Holding on and letting go:
Facts, counterfactuals, and *'before'*

Cleo Condoravdi
PARC and Stanford University

Stefan Kaufmann
Northwestern University

Tenth Stanford Semantics Fest
March 14, 2009

Goals of this talk:

- Look into the relationship between counterfactuals and counterfactual *'before'* sentences more closely.
- Characterize the notion of likelihood underlying the felicity of non-veridical *'before'* sentences in the overall context of the analsyis of counterfactuals.

1 *'Before'*

What we are concerned with in this talk

- The interpretation of *'[before B]'* in *'[A [before B]]'*, where *'B'* is a clause.
- The *modal* dimension of the meaning of *'before'* and the relation between a sentence like (1a) and the corresponding counterfactual conditional (1b).

(1) a. The police defused the bomb before it exploded.
b. If the police hadn't defused the bomb, it would have exploded.

What we are not concerned with in this talk

- *'Before'* as a preposition with a DP complement.
- The fine details of its *temporal* semantics. For more on the latter, see Beaver and Condoravdi (2003) and references therein.

1.1 Modal implications of *'before'*-sentences

- *'A before B'* does not entail *'B'*:

(2) a. Mozart died before he finished the Requiem.
b. $\not\Rightarrow$Mozart finished the Requiem.

- The truth of *'A'* and the falsity of *'B'* are not sufficient for the truth/felicity of *'A before B'*. E.g. from Beaver and Condoravdi (2003):

(3) a. David ate lots of ketchup
b. David never made a clean sweep of all the gold medals in the Sydney Olympics.

c. ⇏David ate lots of ketchup before he made a clean sweep of all the gold medals in the Sydney Olympics.

- What exactly is implied about *'B'* depends on the *context* (Heinämäki (1972), Ogihara (1995), Beaver and Condoravdi (2003)). There are three possibilities, corresponding to the following "readings" of *'before'*:[1]

Veridical: *B* is implied to be true.

'A before B' has no relevant modal implications. The assertion is strictly about the temporal relation between *A* and *B*.

(4) The police checked out the bomb before it exploded.
The bomb exploded.

Non-committal: *B* is neither implied to be true nor implied to be false.

➽ *B* may be false, but if it is, it was at least reasonably probable.

(5) I left Trafalgar Square about half an hour ago and it started to look scary. The happy crowd that had made its way towards the square for some reason got cut off by the police and the Trade Union Congress march was prevented from getting in for their planned meeting. I gather that McDonalds has been trashed. Tourists seemed to be mingling curiously with the demonstrators in the square, but there was no violence at that time. Yet the police seem to be blocking off all the exits. I decided to leave before there was any trouble. [Google]
Trouble looked likely before I left and may or may not have come about.

Counterfactual: *B* is implied to be false.

➽ *B* was reasonably probable.

(6) The police defused the bomb before it exploded. And a good thing too, because they saved the lives of a trainful of commuters.
No explosion
If the police had not defused the bomb, it would have exploded.

➽ The modal implications we are interested in are most clearly observable in contexts which give rise to the "counterfactual" reading of *'before'*.

Analysis of non-veridical readings

In the analysis of Beaver & Condoravdi (2003) the time contributed by the temporal clause is specified in terms of an *earliest* operator applied to the denotations of the temporal clause. They attribute the modal dimension of *'before'* to the fact that when *earliest* is undefined in the world of evaluation it is relativized to a set of alternative worlds to the world of evaluation.

- Hypothetically "rerun history" to the reference time *t* of *'A'* so as to open up the possibility of *'not A'*
- Restrict this hypothetical exploration to those courses of events that were reasonably probably at *t*

[1] Beaver and Condoravdi (2003) show that this context dependence can be accounted for with a uniform analysis of *'before'* that does not postulate an ambiguity.

- Check the consistency of *'B'* with these courses of events
- Consistency of *'B'* with these courses of events implies the definedess of *earliest* at the relevant points of evaluation. The definedness of *earliest* is calculated point-wise.

➽ Facts at times later than the reference time of *'A'* are given up in the process of "rerunning history."

