Subjunctive conditionals

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Subjunctive Conditionals

1 Introduction

Conditional sentences, canonically of the form “if p, q”, whisk us away to a scenario described by their antecedent and then make a claim about it in their consequent. There are two main kinds of conditionals, as illustrated in the well-known Oswald/Kennedy minimal pair (due to Adams 1970):

(1) If Oswald didn’t kill Kennedy, someone else did.
(2) If Oswald hadn’t killed Kennedy, someone else would have.

Clearly, the two conditionals differ in meaning. The conditional in (1) signals that it is an open possibility that Oswald didn’t kill Kennedy and will be judged true by anyone who knows that Kennedy was in fact assassinated. The conditional in (2), in contrast, signals that it is taken for granted that Oswald did in fact kill Kennedy and makes the somewhat dubious claim that Kennedy’s assassination was inevitable, perhaps based on a vast conspiracy. The same difference in meaning can be illustrated with a similar pair (due to Bennett or Stalnaker?), where it is even harder to hear the second conditional as making a plausible claim:

(3) If Shakespeare didn’t write Hamlet, someone else did.
(4) If Shakespeare hadn’t written Hamlet, someone else would have.

This chapter will be concerned with the meaning of conditionals of the second kind.

Conditionals of the first kind are usually called “indicative” conditionals, while conditionals of the second kind are called “subjunctive” or “counterfactual” conditionals. The “indicative” vs. “subjunctive” terminology suggests that the distinction is based in grammatical mood, while the term “counterfactual” suggests that the second kind deals with a contrary-to-fact assumption. Neither terminology is entirely accurate.

It is clear that the outward difference between conditionals of the two kinds lies in their tense/mood/aspect syntax, but it is not reliably a difference in indicative vs. subjunctive mood properly speaking. In languages that have a past subjunctive (such as German), antecedents of conditionals of the second kind do indeed appear in the subjunctive mood but if a language does not have a past subjunctive, some other form is used. English uses an indicative antecedent with an additional...
layer of past tense morphology (“if Oswald didn’t kill Kennedy” becomes “if Oswald hadn’t killed Kennedy”). It is that additional layer of past tense morphology (which doesn’t obviously contribute an actual past meaning) that quite reliably signals conditionals of the second kind. So, instead of calling them subjunctive conditionals, one might call them “additional past” conditionals.

While our initial examples did appear to signal that their antecedent was false, conditionals of the second kind do not always carry the suggestion of counterfactuality:

(5) If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show. (Anderson 1951)

A doctor who utters (5) might be prepared to use it as part of an argument that the antecedent is in fact true, so the conditional could not be conveying counterfactuality as part of its meaning.

Conversely, there are some formally indicative conditionals that express counterfactuality:

(6) If he has solved this problem, I’m the Queen of England.

(7) If Messi waits just a second longer, he scores on that play.

[Indicative counterfactuals of the kind in (7) are common in sportscast play-by-play commentary. They have not (yet) been studied in the semantic literature.]

Even though inaccurate, the terms indicative/subjunctive-counterfactual are so entrenched that it is presumably futile to try to reform usage. The association of conditionals of the second kind with counterfactuality is somewhat tenuous. In comparison, the connection of the distinction to tense/mood/aspect morphology is undeniable even if the morphology in conditionals of the second kind isn’t strictly speaking subjunctive in many languages. So, the indicative vs. subjunctive terminology is the slightly less inappropriate one and will be the one we adopt here.

### 2 The meaning of subjunctives

Conditionals whisk us away to a scenario where the antecedent holds. The most influential semantic theories about subjunctive conditionals, in particular the pioneering works of Stalnaker and Lewis, treat them as making claims about “possible worlds” that may be quite different from the actual world. Nevertheless, when one looks at examples like the Oswald/Kennedy conditionals, one sees that even the counterfactual version makes palpable claims about reality: “if Oswald hadn’t killed Kennedy, someone else would have” makes the claim that in actuality, Kennedy’s assassination was inevitable, perhaps because there was a vast conspiracy in place or because one thinks that Kennedy’s actions and policies inexorably provoked an assassination. How does talk about other possible worlds connect to the actual world?
All of the possible worlds semantic approaches answer this concern by closely
tying the identification of the relevant set of possible worlds to facts in the actual
world. Stalnaker 1968 grounds the basic idea in the same “Ramsey Test” that
has inspired accounts of indicative conditionals. For Stalnaker, the test describes
how we evaluate conditionals: we hypothetically add the antecedent to our in-
formation state and assess the consequent in the resulting state. While canonical
indicatives have antecedents that are compatible with the relevant information
state, canonical subjunctives involve antecedents that contradict some prior in-
formation. Thus, the adding of the antecedent will necessitate some revisions of
the information state. (It should be noted that the proper analysis of Anderson-
type examples, as in (5), in this picture is not entirely obvious. See von Fintel 1998
for some discussion.)

