light would be off. ✗

(13) If switch A and switch B were not both up, the light would be on. ✗

What about the remaining counterfactual, (14)?

(14) If switch A or switch B was down, the light would be off. ✓

$\neg A \vee \neg B > \mathit{Off}$

- As we saw in Class 8, teasing apart $\neg A \vee \neg B > \mathit{Off}$ and $\neg(A \wedge B) > \mathit{Off}$ requires a semantic framework that goes beyond truth-conditions and breaks de Morgan's law.
- This can be done in inquisitive semantics: the disjunction $\neg A \vee \neg B$ introduces two separate propositions as counterfactual assumptions.
- As a result, $\neg A \vee \neg B > \mathit{Off} \equiv (\neg A > \mathit{Off}) \wedge (\neg B > \mathit{Off})$ is predicted true.
- The predictions about the other sentences are not affected, since these sentences involve no inquisitive operators.
- So, combining the background theory with inquisitive semantics we get truth-conditions that accord with the majority judgments.

Evidence for non-maximal background

- In our main experiment, each participant saw one of the target sentences, as well as the filler, (15):

  (15) If switch A and switch B were both down, the light would be off.

  $\overline{A} \wedge \overline{B} > \mathit{Off}$

- Some participants saw the target item first, while others saw the filler first.

Ordering effects

Table 7: Order effects in the main experiment: target precedes filler

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\overline{A}$ > OFF	125	100	80%	3	2.4%	22	17.6%
$\overline{B}$ > OFF	124	94	75.81%	4	3.22%	26	20.97%
$\overline{A} \vee \overline{B}$ > OFF	185	146	78.92%	9	4.86%	30	16.22%
$\neg(A \wedge B)$ > OFF	193	38	19.69%	82	42.49%	73	37.82%
$\neg(A \wedge B)$ > ON	102	21	20.59%	35	34.31%	46	45.10%

Table 8: Order effects in the main experiment: filler precedes target

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\overline{A}$ > OFF	131	69	52.67%	3	2.29%	59	45.04%
$\overline{B}$ > OFF	111	59	53.15%	3	2.70%	49	44.14%
$\overline{A} \vee \overline{B}$ > OFF	177	105	59.32%	5	2.82%	67	37.85%
$\neg(A \wedge B)$ > OFF	179	44	24.58%	54	30.17%	81	45.25%
$\neg(A \wedge B)$ > ON	98	22	22.45%	28	28.57%	48	48.98%

- The order effects can be explained assuming that, in addition to the antecedent, other factors may lead to foregrounding certain facts.
- In our case, the filler is: $\overline{A} \wedge \overline{B} > \mathit{Off}$:

  (16) If switch A and switch B were both down, the light would be off.

- The assumption $\overline{A} \wedge \overline{B}$ foregrounds the position of both switches.
- To some participants, once the position of B has been foregrounded, it remains foregrounded when interpreting the antecedent $\overline{A}$.
- This leads to empty background, and to judge $\overline{A} > \mathit{Off}$ as 'indeterminate'.
- The same explanation works for $\overline{B} > \mathit{Off}$ and for $\overline{A} \vee \overline{B} > \mathit{Off}$.

- Why no order effects for $\neg(A \wedge B) > \mathit{Off}$ and $\neg(A \wedge B) > \mathit{On}$?
- The assumption $\neg(A \wedge B)$ already foregrounds all the facts.
- Interpreting the filler cannot lead to more facts being foregrounded.

- Why no order effects for $\neg(A \wedge B) > \mathit{Off}$ and $\neg(A \wedge B) > \mathit{On}$?
- The assumption $\neg(A \wedge B)$ already foregrounds all the facts.
- Interpreting the filler cannot lead to more facts being foregrounded.
- These data indicate that in order to decide what to background, we take into account not just the antecedent, but also the preceding discourse.
- This is reminiscent of von Fintel's expanding modal horizon: once certain possibilities have been made salient, they can no longer be ignored.

Breaking causal laws

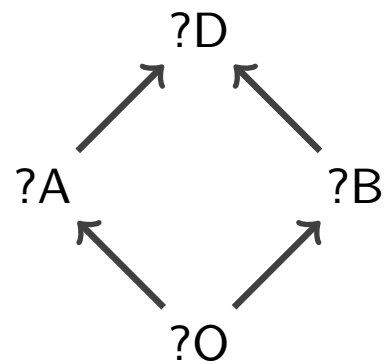
Pearl's (simplified) firing squad scenario

Variables

- Whether officer orders execution ($?O$)
- Whether rifleman A fires ($?A$)
- Whether rifleman B fires ($?B$)
- Whether prisoner dies ($?D$)

Laws

- $A \leftrightarrow O$
- $B \leftrightarrow O$
- $D \leftrightarrow A \vee B$



Actual world

$\mathcal{F}_w = \{O, A, B, D\}$

What if A had not shot?