Veridical and non-veridical implications as contextual entailments

- In order to account for the fact that *before/after*-clauses can be informative and not necessarily pragmatically presupposed Beaver & Condoravdi (2003) do not make the semantic definedness condition of *earliest* a pragmatic presupposition of the connectives.
- Instead they assume that in updating a context with a sentence, the truth conditions are checked pointwise for each world in the context and worlds in which the sentence is true are retained, while worlds in which it is false or has no truth value are discarded.
- This allows them to characterize veridical and non-veridical readings as contextual entailments, that is entailments of particular types of contexts when updated with a *'before/after'* sentence.
- Let $rph_c(w, t)$ designate the reasonably probable, given the contextual assumptions in c, amongst a world's w historical alternatives at a given time t.
- In order for *'A before B'* sentence to be felicitous relative to context c it has to lead to a consistent update. The update will be consistent only if
  - there is some $w \in c$ such that the *'B'* is true in w, OR
  - there is some $w \in c$ such that $rph_c(w, t)$ is compatible with *'B'*

- "For the counterfactual reading, the input context has to entail that A's occurrence makes B's later occurrence impossible and that prior to A's occurrence there was a process that made B's occurrence at least reasonably probable." (Beaver & Condoravdi 2003:51)
- An intuitively plausible and often observed parallelism links the truth of "counterfactual" *'A before B'* with that of the corresponding counterfactual conditional *'If had not been A, would have been B'*.

(7) a. The police defused the bomb before it exploded.
b. If the police had not defused the bomb, it would have exploded.

Questions:

- What does "reasonably probable" mean?
- What is the relationship between (7a) and (7b)?
- Is "reasonable probability" involved in both?

2 Counterfactual conditionals

A first hypothesis based on the preceding section:

Counterfactual *'A before B'* and the counterfactual conditional *'not-A □→B'* go hand in hand.

I.e., whatever makes one true (false) also makes the other true (false).

But this is not correct: Prior likelihood does not play the same role for counterfactual conditionals as it does for counterfactual *'before'*-sentences.

2.1 Likelihood and counterfactuals

Consider (8), and suppose the speaker got off a bus which subsequently was involved in an accident in which everyone on board got injured.

(8) a. ?I got off the bus before I got injured.
b. If I hadn't gotten off the bus, I would have been injured.

- (8a) is infelicitous unless the speaker expected the accident at the time she got off.

➽ *'A before B'* implies that *'B'* is/was *likely/predictable* at the time of *'A'* (as noted above).

- (9b) does not carry this requirement.

➽ *'not-A □→B'* may be true even if *'B'* was not likely at the relevant time.

Problem for the first hypothesis:

In cases like (8), while counterfactual *'A before B'* implies that *'B'* was likely, the counterfactual conditional *'A □→B'* does not.

Main difference: The truth of the counterfactual depends on "posterior" developments (i.e., ones after the reference time of *'A'* whereas the *'before'*-sentence does not.

Q: But maybe what matters is that the plane crash was likely, not from the speaker's subjective perspective, but somehow "objectively," i.e., with respect to facts that held at the time unbeknownst to the speaker?

A: Not likely. Very similar effects are observed with posterior facts that happen purely by chance, and therefore *cannot* have been predictable.

Here's another example from the philosophical literature on the counterfactuals.

Time 1: A fair coin is about to be tossed, and you are asked to make a bet. Since the coin is fair, (9) is false.[2]

[2]This example was attributed to Sidney Morgenbesser by Slote (1978) and was also discussed by Bennett (1984, 2003); Edgington (1995); Barker (1998); Kaufmann (2005) and others.

(9) If I bet on tails, I will lose. *[was false then]*

Time 2: You bet on heads. The coin is tossed and comes up heads. You win. Now (10) is true:

(10) If I had bet on tails, I would have lost. *[is true now]*

The counterfactual (10) is now true. But there was no earlier time at which the indicative (9) was true.

Examples like these suggest that the *'before'*-sentence is more closely related to the past indicative than to the present counterfactual.

- *'A before B'*:
  B was predictable/likely at the time of *A*;
  'If not-A then B' was true at the time.
- *'If had been not-A, would have been B'*:
  B may have been very unlikely at the time of *A*;
  Does not imply that *'if not-A then B'* was true at the time.

