Antecedents of counterfactuals violate de Morgan's law

Lucas Champollion
champollion@nyu.edu
Joint work with Ivano Ciardelli and Linmin Zhang

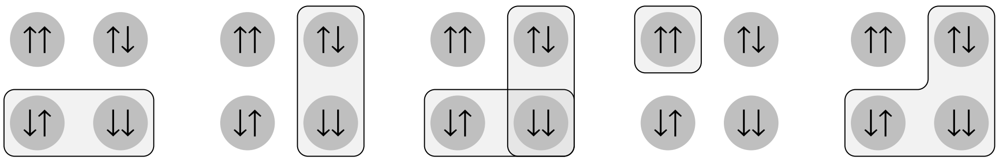
Fourth Workshop on Natural Language and Computer Science (NLCS 2016)
July 10, 2016, Columbia University, New York City, NY

1 Introduction

- There are two ways to look at the relation between meaning and truth conditions.
  - "To know the meaning of a sentence is to know its truth conditions" (Heim & Kratzer, 1998, p. 1)
  - This talk: Truth conditions do not completely determine the meaning of a sentence.
- Imagine two switches A and B: do the sentences in (1) have the same meaning?

(1) a. Switch A or switch B is down. $\neg A \vee \neg B$
b. Switch A and switch B are not both up. $\neg(A \wedge B)$

- They are equivalent in classical propositional logic (de Morgan's law), since they are true at the same possible worlds. In other systems, such as alternative semantics (Alonso-Ovalle, 2009) and inquisitive logic (Ciardelli et al., 2013), they are not.
- If $\neg A \vee \neg B$ and $\neg(A \wedge B)$ have the same meaning, then given compositionality, they should be interchangeable as antecedents of counterfactuals.



(a) $\neg A$ (b) $\neg B$ (c) $\neg A \vee \neg B$ (d) $A \wedge B$ (e) $\neg[A \wedge B]$

Figure 1: Inquisitive meanings of some simple sentences. ↑↑ represents a world where both switches are up, ↑↓ a world where A is up but B is down, etc. Only alternatives are shown.

- But we'll see that replacing one by the other can affect the counterfactual's truth value.
- By compositionality, we'll conclude that their meaning is different.
- But we'll also see that their truth conditions are the same.
- So we'll conclude that meaning is not completely determined by truth conditions.

2 Experiment

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 this 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 . . .

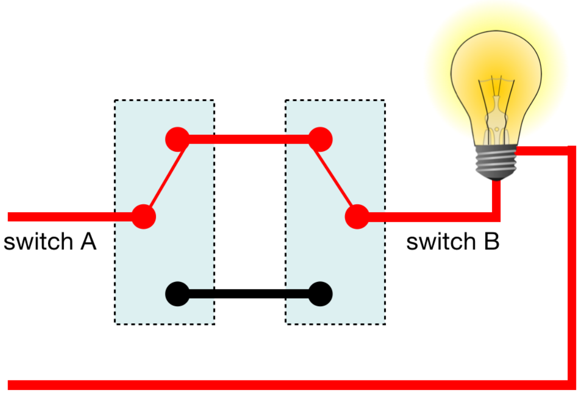


Figure 2: Lifschitz' switches. (Adapted from Lifschitz (1990) via Schulz (2007).)

- Test your intuitions (you may want to write them down; options are *true/false/indeterminate*):

(2) a. If switch A was down, the light would be off. _____
b. If switch A or switch B was down, the light would be off. _____
c. If switch A and switch B were both down, the light would be off. _____
d. If switch A and switch B were not both up, the light would be off. _____

- Figure adapted from multiway switches (c) Cburnett (`https://en.wikipedia.org/wiki/Multiway_switching#/media/File:3-way_switches_position_2.svg`) CC BY-SA 3.0.
- We used MTurk to collect truth value judgments in the context shown in Figure 2, including the picture.
- Participants could choose between *true*, *false* and *indeterminate*.
- We only showed people one of the sentences above at a time.
- We also showed everyone a filler sentence that we take to be uncontroversially false:

(3) If switch A and switch B were both down, the light would be off. $(\neg A \wedge \neg B) > \textsc{off}$

- All our subjects were based in the US. We eliminated subjects who participated more than once or didn't finish (less than 1%), didn't speak American English natively (about 4%), or didn't judge the filler sentence false (about 38%).

- The results are shown in Table 1.

Table 1: Results of the main experiment

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\neg A > \textsc{off}$	255	169	66.3%	6	2.4%	80	31.4%
$\neg B > \textsc{off}$	234	153	65.4%	7	3.0%	74	31.6%
$\neg A \vee \neg B > \textsc{off}$	346	242	69.9%	12	3.5%	92	26.6%
$\neg(A \wedge B) > \textsc{off}$	356	80	22.5%	129	36.2%	147	41.3%
$\neg(A \wedge B) > \textsc{on}$	200	43	21.5%	63	31.5%	94	47.0%

- Differences across the dashed line were highly significant ($p < 0.0001$ on a chi-square test).