Correlated with this psychological process of assessing a conditional, Stalnaker
proposes a truth-conditional semantics that starts from the actual world, considers
the antecedent, and looks for worlds that differ minimally from the actual world
while making the antecedent true. It is in those worlds that the consequent is then
evaluated. So, in the Oswald/Kennedy case, we keep fixed all actual facts that are
not strictly tied to Oswald killing Kennedy. The subjunctive conditional claiming
that if Oswald hadn’t killed Kennedy, someone else would have, thus amounts to
saying that there were facts in the actual world that would have led to Kennedy’s
assassination one way or the other.

Tying the semantics of conditionals to the notion of comparative similarity
between possible worlds may seem like it doesn’t help us nail down the meaning of
conditionals all that much, since similarity obviously is quite a vague and context-
dependent notion. In a brilliant and influential move, Lewis (1973: 91ff) defended
the use of this notion by pointing out that conditionals themselves are inherently
vague and context dependent. Therefore a proper analysis of conditionals needs
to correctly reflect their vagueness and context dependence. He argued that the
comparative similarity relation between possible worlds is just the right tool to do
so.

As an illustration of the vagaries of judging similarity and difference between
worlds, consider Quine’s famous pair of conditionals (1960: 221, NB: he actually
calls the pair “Goodman’s, nearly enough”):

(8) If Caesar were in command, he would use the atom bomb.
(9) If Caesar were in command, he would use catapults.

Quine wrote: “we feign belief in the antecedent and see how convincing we then
find the consequent. What traits of the real world to suppose preserved in the
feigned world of the contrary-to-fact antecedent can be guessed only from a symp-
thetic sense of the fabulist’s likely purpose in spinning his fable.” His “pre-
serving of traits of the real world in the feigned world of the antecedent” cor-
responds quite directly with the decision about which properties of worlds carry
more weight in the judgment of similarity between worlds. In (8), we seem to keep
constant Caesar’s ruthlessness, while in (9) we (also) care about the technologies
that were actually at his disposal in his time.
The similarity/difference-based semantics remains quite a schematic framework, to be filled in with contextually relevant considerations for assessing similarities and differences. In later work, in response to several worries about the account, Lewis and others suggested constraints on the kind of similarity relation underlying conditionals. This debate continues rather unabated.

A variant of the semantics based on a similarity ordering is given by “premise semantics”, inspired by Goodman (1947) and Rescher (1964) and developed in rival forms by Kratzer (1977, 1979, 1981) and Veltman (1976). Here, we start with a description of the actual world (a set of sentences or propositions) and revise it to make it accommodate the addition of the antecedent. Then, we check whether the consequent follows from the resulting set of premises. The revision process in this account intuitively parallels the determination of the antecedent worlds most similar to the actual world in the Stalnaker/Lewis framework. And in fact, Lewis (1981) showed that technically the two approaches are intertranslatable. More recent work in the premise semantic tradition includes Kratzer 1989, Veltman 2003, Kanazawa et al. 2005, Kratzer 2005.

3 Nonmonotonicity

There is an austere way of spelling out a possible worlds analysis of conditionals according to which the context supplies us with a set of relevant possible worlds, those that in all relevant respects are similar enough to the actual world, and that the conditional then makes a claim about all of the antecedent worlds in that set. This is known as “strict implication”. In contrast, the analysis proposed by Stalnaker and Lewis is a “variably strict” analysis (terminology due to Lewis), in which we start from the antecedent and identify among the antecedent worlds those worlds that are as similar as possible to the actual world. The variably strict semantics in distinction to the strict implication analysis predicts a variety of nonmonotonic behaviors, and Stalnaker and Lewis argued that those predictions are borne out.

