On Stalnaker’s unified theory of conditionals

Justin Khoo

February 18, 2012

Abstract

A problem for Stalnaker (1975, 2011)’s unified theory of indicative and subjunctive conditionals is raised. Several solutions to the problem are discussed and rejected. The tentative conclusion is that we shouldn’t account for the semantic differences between indicative and subjunctive conditionals in terms of the mutual presuppositions of speakers in the discourse.

Broadly speaking, philosophers have been interested in two “kinds” of conditionals: indicative and subjunctive. Here are two paradigm cases:

(1) If the butler didn’t do it, someone else did.  Indicative
(2) If the butler hadn’t done it, someone else would have.  Subjunctive

(1) and (2) differ grammatically: the subjunctive, but not the indicative, contains an extra layer of “fake past” tense and the tense auxiliary would in its consequent. And these grammatical differences yield a difference in what the conditionals mean. Consider the following scenario:

By “fake past” we mean that the past tense in (2) doesn’t shift the event time of the antecedent to past. To see this clearly, notice that “tomorrow” is felicitous in the antecedent of the following subjunctive, which is not the case when the same sentence occurs outside of the antecedent:

(i)  a. The contest was held today. If you had entered tomorrow, you would have missed it.
    b. Luckily, Sue had entered the contest yesterday.
    c. #Unfortunately, John had entered the contest tomorrow.

This feature of subjunctive conditionals has been called “forward time shift” in the conditionals literature (c.f. Gibbard (1981), Dudman (1983, 1984), Edgington (1995), Bennett (2003)).
Mystery!

A murder has been committed. Here are the known facts: whoever did it acted alone, and it was a crime of passion (not premeditated)—if the actual culprit hadn’t done it, no other person would have. The butler is the prime suspect.

In this scenario, (1) seems true—after all, we know the murder was committed, so if the butler didn’t do it, someone else must have. And since it was a crime of passion, if the actual culprit (who we think is the butler) hadn’t done it, no one else would have—so (2) seems false. Thus, the two conditionals seem to differ in their truth (or assertability) conditions. Furthermore, their semantic difference is philosophically important: broadly speaking, indicatives express features of our knowledge, while subjunctives express metaphysical dependencies in the world—in slogan form, indicatives are epistemic while subjunctives are metaphysical.\(^2\)

Given its philosophical importance, we might wonder what accounts for the semantic difference between indicative and subjunctive conditionals. It’s not enough to say they involve different conditional operators and leave it at that (c.f. Lewis (1973), Gibbard (1981), Bennett (2003))—this simply rephrases our question as follows: why does subjunctive morphology and would yield the “subjunctive” operator while their absence results in the “indicative” operator? There’s some reason why the grammatical differences give rise to the semantic difference between indicative and subjunctive conditionals, and articulating that connection is the aim of this paper. Getting clearer about why they mean different things is one step along the way to better understanding what these conditionals mean.

One account of why indicatives and subjunctives mean different things is due to Robert Stalnaker, who explains their semantic differences in terms of the relationship between the proposition the conditional expresses in a context and the presuppositions of the speakers of that context (see Stalnaker (1975, 2011)). In this paper, I raise some trouble for

\(^2\)It’s this semantic difference why subjunctives but not indicatives figure into talk about causality (see Lewis (1986, 2000), Woodward (2003)) and bear an intimate connection to laws of nature (see Goodman (1947), Chisholm (1955), Maudlin (2007), Lange (2009)). It’s also this difference that drives causal decision theorists to insist that it’s the probability of subjunctives (rather than indicatives) that ought to guide rational deliberation (see Gibbard & Harper (1981), Lewis (1981a,b), Williamson (2007)).
Stalnaker’s theory and explore how it might be patched up to avoid such problems. Although no conclusive refutation of the theory is forthcoming, the lack of a solution to these problems should motivate us to look elsewhere for an account of the differences between indicative and subjunctive conditionals, one not in terms of the mutual presuppositions of speakers in a discourse.