- The similarity between $(\neg A) > \textsc{off}$, $(\neg B) > \textsc{off}$, and $(\neg A \vee \neg B) > \textsc{off}$ is striking. None of these conditions differed significantly.

3 Discussion

- We draw the following pretheoretical conclusions from the experiment:

- "If A or B then C" is interpreted in the same way as "If A then C, and if B then C" (this is known as *simplification of disjunctive antecedents*).

- When "If A or B then C" is interpreted, A and B are changed only one at a time. We consider two counterfactual scenarios: what if A but not B; what if B but not A.

- When "If (not both A and B) then C" is interpreted, we consider three counterfactual scenarios: what if A but not B; what if B but not A; what if neither A nor B.

- These results constrain our theories: Meaning is not completely determined by truth conditions, and de Morgan's law does not hold in the antecedents of counterfactuals.

- Propositional logic needs to be either supplemented or replaced in order to account for this.

4 What about exhaustive *or*?

- Do $\neg A \vee \neg B$ and $\neg(A \wedge B)$ really have identical truth conditions?

- Maybe $\neg A \vee \neg B$ is interpreted exclusively?

- We presented (1a) and (1b) in a context in which both switches are down
- We included the sentence *Switch A is up* as a filler item, and we excluded data from participants who failed to judge it false (this resulted in 16% of data being discarded).

Table 2: Results of Pretest I

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
(1a): $\neg A \vee \neg B$	145	118	81.4%	23	15.9%	4	2.8%
(1b): $\neg(A \wedge B)$	130	117	90.0%	11	8.5%	2	1.5%

- We conclude that $\neg A \vee \neg B$ and $\neg(A \wedge B)$ have the same truth conditions.

5 Quick primer on counterfactuals

(4) If kangaroos had no tails, they would topple over. (Lewis, 1973)

- Material implication at the actual world is not an option.
- Nor is material implication at all possible worlds:

(5) a. If kangaroos had no tails but used crutches, they would topple over.
b. $A > C \not\Rightarrow (A \wedge B) > C$

  - There will generally be some A-worlds which are very odd and remote
  - It should not matter whether such worlds are also C-worlds

- Stalnaker (1968), Lewis (1973): worlds are ordered based on how similar they are to the actual world
- "In any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over." (Lewis, 1973)
- Simplifying Lewis's proposal a bit, it says: $A > C$ is true iff C is true at every closest A-world.
- In case there is exactly one closest A-world, this amounts to: $A > C$ is true iff C is true at the closest A-world (Stalnaker, 1968)
- One problem: counterfactuals involving drastic changes should never be true (Fine, 1975)

(6) a. If Nixon had pressed the button, there would have been a nuclear holocaust.
b. The closest A-worlds will be worlds where the wire is cut etc.

- For this to work, a cut wire must mean a bigger difference than a nuclear explosion.
- Another problem (Schulz, 2007): similarity needs to conspire to simulate causal effects

(7) a. If switch A was down, the light would be off. *true*
b. If switch A was down, switch B would be down as well. *false*

- For this to work, a changed switch must mean a bigger difference than a changed light.
- Only the light causally depends on switch A. Switch B does not. But this fact is not relevant for Stalnaker and Lewis.
- Lewis (1979) responded by introducing a system of weights that makes the nuclear-holocaust world more similar to ours than the world in which the wire is cut.
- But our results are incompatible with the original Stalnaker/Lewis system on any similarity metric.

(8) a. Most of our speakers judged $(\neg A) > C$ and $(\neg B) > C$ true.
b. So the closest $\neg A$-worlds and the closest $\neg B$-worlds are C-worlds.
c. Most of our speakers judged $(\neg(A \wedge B)) > C$ false or indeterminate.
d. So not every closest $\neg(A \wedge B)$-world can be a C-world.
e. But every closest $\neg(A \wedge B)$-world is either a closest $\neg A$-world or a closest $\neg B$-world (or both), and by (8b) must be a C-world.

6 Explaining our findings

- We build on causal accounts (Pearl, 2000; Schulz, 2007; Kaufmann, 2013)
- Causal dependencies are described by a causal structure and a set of causal laws.
- The causal structure is a directed acyclic graph whose nodes (the *variables*) are partitions over the set of possible worlds.
- Variables are called *dependent* or *independent* based on whether they have ancestors.
- Figure 3 shows the causal network for the switches scenario. Its variables are “whether switch A is up” (independent), “whether switch B is up” (independent), and “whether the light is on” (dependent).

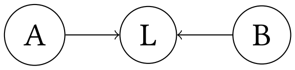


Figure 3: The causal structure for the switches scenario

- The basic recipe is as follows:

    1. start from the actual setting of the variables;
    2. change this minimally to make the antecedent true;
    3. propagate this change according to causal laws;
    4. check if the consequent is true.