- Consider the counterfactual assumption $\neg A$.
- The only fact that contributes to its falsity is A.
- The maximal factual background is $\{O, B\}$.
- But $\{\neg A, O, B\}$ is inconsistent with the law $A \leftrightarrow O$.
- To make room for the assumption $\neg A$, we must discard the law $A \leftrightarrow O$.

What if A had not shot?

- Consider the counterfactual assumption $\neg A$.
- The only fact that contributes to its falsity is A.
- The maximal factual background is $\{O, B\}$.
- But $\{\neg A, O, B\}$ is inconsistent with the law $A \leftrightarrow O$.
- To make room for the assumption $\neg A$, we must discard the law $A \leftrightarrow O$.

Discarding laws

- In making an assumption a, we discard a law l with effect E in case E_w contributes to the falsity of a.
- The hypothetical context created by a consists of those worlds where a is true, the background facts are true, and the remaining laws are obeyed.

What if A had not fired?

- The assumption is $\neg A$.
- The only fact that contributes to its falsity is A.
- Laws:
  - $A \leftrightarrow O$
  - $B \leftrightarrow O$
  - $D \leftrightarrow A \vee B$
- We discard the law $A \leftrightarrow O$.
- The hypothetical context is given by:
  - assumption $\neg A$
  - background facts $\{O, B\}$
  - background laws $\{B \leftrightarrow O, D \leftrightarrow A \vee B\}$.
- This entails D.
- Therefore, (17) is predicted to be true.

(17) If A had not shot, the prisoner would still have died.

Firing squad, example 2

(18) If A and B had not both shot, the prisoner would still have died.

- The counterfactual assumption is $\neg(A \wedge B)$.
- The fact that contribute to the falsity of the assumption are A and B.
- The maximal factual background is $\{O\}$.
- The antecedent intervenes on the laws $A \leftrightarrow O$ and $B \leftrightarrow O$.
- The hypothetical context is given by the facts $\{O, \neg(A \wedge B)\}$ and the law $D \leftrightarrow A \vee B$.
- This does not entail D.
- Therefore, (18) is not predicted to be true.

More evidence for the background/foreground divide

More evidence for the background/foreground divide

(19) If I suddenly became taller than I am, you would not notice.

More evidence for the background/foreground divide

(19) If I suddenly became taller than I am, you would not notice. ?

More evidence for the background/foreground divide

(19) If I suddenly became taller than I am, you would not notice. ?

- My height is foregrounded: no pressure to keep it close to actual.

More evidence for the background/foreground divide

(19) If I suddenly became taller than I am, you would not notice. ?

- My height is foregrounded: no pressure to keep it close to actual.
- On the other hand, your ability to pick up differences in height is backgrounded, and held fixed. This is crucial:

(20) If I suddenly became much taller than I am, you would notice.

More evidence for the background/foreground divide

(19) If I suddenly became taller than I am, you would not notice. ?

- My height is foregrounded: no pressure to keep it close to actual.
- On the other hand, your ability to pick up differences in height is backgrounded, and held fixed. This is crucial:

(20) If I suddenly became much taller than I am, you would notice. ✓

More evidence for the background/foreground divide

(19) If I suddenly became taller than I am, you would not notice. ?

- My height is foregrounded: no pressure to keep it close to actual.
- On the other hand, your ability to pick up differences in height is backgrounded, and held fixed. This is crucial:

(20) If I suddenly became much taller than I am, you would notice. ✓

- Of course, we can make the right predictions in minimal change semantics: stipulate that people's ability matters for similarity, but my height doesn't.

More evidence for the background/foreground divide

(19) If I suddenly became taller than I am, you would not notice. ?

- My height is foregrounded: no pressure to keep it close to actual.
- On the other hand, your ability to pick up differences in height is backgrounded, and held fixed. This is crucial:

(20) If I suddenly became much taller than I am, you would notice. ✓

- Of course, we can make the right predictions in minimal change semantics: stipulate that people's ability matters for similarity, but my height doesn't.
- But why?

More evidence for the background/foreground divide

(19) If I suddenly became taller than I am, you would not notice. ?

- My height is foregrounded: no pressure to keep it close to actual.
- On the other hand, your ability to pick up differences in height is backgrounded, and held fixed. This is crucial:

(20) If I suddenly became much taller than I am, you would notice. ✓

- Of course, we can make the right predictions in minimal change semantics: stipulate that people's ability matters for similarity, but my height doesn't.
- But why?
- Also, with a different antecedent (e.g., if you had been hallucinating), we might have to assume the opposite: but similarity is not supposed to be sentence-relative.