1 Stalnaker’s theory

On Stalnaker’s theory, both indicative and subjunctive conditionals have the same underlying semantics:

\[(S) \text{ Semantics} \]
\[\text{if } p, q \text{ is true at } w \text{ iff } f(p, w) \in q\]

Here, \(f\) is a function from a world \(w\) and proposition \(p\) (for simplicity, I assume that propositions just are sets of worlds) to the closest \(p\)-world to \(w\).\(^3\) \((S)\) says that if \(p, q\) is true at \(w\) just in case the closest \(p\)-world to \(w\) is a \(q\)-world. To spell out the difference between indicatives and subjunctives on this theory, we need the notion of a common ground (Stalnaker (1978, 1999, 2002))—the set of propositions which are mutually presupposed (mutually taken for granted) by the discourse participants. The common ground determines a set of worlds which are the worlds true at each proposition in the common ground—this is called the context set, \(C\). Thus, if \(p\) is mutually presupposed in some conversation, where \(C\) is the context set of that conversation: \(C \subseteq p\). In a Stalnakerian framework of conversation, assertions aim to update the context set by eliminating from it all the worlds incompatible with the asserted proposition (equivalently: assertions aim to make the asserted proposition mutually presupposed, or part of the common ground).

For Stalnaker, indicative conditionals carry a default pragmatic constraint on \(f\) that every world in \(C\) is closer to every other world in \(C\) than it is to any world outside of \(C\) (I’ll mark when \(f\) is subject to the pragmatic constraint with a subscript _c_):

\[^3\text{See Stalnaker (1968), Lewis (1973, 1979), Bennett (2003) for more discussion on the closeness metric.}\]
Pragmatic Constraint
\[ \forall p : \forall w \in C : f_c(p, w) \in C \]

Assuming that the conditional is infelicitous when \( f \) is undefined, this nicely predicts that indicative conditionals are felicitous only if \( C \cap p \neq \emptyset \), which delivers the correct verdict about (4):

(4) A: Booth shot Lincoln—here’s the proof.
    B: I am convinced. #I also think that if Booth didn’t shoot Lincoln, someone else did.

On this account, subjunctive morphology is understood as a conventional signal that the pragmatic constraint is lifted. Hence, for a subjunctive conditional there might be a world in the context set at which the closest worlds to it are not in the context set. By lifting the pragmatic constraint, subjunctives are felicitous in contexts in which their antecedents are presupposed false, hence the felicity of (5):

(5) A: Booth shot Lincoln—here’s the proof.
    B: I am convinced. I also think that if Booth hadn’t shot Lincoln, someone else would have.

Thus, on Stalnaker’s theory, indicative morphology is the default (unmarked) case, and signals that the pragmatic constraint governs \( f \). Hence, when \( w \in C \), an indicative conditional \( \text{if } p, q \) is true at \( w \) iff \( f_c(p, w) \in q \) iff the closest \( p \)-world to \( w \) that is also in \( C \) is a \( q \)-world. Subjunctive morphology signals that the pragmatic constraint is lifted, and hence, even when \( w \in C \), a subjunctive conditional \( \text{if } p, q \) is true at \( w \) iff \( f(p, w) \in q \) iff the closest (overall) \( p \)-world to \( w \) is a \( q \)-world. The intuitive idea is that the proposition expressed by an indicative conditional is constrained by the mutual presuppositions in our context—this is what makes indicatives epistemic. The proposition expressed by a subjunctive conditional is not so constrained, which is why subjunctives are metaphysical—this metaphysical relation is cashed out in terms of overall similarity between worlds.
1.1 The problem

Recall our indicative/subjunctive minimal pair:

(1) If the butler didn’t do it, someone else did.

(2) If the butler hadn’t done it, someone else would have.

Let \( \neg b \) be the proposition that the butler didn’t do it, and \( s \) be the proposition that someone else did. The intuition is that (1) is true but (2) is false. Can Stalnaker’s theory predict this intuition? Kinda-sorta. Notice that, according to Stalnaker’s theory, the propositions expressed by (1)/(2) in \( C \) are (respectively):

(6) a. \( \lambda w.f_c(\neg b, w) \in s \)

b. \( \lambda w.f(\neg b, w) \in s \)

Do these propositions have different truth values at the actual world \( \alpha \)? Not if \( \alpha \) is not in the context set (as may happen if something false is presupposed). For the pragmatic constraint only applies to worlds in the context set—hence, if \( \alpha \not\in C \) then \( f_c(\neg b, \alpha) = f(\neg b, \alpha) \).\(^4\) Thus, if \( \alpha \not\in C \), both the indicative and subjunctive versions of if \( \neg b \), \( s \) are true just in case \( f(\neg b, \alpha) \in s \). Now, if \( \alpha \in C \), then their actual truth value might differ, because then the pragmatic constraint will kick in, allowing for the possibility that \( f_c(\neg b, \alpha) \in s \) and \( f(\neg b, \alpha) \not\in s \); hence, in that case the theory can predict that the indicative (1) is (actually) true but subjunctive (2) is (actually) false.