Under a strict implication analysis, the pattern known as Strengthening the Antecedent, is predicted to be valid:

(10)  Strengthening the Antecedent
if p, q ⇒ iif p&r, q

If all (contextually relevant) p-worlds are q-worlds, then a fortiori all p&q-worlds, a subset of the p-worlds, have to be q-worlds. This pattern becomes invalid in the Stalnaker/Lewis analyses. If the p-worlds that are most similar to the evaluation world are all q-worlds, that does not necessitate that the most similar p&q-worlds are also all q-worlds. Lewis (1973) gives a humorous example:

(11)  Failure of Strengthening the Antecedent
If kangaroos had no tails, they would topple over. ∴
If kangaroos had no tails but used crutches, they would topple over.
The variably strict analyses explain why the inference in (11) fails: the worlds where kangaroos have no tails but that are otherwise as similar as possible to the evaluation world are not worlds where kangaroos use crutches, so the first conditional does not connect logically to the second conditional.

Other patterns that are expected to be valid under the strict implication analysis but arguably aren’t are Hypothetical Syllogism and Contraposition:

\[(12) \text{ Failure of the Hypothetical Syllogism (Transitivity)}
\]

If Hoover had been a Communist, he would have been a traitor.
If Hoover had been born in Russia, he would have been a Communist.
\(\not\Rightarrow\) If Hoover had been born in Russia, he would have been a traitor.

\[(13) \text{ Failure of Contraposition}
\]

If it had rained, there wouldn’t have been a terrific cloudburst. \(\not\Rightarrow\)
If there had been a terrific cloudburst, it wouldn’t have rained.

[The example in (12) is from Stalnaker 1968, and the example in (13) is a subjunctive version of an example from Adams 1975.]

The non-monotonic analyses predict both of these invalidities correctly.

4 The dynamic strict analysis

Let us return to the apparent failure of Strengthening the Antecedent. Lewis tried to forestall the idea that what is treated as semantic non-monotonicity in his account could actually be explained in a strict implication account by saying that the contextually relevant set of worlds that the conditional quantifies over is easily shifted in a sequence of sentences. He argued that this move would not be able to explain the well-formedness of what became known as Sobel Sequences:

\[(14) \text{ If the USA threw its weapons into the sea tomorrow, there would be war; but if all the nuclear powers threw their weapons into the sea tomorrow, there would be peace.}
\]

Lewis deliberately put this example “in the form of a single run-on sentence, with the counterfactuals of different stages conjoined by semicolons and \textit{but},” suggesting that it would be a “defeatist” move to say that in such a tight sequence the context could shift in response to the introduction of a new antecedent clause.

Defeatist or not, based on an observation by Heim (MIT class handout), von Fintel (2001) develops such an account. Heim had noted that Lewis’ Sobel Sequence cannot be reversed:

\[(15) \text{ ?If all the nuclear powers threw their weapons into the sea tomorrow, there would be peace; but if the USA threw its weapons into the sea tomorrow, there would be war.}
\]

This is unexpected from the point of view of a semantically non-monotonic analysis. In von Fintel’s paper, a dynamic strict analysis is developed in which the antecedent has the potential to expand the “modal horizon”, the set of contextually
relevant possible worlds which the conditional then ranges over. It is shown that if the expansion of the modal horizon is governed by the same similarity ordering used in the Stalnaker/Lewis systems, the analysis replicates the truth-conditions of those systems for isolated or discourse-initial conditionals. The context shifts become only relevant in sequences of conditionals and then create the appearance of semantic non-monotonicity. One crucial argument von Fintel gives for his account is that negative polarity items are licensed in the antecedent of conditionals and that therefore we would prefer a monotonic analysis. It turns out, however, that only a very special notion of monotonicity (dubbed Strawson Downward Entailingness) holds for von Fintel’s conditionals: these conditionals are downward monotone in their antecedent only under the assumption that the initial context is such that the modal horizon is already large enough to be unaffected by any of the conditionals in the sequence. This idea is explored for other puzzles for NPI-licensing in von Fintel 1999. The dynamic strict analysis is developed further by Gillies (2007) and critically compared to a pragmatically supplemented non-monotonic analysis by Moss (2010).