More evidence for the background/foreground divide

(19) If I suddenly became taller than I am, you would not notice. ?

- My height is foregrounded: no pressure to keep it close to actual.
- On the other hand, your ability to pick up differences in height is backgrounded, and held fixed. This is crucial:

(20) If I suddenly became much taller than I am, you would notice. ✓

- Of course, we can make the right predictions in minimal change semantics: stipulate that people's ability matters for similarity, but my height doesn't.
- But why?
- Also, with a different antecedent (e.g., if you had been hallucinating), we might have to assume the opposite: but similarity is not supposed to be sentence-relative.
- Background semantics provides a simple explanation: we only manipulate those variable that are currently at stake, and their causal consequences, holding the rest fixed.

Conclusions:

- Our observations are incompatible with the idea that making a counterf. assumption involves maximizing similarity to the world of evaluation.
- This holds regardless of how similarity is construed:
  our observations violate the logic of minimal-change accounts.
- Proposal: replace the minimal change idea with a distinction between foreground and background for a given assumption.
- Background is held fixed; foreground can change non-minimally.
- This yields a natural account of the judgments on the switches scenario, and of other observations (ordering effects in data, height examples).
- At the same time, background semantics can be seen as a generalization of Pearl's semantics which allows us to interpret arbitrary counterfactuals.

Appendix A

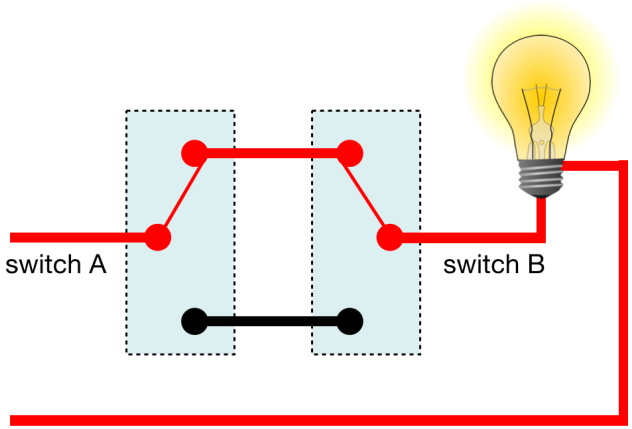
Experimental results

Main experiment

- The survey was conducted using Amazon's Mechanical Turk.
- Each participant saw the text and figure below.

The context

Imagine a long hallway with a light in the middle and with two switches, one at each end. One switch is called switch A and the other one is called switch B. As the following wiring diagram shows, the light is on whenever both switches are in the same position (both up or both down); otherwise, the light is off. Right now, switch A and switch B are both up, and the light is on. But things could be different. . .



Main experiment

Each participant saw a target item and the filler, in random order,
and was asked to judge these sentences as 'true', 'false', or 'indeterminate'.

Targets

(1) If switch A was down, the light would be off.

(2) If switch B was down, the light would be off.

(3) If switch A or switch B was down, the light would be off.

(4) If switch A and switch B were not both up, the light would be off.

(5) If switch A and switch B were not both up, the light would be on.

Filler

(6) If switch A and switch B were both down, the light would be off.

Main experiment

- We rejected data from participants whose native language was not English, judged the filler incorrectly or took the survey more than once.
- The remaining 1425 responses as summarized in the following table.

Table 3: Results of the main experiment

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\overline{A} >$ OFF	256	169	66.02%	6	2.34%	81	31.64%
$\overline{B} >$ OFF	235	153	65.11%	7	2.98%	75	31.91%
$\overline{A} \vee \overline{B} >$ OFF	362	251	69.33%	14	3.87%	97	26.80%
$\neg(A \wedge B) >$ OFF	372	82	22.04%	136	36.56%	154	41.40%
$\neg(A \wedge B) >$ ON	200	43	21.50%	63	31.50%	94	47.00%

Pre-test I

- Our aim was to test whether the antecedents of (3) and (4) are indeed perceived as truth-conditionally equivalent.

  (7) Switch A or switch B is down.

  (8) Switch A and switch B are not both up.

- More specifically, we wanted to check if (7) is interpreted exclusively.
- We presented participants with the picture below, and asked them to judge these sentences as 'true', 'false', or 'indeterminate'.