However, it’s not clear that predicting that (1) and (2) actually differ in truth value only if \( \alpha \in C \) captures our intuitions about the difference in what (1) and (2) mean. After all, our intuition that (1) is true and (2) is false remains even if we are sure that something that

---

\(^4\)It’s crucial that, when \( \alpha \not\in C \), the pragmatic constraint not apply to indicative conditionals evaluated at \( \alpha \), otherwise the theory would predict that modus ponens is invalid. The reason is that, if the pragmatic constraint applies to indicatives evaluated at worlds outside of the context set, even though the closest \( p \)-world to \( w \) is itself whenever \( p \) is true at \( w \), if \( w \not\in C \) it may still be the case that \( p \) is true at \( w \) and the closest \( p \)-world to \( w \) that is in \( C \) is not \( w \). Thus, the following scenario, which exploits this fact, would be possible: \( \alpha \not\in C \), \( \alpha \in p \), \( \alpha \not\in q \) and \( f_c(p, \alpha) \in q \). On this scenario, the indicative if \( p \), \( q \) is true, but \( p \) is true and \( q \) is false. Holding that the pragmatic constraint does not apply to \( f \) when the conditional is evaluated at a world outside of the context set avoids this problem. This is nicely illustrated in Starr (2010).
has been presupposed is false. Stalnaker (2011) discusses a related problem (raised by Andy Egan and Brian Weatherson), and recommends thinking of the common ground in a slightly different way such that some propositions in the common ground are discourse-relevant and others discourse-irrelevant. It is only those propositions which are discourse-relevant that determine the context set. Thus, if we end up presupposing something false but irrelevant to the topic at issue, this won’t have the result that \( \alpha \notin C \). Thus, even in a context with some false presuppositions, the theory can predict that each member of an indicative/subjunctive minimal pair has different actual truth conditions.

I’m not sure this fix is enough to avoid our current problem, since there can still be circumstances in which we end up presupposing a discourse-relevant falsehood, and in which we still have the intuition that (1) is true and (2) false. Consider: Holmes and Watson are hot on the trail of the killer, and they’re in a heated discussion about who did the deed and how they did it. Thus, two questions are relevant in their discussion: who killed the heiress, and what was the murder weapon? Even if they’ve come to mutually presuppose a false answer to the second question, since they suspect the butler did it, know that someone did it, and also know that no one other than the actual culprit would have stepped in had he or she failed, in this situation (1) still seems true while (2) seems false.

But the Stalnakerian has a fallback strategy. Although according to Stalnaker’s theory, the propositions expressed by (1) and (2) can differ in truth value only if \( \alpha \in C \), the conditions under which the propositions they express are accepted differ regardless of whether \( \alpha \in C \)—hence, this difference in their acceptance conditions might be enough to account for our intuitions about the difference in what (1) and (2) mean. Adopting this

---

5 Why would we allow the common ground to contain a falsehood? We might accept some proposition for the sake of moving the conversation forward in a smooth manner, especially if it doesn’t matter to the main point of the discourse.

6 The idea that only some propositions are discourse-relevant is independently motivated—see Yalcin (2011)’s notion of a “modal resolution” and Schaffer (2004) on contrastivism about knowledge, along with the semantic and philosophical literature on questions, especially Higginbotham (1996), Groenendijk & Stokhof (1997).

7 Carrying out this line of thought, the Stalnakerian might respond that the false presupposed proposition about the murder weapon is not discourse-relevant. But this now looks more like tailoring one’s account of discourse-relevance to save the theory rather than employing an independently motivated mechanism to resolve a problem. The threat of ad-hocity looms.

8 I’m using “accept” in the semi-technical sense of Stalnaker (1975) in which a proposition \( p \) is accepted in a context set \( C \) iff \( p \) is true at every world in \( C \) iff \( p \) is presupposed in \( C \).
account of the difference between (1) and (2) is analogous to how Stalnaker handles the “direct argument” (in Stalnaker (1975)):

(7)    a. Either the butler or the gardner did it.
       b. If the butler didn’t do it, the gardner did.