5 The Restrictor View

The dominant approach to the semantics of conditionals in linguistics is not so much an alternative to the accounts we have discussed so far, and in particular not to the Stalnaker/Lewis analysis, but a radical rethinking of the compositional structure of conditional sentences. It began with Lewis’s 1975 paper on adverbial quantification, which dealt with sentences like

(16) If it is sunny, we always/usually/mostly/rarely/sometimes/never play soccer.

Lewis argued that there was no plausible semantics for the conditional connective that would interact compositionally with the adverbs of quantification to give correct truth-conditions for these sentences. Instead, he argued that the if-clause added no conditional meaning of its own to the construction. The idea is that the only “conditional” operator in the structure is the adverb and that if merely serves to introduce a restriction to that operator. In other words, where naively one would have thought that (16) involved the combination of an adverbial quantificational operator with the conditional expressed by if, Lewis argued that there was just one operator and that if didn’t express any kind of conditional operator of its own.

Lewis himself did not generalize this idea; nowhere else in his writings does he give any indication that if’s found elsewhere are to be treated on a par with the if in adverbially quantified sentences. (It should be noted that in the adverbial quantification paper, Lewis does suggest that the if found in construction with probability operators is also not a conditional operator of its own, although he doesn’t say whether it is to be seen as a restrictor in those cases. It is a shame that Lewis did not connect his insights in the adverbial quantification paper to the problems surrounding conditional probability, as discussed for example in Lewis
Krater took the logical step and argued that Lewis’ idea should be applied to all conditional constructions. She put the point very concisely in Krater 1986: “The history of the conditional is the story of a syntactic mistake. There is no two-place if … then connective in the logical forms for natural languages. If-clauses are devices for restricting the domains of various operators. Whenever there is no explicit operator, we have to posit one.”

The central idea is that if itself does not carry any distinctive conditional meaning, rather it is, so to speak, a helper expression that modifies various quantificational/modal operators. This doesn’t just apply to when an overt operator combines with an if-clause but also when an if-clause occurs on its own with no overt operator in sight. In that case, Krater suggest, there must a covert, or at least not obviously visible, operator. What one might call bare indicative conditionals either contain a covert epistemic necessity modal or a covert generic frequency operator (≈ usually/always):

(17) If he left at noon, he’s home by now. [epistemic necessity]
(18) If he leaves work on time, he has dinner with his family. [generic frequency]

In bare subjunctive conditionals, one should consider the possibility that the modal form would is the operator restricted by the if-clause, an idea bolstered by the fact that there are if-less would-sentences (see Kasper 1992 and Schueler 2008):

(19) I would have beaten Kasparov.

It should be pointed out that while one may have a desire to have a uniform(ish) analysis of indicative and subjunctive conditionals, partially because both are if … then constructions, the restrictor analysis opens up a potentially large gap between them. The uniform presence of if would be almost entirely beside the point: how big the difference between the two kinds is depend on what, if any, difference there is between the modal operators present in them.

6 The Limit and Uniqueness Assumptions, Conditional Excluded Middle, and Might

Lewis and Stalnaker differ in their assumptions about the similarity ordering. Stalnaker assumes that for any (non-contradictory) antecedent and any evaluation world, there will be a unique most similar antecedent world. Lewis neither makes this Uniqueness Assumption (he calls it “Stalnaker’s Assumption”) nor the weaker Limit Assumption (that for any antecedent and evaluation world, there is a set of most similar antecedent worlds). Informally, here, we have been using the Limit Assumption but not the Uniqueness Assumption when we talk about the most similar or closest antecedent worlds. For discussion of this difference, see Lewis (1973: 19–21) and Stalnaker (1984: Chapter 7, esp. 140–142); Pollock
(1976), Herzberger (1979), and Warbrod (1982) argue for the Limit Assumption as well. More recently, Williams (2010b) and Swanson (2010) have revisited the topic.