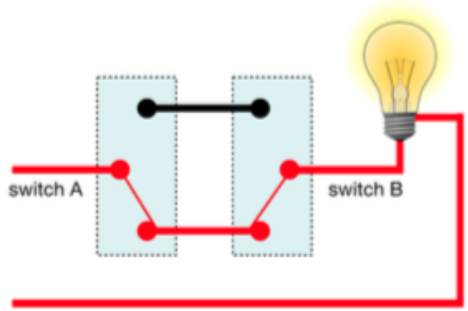
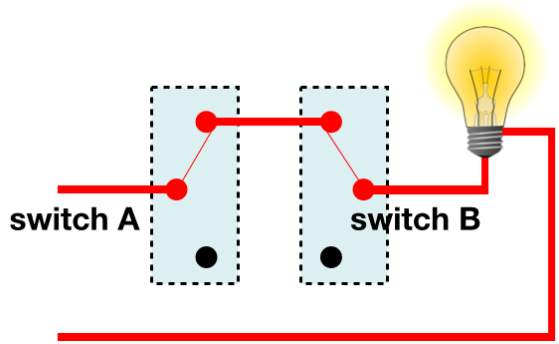


Table 1: Results of Pretest I

Sentence	Number	True	(%)	False	(%)	Indeterminate	(%)
$\overline{A} \vee \overline{B}$	145	118	81.38%	23	15.86%	4	2.76%
$\neg(A \wedge B)$	130	118	90.77%	11	8.46%	1	0.77%

Post-hoc test I

- We wanted to make sure that (4) are (5) are not rejected based on context-independent reasons (e.g., excessive processing load).
- For this, we tested our sentences in a different scenario: the light is on if and only if the switches are both up.



Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\overline{A}$ > OFF	52	41	78.85%	5	9.61%	6	11.54%
$\overline{B}$ > OFF	68	60	88.24%	5	7.35%	3	4.41%
$\overline{A} \vee \overline{B}$ > OFF	110	104	94.55%	1	0.91%	5	4.54%
$\neg(A \wedge B)$ > OFF	116	99	85.34%	9	7.76%	8	6.90%
$\neg(A \wedge B)$ > ON	103	19	18.45%	79	76.70%	5	4.85%

Post-hoc test II

- We wanted to make sure that our assumption of treating *down* as equivalent to *not up* was innocent.
- For this, we tested versions of our sentences where *down* is replaced by *not up*, in the setting of our main experiment.

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\neg A >$ OFF	36	27	75.00%	1	2.78%	8	22.22%
$\neg B >$ OFF	43	28	65.12%	7	16.28%	8	18.60%
$\neg A \vee \neg B >$ OFF	80	48	60.00%	16	20.00%	16	20.00%
$\neg(A \wedge B) >$ OFF	372	82	22.04%	136	36.56%	154	41.40%
$\neg(A \wedge B) >$ ON	200	43	21.50%	63	31.50%	94	47.00%

Appendix D

Premise semantics

Premise semantics (Veltman 76, Kratzer 81)

- Worlds are equipped with a set of true propositions, the premisses.
- To evaluate a counterfactual, we proceed as follows:
  1. add to the antecedent as many premisses as consistency permits;
  2. check if, for each way of doing so, the consequent follows.

Premise semantics (Veltman 76, Kratzer 81)

- Worlds are equipped with a set of true propositions, the premisses.
- To evaluate a counterfactual, we proceed as follows:
  1. add to the antecedent as many premisses as consistency permits;
  2. check if, for each way of doing so, the consequent follows.
- Adding as many premises as consistency permits is an implementation of the minimal change principle.
- Indeed premise semantics is translatable to (general) ordering semantics.
- Therefore, it cannot accommodate our data either.

Appendix E

Inquisitive lifting

Inquisitive lifting

- We implement this idea by means of a recipe called inquisitive lifting.
- Let f embody the truth-conditional account to be lifted:

  $s \models \varphi > \psi \iff \forall a \in \mathrm{Alt}(\varphi)\ \exists b \in \mathrm{Alt}(\psi)$ such that $s \subseteq f(a, b)$

- In the absence of inquisitiveness, the lifted account coincides with the base:

$$\mathrm{Alt}(\varphi > \psi) = \{\ f(|\varphi|, |\psi|)\ \}$$

- This implements Alonso-Ovalle's idea, in particular:

$$\begin{aligned} \mathrm{Alt}(\overline{A} \vee \overline{B} > \mathit{Off}) &= \{ f(|\overline{A}|, |\mathit{Off}|) \cap f(|\overline{B}|, |\mathit{Off}|)\ \} \\ &= \mathrm{Alt}((\overline{A} > \mathit{Off}) \wedge (\overline{B} > \mathit{Off})) \end{aligned}$$

Inquisitive lifting vs. Alonso-Ovalle's own account:

- builds on inquisitive semantics – we can draw on the inquisitive theory of propositional connectives;
- is modular with respect to the base account of counterfactuals, can be combined with a (fairly) arbitrary theory.