2.2 Consequences for the analysis of counterfactuals: Holding on

Naïve strategy

To evaluate a counterfactual *'A* $\Box\!\!\rightarrow$ *B'*, where *'A'* is false and the reference time of *'A'* precedes that of *'B'* (Lewis, 1979):

- Hypothetically "rerun history" from a past time at which *'A'* was still a historical alternative (i.e., a "live possibility")
- Restrict this hypothetical exploration to those courses of events in which *'A'* does occur.
- The counterfactual is true iff all of those courses of events lead to *'B'*.
- Modify the preceding clause if some weaker notion of necessity is called for (e.g., Kratzer's "human necessity" relative to an ordering source).

➽ Facts at times later than the reference time of *'A'* are given up in the process of "rerunning history."

Proper strategy

Make reference to a relation of *causal (in)dependence*:

- *Bus example:* Whether the accident occurs does not causally depend on my presence or absence on the bus.

  ➽ The accident was unlikely, but it did occur. This fact is held constant.
- *Coin example:* Whether the coin lands heads or tails does not causally depend on the bet.

  ➽ The coin's coming up heads was not predictable, but it did occur. This fact is held constant.

Counterfactuals *vs.* *'before'*:

- Posterior facts that are causally independent of *'A'* are held on to in the interpretation of the counterfactual *'$A \mathrel{\Box\!\!\rightarrow} B$'*.
- Such posterior facts are given up in the interpretation of *'A before B'*.

3 *'Before'* again

A second hypothesis:

> What matters for the question whether *'A'* happened "before *B*" or not is whether *'B'* was likely at the time.

I.e., ignore everything that came after *'A'* and consider only the state of the world at the reference time of *'A'*.

But this is still not enough: Certain facts about the state of the world at the reference time of *'A'* must be given up as well.

3.1 Likelihood and *'before'*

In (11) the use of *before* is felicitous even though the meteor was unlikely to reach the ground; the counterfactual is true as well.

(11) a. *[As meteors usually do / As scientists expected]*
The meteor burned up before it hit the ground.
b. If the meteor had not burned up, it would have hit the ground.

The problem is that (11) explicitly denies that the meteor was likely to hit the ground.

4 Outline of a formal analysis

Goals of this section:

- Start with an off-the-shelf semantic framework for the interpretation of counterfactuals (in our case, Kratzer's (1981) premise semantics)
- Add a priniciple (and also off-the-shelf) way to deal with (many) counterfactual conditionals
- Against this background, propose a principled way to deal with *'before'*-clauses which highlights the contrast.

4.1 Premise semantics, similarity, and "human necessity"

Main points (accepted here largely without argument):

- Counterfactuals and *'before'*-sentences alike call for a modal interpretation in terms of what's likely, necessary, etc.
- The standard interpretation of necessity as truth at *all* A-worlds is too strong for our purposes.

  For instance, if *'B'* was likely but did not happen, its likelihood cannot have been truth at all possible courses of events. Similar arguments abound for conditionals.
- "Human necessity" is a weaker notion of (roughly speaking) truth at all "relevant" worlds, e.g., all worlds following the "normal" course of events.

➽ Goal: Encode the difference between counterfactuals and *'before'*-clauses in terms of different conditions on the "relevant" worlds to consider.

Some formal background

Let us adopt the formal tool of an *ordering source* imposing a relation of *relative likelihood* on the possible worlds (or states of affairs). For details, see Kratzer (1981).

- A *proposition* is a set of possible worlds.
- A *modal base* is a function f from worlds to sets of propositions. For concreteness, let $f(w)$ be the set of those propositions that are contextually taken for granted. $\bigcap f(w)$ is the set of those worlds which, are compatible with everything that is taken for granted.
- An *ordering source* is a function o from worlds to sets of propositions. Intuitively, in our case, $o(w)$ is the set of those propositions that are "normally" true (or more likely to be true than not) given the facts at w.
- At each world w of evaluation, $o(w)$ induces a relation of *relative likelihood* between worlds:

  (12) $$u \leq_{o(w)} v \iff \{p | p \in o(w) \wedge v \in p\} \subseteq \{p | p \in o(w) \wedge u \in p\}$$

- This offers a new notion of "human" necessity that is weaker than "simple" necessity. Simplifying, p is a human necessity at w relative to f and g iff those worlds in the modal base that are most likely under $g(w)$ are p-worlds.