- Problem for the interpretation of $\neg(A \wedge B) > \textsc{off}$:

- There are two minimal ways to make the antecedent of $\neg(A \wedge B) > \textsc{off}$ true: {A=up,B=down} and {A=down,B=up}.

- Each of these settings implies that the light is on.

- So $\neg(A \wedge B) > \textsc{off}$ is erroneously predicted true.

7 Complex antecedents

- To model our results we will need to modify the procedure somewhat.

7.1 Disjunctive antecedents

- We have seen that we got the same results for $\neg a \vee \neg b$ as we did for $\neg a$ and for $\neg b$.

- This suggests that a separate instance of the counterfactual is run on each disjunct.

- Suppose we have a counterfactual operator $A \blacktriangleright C$ that is defined only for noninquisitive meaning, e.g. that of Stalnaker, Lewis, or Pearl.

- We can lift it into inquisitive semantics in the following way (where $\textsc{Alt}(\varphi)$ returns the set of maximal subsets of φ):

(9) Inquisitive counterfactual conditional
$$[\![\varphi > \psi]\!] = \{p \,|\, \text{for all } A \in \textsc{Alt}(\varphi) \text{ there is a } C \in \textsc{Alt}(\psi) \text{ such that } p \subseteq A \blacktriangleright C\}$$

- In the cases of interest, ψ has only one alternative C (“the light is off”), but φ may have two alternatives (A_1 =“switch A is down”, A_2 = “switch B is down”).

- Then this definition amounts to separately testing $A_i \blacktriangleright C$ for each A_i.

7.2 Conjunctive antecedents

- The case of $\neg(a \wedge b)$ needs more work.

(10) If switch A and switch B were not both up, the light would be off. – *judged false*

- We will consider separately each of the three "ways of making A true", or what we will call the grounds for A: $\{a, \overline{b}\}, \{\overline{a}, b\}, \{\overline{a}, \overline{b}\}$. This forces attention to the case where both switches are down.
- Our updated recipe, in somewhat simplified form:
- $A > C$ is true given a causal network N iff for each ground G of A given the set of variables of N, the causal laws of N together with G entail C.
- Now (10) requires us to consider three grounds for the antecedent:

(11) a. Ground 1: Switch A is up and switch B is down.
b. Its only maximal causal premise set says that switch A is up.
c. Given the causal laws, the light is off.

(12) a. Ground 2: Switch A is down and switch B is up.
b. Its only maximal causal premise set says that switch B is up.
c. Given the causal laws, the light is off.

(13) a. Ground 3: Both switches are down.
b. Its only maximal causal premise set does not say anything.
c. Given the causal laws, the light is on.

- Since on ground 3 the light is on, (10) is predicted false, as desired.

8 Conclusion

- Our experiment shows that de Morgan's law fails to hold in the antecedents of counterfactuals.
- Although $\neg A \vee \neg B$ and $\neg(A \wedge B)$ have the same truth conditions, they still differ in meaning.
- Inquisitive semantics provides us with an intuitive explanation for this contrast.
- Our experiment also shows that $(\neg A) > C$ and $(\neg B) > C$ do not entail $(\neg(A \wedge B)) > C$, contra the unmodified Stalnaker/Lewis account.
- An antecedent of the form $\neg(A \wedge B)$ invites contemplation of three "grounds": A but not B, B but not A, and crucially, neither A nor B.
- An antecedent of the form $\neg A \vee \neg B$ only invites contemplation of the first two cases.
- Meaning is not completely determined by truth conditions.

References

Alonso-Ovalle, Luis. 2009. Counterfactuals, correlatives, and disjunction. *Linguistics and Philosophy* 32(2). 207–244. doi:10.1007/s10988-009-9059-0.

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2013. Inquisitive semantics: A new notion of meaning. *Language and Linguistics Compass* 7(9). 459–476. doi:10.1111/lnc3.12037.

Fine, Kit. 1975. Critical notice. *Mind* 84(335). 451–458. doi:10.1093/mind/LXXXIV.1.451.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Blackwell Publishing.

Kaufmann, Stefan. 2013. Causal premise semantics. *Cognitive Science* 37(6). 1136–1170. doi:10.1111/cogs.12063.

Lewis, David. 1973. *Counterfactuals*. Blackwell.

Lewis, David. 1979. Counterfactual dependence and time's arrow. *Noûs* 13(4). 455–476. doi:10.2307/2215339.

Lifschitz, Vladimir. 1990. Frames in the space of situations. *Artificial Intelligence* 46(3). 365–376. doi:10.1016/0004-3702(90)90021-q.

Pearl, Judea. 2000. *Causality: Models, reasoning, and inference*. Cambridge University Press. doi:10.1017/cbo9780511803161.

Schulz, Katrin. 2007. *Minimal models in semantics and pragmatics: Free choice, exhaustivity, and conditionals*: University of Amsterdam dissertation.

Stalnaker, Robert C. 1968. A theory of conditionals. In Nicholas Rescher (ed.), *Studies in logical theory*, 98–113. Blackwell. doi:10.1007/978-94-009-9117-0_2.