In his argument against the Limit Assumption, Lewis adds cases that seem to show that the closeness to the actual world can get asymptotically closer:

(20) If this one inch line were more than an inch long, …

For any world in which the line is 1+x inches long, there will be a world where the line is just a little bit shorter but still more than an inch long. So, there will be no world where the line is as close to its actual length as possible while still more than an inch long. Stalnaker argues that in actual use, natural language would not make the fine-grained distinctions needed to threaten the Limit Assumption in such cases.

Cases that throw doubt on the Uniqueness Assumption are less recherché. Quine’s example (1950: 14) will do:

(21) If Bizet and Verdi had been compatriots, Bizet would have been Italian.
(22) If Bizet and Verdi had been compatriots, Verdi would have been French.

If the Uniqueness Assumption were correct, exactly one of (21) and (22) would be true. But they both seem dubious.

A virtue of the Uniqueness Assumption is that it validates the principle of Conditional Excluded Middle: either if p, q or if p, not q. Surprisingly, this validity seems to persist even in the problematic Bizet-Verdi case:

(23) Either if Bizet and Verdi had been compatriots, Bizet would have been Italian, or (if Bizet and Verdi had been compatriots) Verdi would have been French.

Stalnaker suggests that (21) and (22) are semantically indeterminate (because it is indeterminate which resolution of the similarity ordering is contextually salient), but that (23) is true nevertheless because it would be true under any reasonable resolution of the indeterminacy. He proposes implementing this suggestion in a supervaluation framework. Klinedinst (2011) explores this further and shows how it can make good on the suggestion by von Fintel & Iatridou (2002) and Higginbotham (2003) that a CEM-validating semantics for conditionals is behind the intuitive equivalence of pairs of conditionals under quantifiers (contra Leslie 2009; see also Huitink 2010):

(24) Every student would have failed if he had goofed off.
(25) No student would have passed if he had goofed off.

The validity of CEM also is involved in the intuitive relation between if-conditionals and only if-conditionals:

(26) Only if the Queen had been home, would the flag have flown.
(27) If the flag had flown, the Queen would have been home.
See Barker 1993 and von Fintel 1997 for discussion.

Another consequence of making the Uniqueness Assumption, is that Stalnaker cannot treat *would* and *might* as duals (as proposed by Lewis). Instead, he suggests that *might* is a higher (usually epistemic) operator that takes a whole counterfactual in its scope. And the embedded *would*-counterfactual is of course analyzed as usual. He gives the following paraphrase relation:

(28) If John had been invited, he might have come to the party.
(29) It might be the case that if John had been invited, he would have come to the party.

Consider the following contrast:

(30) It’s not the case that John must come to the party. But he might.
(31) #It’s not the case that John would have come to the party if he had been invited. But he might have.

(31) should be fine according to Lewis, it’s like saying *not all but some*. Stalnaker explains its anomaly this way: in the first sentence the speaker says that he knows that John wouldn’t have come, in the second she says that it’s compatible with her knowledge that he would have.

Lewis has a purported counterexample to this analysis. Suppose there is in fact no penny in my pocket, although I do not know it since I did not look. “Then ‘If I had looked, I might have found a penny’ is plainly false.” But it is true that it might be, for all I know, that I would have found a penny if I had looked.

Stalnaker contends that under the epistemic reading of *might*, the *might*-conditional is not in fact plainly false. He concedes a non-epistemic reading where the statement is false, but proposes to capture it by relativizing might to a situation where the speaker knows all the relevant facts. “This will yield a kind of quasi-epistemic possibility — possibility relative to an idealized state of knowledge.” This reading, as he points out, comes out almost identical to Lewis.

The stakes in this debate are increased by the observation that *might*-conditionals are very easy to read as true. If *might* and *would* are duals, then *would*-conditionals are predicted to be very hard to read as true. In fact, Hájek (2009) claims that (almost) all counterfactuals are false because of this. See DeRose 1999 for an earlier discussion and Hawthorne 2005, Williams 2008, 2010a for more recent relevant work. One might hope that thinking about the contextual dynamics of *might*-conditionals (see Gillies 2007 and Moss 2010) would defuse the argument.

7 Tense and Aspect

We will not be able to discuss the syntax of conditionals in this chapter (cf. Bhatt & Pancheva 2006) but we should take a look at the morphological fine structure of conditionals. It is quite apparent that in English at least, the indicative/subjunctive classification of conditionals is marked by tense & aspect morphology:
If Griepstra played his drums, de Gier played his flute.