  (13) $$\forall u \in \bigcap f(w) \exists v \in \bigcap f(w)[v \leq_{o(w)} u \wedge \forall z \in \bigcap f(w)[z \leq_{o(w)} v \rightarrow z \in p]]$$

4.2 Counterfactuals

Main idea

- In the evaluation of '$A \boxrightarrow C$' at world w, the "relevant" worlds are those worlds that satisfy two conditions:
  - A is true at them; and
  - they are at least as similar to w as any other A-worlds.
- Roughly: '$A \boxrightarrow C$' is true at w iff among all A-worlds, those that are most similar to w are C-worlds (Stalnaker, 1968; Lewis, 1973).[3]

- $f(w) = \emptyset$ for all w. Thus $\bigcap f(w) = W$, the set of all worlds.
- Since the relevant ranking is based on *relative similarity to w*, $o(w)$ contains propositions that are true at w.

Q: Which true propositions?

How similarity is determined

Much recent work in philosophy, psychology and artificial intelligence has explored the role of *causal (in)dependencies* in the interpretation of counterfactuals (Glymour et al., 1993; Pearl, 2000; Hiddleston, 2005; Schulz, 2007, among many others)
To illustrate:

(14) *[I bet on heads, the coin came up heads, I won]*
If I had bet on tails, I would have lost.

- True if the speaker's bet has no (causal) influence on the outcome of the toss.
- False if it does (e.g., if a different fair coin is used depending on the bet).

(15) *[I got off the bus, the bus had an accident, I was uninjured]*
If I hadn't gotten off the bus, I would have been injured.

- True if the speaker's presence or absence on the bus has no (causal) influence on the accident.
- False if it does.

➽ If a relation of *causal dependence* (or causal influence) is given as basic, then the interpretation of counterfactuals can be explained in terms of it.

Lots of people nowadays think that this is a more fruitful approach than Lewis's attempt to go the other way and reduce causality to counterfactual (in)dependence.

[3] The above presupposes that there is a set of most similar A-worlds. In fact, neither Stalnaker nor Lewis make exactly that assumption. Stalnaker: iff *the* most similar A-world (to w) is a C-world. Lewis: iff any sequence of A-worlds whose similarity to w grows monotonically, ends in a sequence of C-worlds.

Some formal background

- Causal dependence is a relation between *variables* (in the statistical sense).
- The relation is transitive and asymmetric (often given as a *directed acyclic graph*).
- In our framework, a "variable" is a *partition* of the set of all worlds (like a question denotation).
- For simplicity, assume that these variables are binary (i.e., yes-no questions): Each corresponds to a proposition and its negation.
- At each world, all relevant variables take some value or other.

  (16) Let X_w be the cell in partition X that contains world w.

- Impose the following constraint on ordering sources for counterfactuals:

  (17) For all X, Y such that $X \rightarrow Y$ (i.e., X causally influences Y), and all w, $o_{cf}(w)$ contains only $Y_w \cap X_w$ (not Y_w).

  Intuitively, the partition Y does not really play a role in the intepretation; only its refinement by X does.

Upshot:

- Suppose X, Y are both true at w.
- Without causal dependence: Among the worlds at which X is false, those at which Y is true are more similar to w than those at wich Y is false.
- With causal dependence $X \rightarrow Y$: Among the worlds at which X is false, Y does not count towards similarity to w.

➽ We derive the following asymmetry:

(18) a. If Y were false, X would (still) be true.
b. If X were false, Y might not be true.

4.2.1 Back to our examples

(19) If I had stayed on the bus, I would have been injured.

- Suppose my presence or absence on the bus had no causal bearing on whether there would be an accident.

➽ At worlds w at which the accident happened, $o(w)$ contains the proposition that the accident happened.

This proposition is consistent with the antecedent of (19), hence (19) is true.

- Suppose my whereabouts did have an influence on the accident.

➽ At worlds w at which the accident happened, $o(w)$ does not contain the proposition that the accident happened, but only the proposition that I got off and the accident happened.

This proposition is not consistent with the antecedent of (19), hence (19) is false.

4.3 *'Before'*-clauses

(20) If I had stayed on the bus, I would have been injured.

(21) The meteor burned up before it hit the ground.

Main intuition

- What mattered for counterfactuals was *similarity to w*.
  Since the accident happened, accident-worlds are more likely than others.
- What matters for *'before'*-clauses is not similarity.
- *Nor* is it (overall) likelihood, however!
  (21) may be true even if the meteor never had a chance to hit the ground.

➽
  - Not only do we not hold on to the fact that the meteor burned up;
  - We even disregard the question of whether it would burn up nor not.

Implementation

- Assume again that we are given information on causal independence.
- Consider some relevant fact, such as the fact that the meteor was falling (along a certain trajectory, with a certain velocity) at time t.
- Worlds at which it burns up in the atmosphere are *more likely* than worlds at which it doesn't.
- This follows from a "stereotypical" ordering source of the usual kind:

  (22) For all X, Y such that $X \nrightarrow Y$, and all w, $o_l(w)$ contains Y_w or its negation, whichever is more "normal" at w.

- BUT the atmosphere presents an external intervention on the meteor's fall that we want to ignore.
- Impose the following constraint on ordering sources for *'before'* in sentences *'A before B'*, supposing there is an X with a cell X_B in which *'B'* is a human necessity:[4]

  (23) For all Y such that $X \nrightarrow Y$, and all $w \in X_B$, $o_b(w)$ contains neither Y_w nor its negation.

- So, for instance, condition (23) will remove any partitions Y that are causally independent of X, that is, it will make ordering sources $o_l(w)$ for $w \in X_B$ irrelevant.
- Consequently, even though in both (24a,b) the *'before'*-clause is not only false but unlikely as well, (24a) is infelicitous (there is no X relative to which David's making a clean sweep of all the gold medals in the Sydney Olympics is a human necessity) while (24b) is felicitous.

[4] We leave as an open question in this talk what the ordering source is relative to which *'B'* is a human necessity .

(24) a. David ate lots of ketchup before he made a clean sweep of all the gold medals in the Sydney Olympics.
b. Mozart, fatally ill, died before he finished the Requiem.

References

Barker, S. J. 1998. Predetermination and tense probabilism. *Analysis*, 58:290–296.

Beaver, D. and C. Condoravdi. 2003. A uniform analysis of *before* and *after*. In *Proceedings of SALT XIII*, pages 37–54.

Bennett, J. 1984. Counterfactuals and temporal direction. *The Philosophical Review*, 93(1):57–91.

Bennett, J. 2003. *A Philosophical Guide to Conditionals*. Oxford University Press.

Edgington, D. 1995. On conditionals. *Mind*, 104(414):235–329.

Glymour, C., P. Spirtes, and R. Scheines. 1993. *Causation, Prediction, and Search*. Springer.

Heinämäki, O. 1972. Before. In *Proceedings of CLS 8*, pages 139–151.

Hiddleston, E. 2005. A causal theory of counterfactuals. *Noûs*, 39(4):632–657.

Kaufmann, S. 2005. Conditional predictions: A probabilistic account. *Linguistics and Philosophy*, 28(2): 181–231.

Kratzer, A. 1981. The notional category of modality. In Eikmeyer, J. and H. Riesner, editors, *Words, Worlds, and Contexts*, pages 38–74. Walter de Gruyter.

Lewis, D. 1973. *Counterfactuals*. Harvard University Press.

Lewis, D. 1979. Counterfactual dependence and time's arrow. *Noûs*, 13:455–476.

Ogihara, T. 1995. Non-factual *before* and adverbs of quantification. In *Proceedings of SALT V*, pages 273–291. CLC Publications, Cornell University.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Schulz, K. 2007. *Minimal Models in Semantics and Pragmatics: Free Choice, Exhaustivity, and Conditionals*. PhD thesis, University of Amsterdam.

Slote, M. A. 1978. Time in counterfactuals. *The Philosophical Review*, 87(1):3–27.

Stalnaker, R. 1968. A theory of conditionals. In *Studies in Logical Theory, American Philosophical Quarterly, Monograph: 2*, pages 98–112. Blackwell.