If Griepstra had played his drums, de Gier would have played his flute.


The central observation is that what is commonly called subjunctive in “subjunctive conditionals” is an additional layer (or two) of past tense morphology, no matter whether the referred to state of affairs is temporally located in the past, present, or future:

a. If Roman comes to the party tomorrow, it will be a grand success.
b. If Roman came to the party tomorrow, it would be a grand success.
c. If Roman had come to the party tomorrow, it would have been a grand success.

Iatridou (2000) discusses this basic pattern (although she doesn’t discuss the two layer pasts in future or present conditionals) and proposes that the additional past does not serve a temporal function. Instead, she argues that the past tense has a schematic semantics that can be applied both temporally and modally: past is an “exclusion feature”. When past is used temporally it marks the times talked about as distinct from the now of the speaker (an additional wrinkle is needed to explain why past means past rather than non-present = past or future). When past is used modally it marks the worlds talked about as distinct from the actual world of the speaker (this does not mean that modally used past is a counterfactuality marker; rather, the intent is to derive something very much like the Stalnaker-analysis of the import of subjunctive marking).

The alternative to Iatridou’s account is to try to maintain that the additional pasts in subjunctive conditionals do after all retain their usual temporal meaning. This idea goes back to Dudman (1983, 1984, 1988, 1989) and has been pursued by Ippolito (2003, 2007) and Arregui (2005, 2009), among others. We do not have the space to survey the details of these accounts. Let’s rather look at a simplified sketch. Suppose that the extra layer of past tense marks that what the conditional quantifies over is a set of worlds that were accessible from the evaluation world at
a past time but may not be anymore. This is typically embedded in a branching futures version of possible worlds semantics. As the time index progresses, more and more open futures are precluded. Imagine that at some point in time, it was an open possibility that Roman would leave before noon, but by the present time it is settled that he did not. Then, assuming that the conditional employs a “historical necessity”-type of accessibility relation, the time index needs to be moved to the past to make sure that the domain of accessible worlds includes at least some worlds where he did leave before noon. Hence, the need for past tense marking on the modal (would = will + PAST) in (36b); the past tense in the antecedent may be a mere agreement phenomenon.

What then about the indicative conditional in (36a)? Clearly, if we assume a historical necessity modal, at the time of utterance it is already settled whether Roman did or did not leave before noon. So, if there need to be at least some antecedent worlds in the domain of the modal, the covert modal in (36a) cannot be a historical necessity modal. Thus, it is not mysterious why (36a) is naturally analyzed as involving a (covert) epistemic necessity modal.

In this story, then, the difference between indicative and subjunctive is two-fold: (i) type of accessibility relation/type of modal (epistemic vs. historical), (ii) time index on the modal (present vs. past). An obvious question is whether these differences cross-cut: are there past epistemic conditionals? are there present historic necessity conditionals? The answer to the second question is possibly yes: If Roman comes to the party tomorrow, it will be a grand success might arguably be a non-epistemic conditional. The answer to the first question might be expected to be no, since it is well-known that epistemic modals resist embedding under past tense.

One possibly problematic fact for the view just sketched comes from hindsight counterfactuals (Barker 1998, Edgington 2003):

(37) [A randomly tossed coin comes up heads.]
   a. If you had bet on heads, you would have won.
   b. If you bet on heads, you will win.

While (37a) seems acceptable and true after the coin has come up heads, there is no time in the past at which (37b) would have been rational to assert. While that doesn’t mean that there wasn’t a time at which the indicative conditional was true, it does throw some doubt on the simple idea that the only difference between (37a) and (37b) is the temporal perspective.

Related topics

indicative conditionals
References


Copley, Bridget. 2006. Temporal orientation in conditionals (or, how I learned to stop worrying and love UFOs). Ms, CNRS/Université Paris 8.

DeRose, Keith. 1999. Can it be that it would have been even though it might not have been? *Noûs* 33(s13). 385–413.


von Fintel, Kai & Sabine Iatridou. 2002. If and when *If*-clauses can restrict quantifiers. Ms, MIT.


[Word count: 6000]