Although the inference from (1) to (2) isn’t valid, on Stalnaker’s theory it is a reasonable inference in the following sense: in any context in which (1) is appropriately asserted and accepted, (2) is accepted. Similarly, the strategy here is to show that (1) and (2) have distinct acceptance conditions—hence going from (1) to (2) is not a reasonable inference.

Let’s convince ourselves that we can accept/reject the propositions expressed by (1)/(2) respectively. The conditions under which (1) is accepted and (2) not (in some context $C$) are such that:

**Accept Indicative (1)**

$$\forall w \in C : f_c(\neg b,w) \in s$$

**Reject Subjunctive (2)**

$$\neg \forall w \in C : f(\neg b,w) \in s$$

Suppose that **Accept Indicative (1)** is true. Can we find a world in $C$ at which $f(\neg b,w) \notin s$? Sure: suppose $w_1 \in C$ and $f_c(\neg b,w_1) = x$ and suppose that $x \in s$. It might still be the case that $f(\neg b,w_1) = y$ where $y \notin s$. (This will happen when the closest $\neg b$-world to $w_1$ in $C$ is $x$ but the closest overall $\neg b$-world to $w_1$ is $y$) But then $w_1$ is a world compatible with **Accept Indicative (1)** being true which is enough to make **Reject Subjunctive (2)** true. Given that this strategy avoids the false presupposition problem, perhaps this is how the Stalnakerian theory should best distinguish the meanings of (1) and (2)—on this strategy, a key feature of their semantic difference is that the propositions they express have different acceptance-conditions.

So far so good. Now, holding fixed the scenario **Mystery!**, notice that if we negate the consequents of (1) and (2), our intuitions shift: the indicative now seems false and the subjunctive true,

(8)    If the butler didn’t do it, no one else did.
(9) If the butler hadn’t done it, no one else would have.

Recall that we know the murder was committed, so even though the butler is our prime suspect, (8) seems false. But it was a crime of passion after all, so (9) seems true. Thus, in *Mystery!*, the indicative (8) seems false and the subjunctive (9) seems true. Can Stalnaker’s theory predict this difference?⁹

Since it faces the same problem (in reverse) predicting a difference in the actual truth values of (8) and (9), let’s see whether Stalnaker’s theory can predict a difference in their acceptance conditions.¹⁰ Unfortunately, as I will show now, this strategy cannot deliver the necessary difference—we cannot reject (8) and accept (9). Take the two propositions respectively:

(8) \[ \lambda w. f_c (\neg b, w) \in n \]
(9) \[ \lambda w. f (\neg b, w) \in n \]

There are similar cases in which we have an indicative/subjunctive pair in which we judge the subjunctive likely and the indicative unlikely. Jane was recently offered a bet on a fair coin flip. She isn’t a big gambler, so you’re pretty sure she didn’t take the bet, but you also think it’s likely that if she took the bet, she bet on heads, since that’s what she usually does. The coin ended up landing heads. After the money is paid out on this bet, you learn that Jane has no more money now than before the bet, so you think the following is extremely unlikely:

(i) If Jane took the bet, she won.

Nonetheless, since she usually bets on heads, you regard the following as very likely:

(ii) If Jane had taken the bet, she would have won.

Adams (1970) discusses several similar cases. One particularly compelling case is that even if you believe:

(iii) If Oswald hadn’t shot Kennedy, Kennedy would be alive today.

you aren’t thereby irrational if you refuse to take the following bet:

(iv) If Oswald didn’t shoot Kennedy, Kennedy is alive today.

¹⁰As before, the theory can predict a difference in the actual truth values of (8) and (9) only if \( \alpha \in C \), since if not then (8)(9) are true at \( \alpha \) iff \( f(\neg b, \alpha) \in n \). If \( \alpha \in C \), then the pragmatic constraint will allow the theory to generate the right result. For, suppose there are only \( \alpha, w_1 \in C \) and that \( f(-b, \alpha) = w_1 \), whereas \( f(-b, \alpha) = w_2 \) and that \( w_1 \not\in n \) but \( w_2 \in n \). Then, (8) will be false and (9) will be true. But this is not enough to account for our intuitions in the case—that (8) seems false and (9) seems true doesn’t depend on whether we aren’t presupposing anything false.
If (8) is rejected while (9) accepted in $C$, $C$ must be such that:

**REJECT INDICATIVE (8)**

$\neg \forall w \in C : f_c(-b, w) \in n$

**ACCEPT SUBJUNCTIVE (9)**

$\forall w \in C : f(-b, w) \in n$

Suppose that **REJECT INDICATIVE** (8) is true. Then there is a world $w_1 \in C$ such that $f_c(-b, w_1) \not\in n$. Let $f_c(-b, w_1) = x$. It must be the case that $x \in C$ and $x \in \neg b$ and $x \not\in n$. But every world is closer to itself than any other world—if $w \in p$ then $f(p, w) = w$. Thus, it follows that $f(-b, x) = x$ and hence that $f(-b, x) \not\in n$ and since $x \in C$ that $\exists w \in C : f(-b, w) \not\in n$ and hence that **ACCEPT SUBJUNCTIVE** (9) is false. More simply, once (8) is rejected in $C$, there is a world in $C$ at which someone other than the butler did it. But since this world is closer to itself than any other world is, it’s a world at which (9) is false, and hence its presence in $C$ ensures that (9) isn’t accepted in $C$.

Hence, we cannot reject (8) and accept (9)—yet this is the pattern of our intuitions about them: in the scenario **Mystery!**, (8) seems false and (9) seems true. Furthermore, as we saw, the theory cannot deliver a difference in the actual truth values of (8) and (9) unless $\alpha \in C$, which is not enough to account for our intuitions about what they mean, since our intuition that one is true and the other false persists even if something relevant and false is presupposed. It seems that the Stalnakerian theory cannot account for our intuitions about indicative/subjunctive minimal pairs like (8)/(9), in which the indicative seems false and the subjunctive true. However, perhaps we’ve simplified things too much. In the

---

11This constraint on closeness is explicitly endorsed in Stalnaker (1968), Lewis (1973).

12The same point can be put slightly differently. If Stalnaker’s theory is right, then (9) is accepted in $C$ only if $C$ contains no worlds at which someone besides the butler did it (no $(\neg b \land s)$-worlds in $C$). Hence, Stalnaker’s theory predicts the following should be a reasonable inference:

(i) a. It is possible that someone besides the butler did it. $\diamond (\neg b \land s)$

b. So, if the butler hadn’t done it, someone else would have. $\text{if } \neg b, s$

(Once (1) is felicitously asserted and accepted, there are some $(\neg b \land s)$-worlds in $C$, so (2) must be accepted.) But this is not a compelling inference at all! Recall the facts of the case: you’re pretty sure (but not positive) it was the butler who did it, so you accept (1), but you’re certain that if the actual culprit hadn’t done it, no one else would have, so you reject (2).
next section, I consider whether more complicated Stalnakerian account of the difference between indicatives and subjunctives can yield the right results.

1.2 Context shifting?

Perhaps the best move for the Stalnakerian is to reject both explanations and instead hold that there is no single context in which we evaluate both (1)/(2) and (8)/(9). The motivation for this thought is that it sometimes feels hard to evaluate a subjunctive conditional in a context in which its antecedent is not presupposed to be false, and this is the kind of context that I assume we can evaluate both (1)/(2) (and (8)/(9)) in. After all, the following sentence sounds odd:

(10) #It’s possible that the butler didn’t do it, and if the butler hadn’t done it, no one else would have.

We might reason as follows: subjunctives presuppose the falsity of their antecedents—hence, once the subjunctive (9) is uttered, if there are any worlds in which the butler didn’t do it, those worlds are eliminated (by presupposition accommodation) from the context prior to computing the proposition expressed by (9). Let \( C^+ \) be the context set that results from adjusting \( C \) to account for the utterance of (9) and accommodating its presuppositions if necessary—crucially, \( C^+ \) will contain no \( \neg b \)-worlds.\(^{13}\) The proposition expressed by (9) in \( C^+ \) is accepted in \( C^+ \) iff

(11) \( \forall w \in C^+ : f(\neg b, w) \in n \)

Let’s suppose this condition is met. Now, recall from §1 that Stalnaker’s theory (correctly) predicts that indicative conditionals are felicitous in \( C \) only if there are some worlds in \( C \) in which their antecedent is true. Let’s suppose that this felicity condition is also a presupposition, or at least in many cases induces presupposition accommodation-like effects—then, uttering the indicative (8) in a context in which there are no worlds in which the butler didn’t do it (no \( \neg b \)-worlds), such as \( C^+ \) will induce accommodation to add some

---

\(^{13}\)This is the post-utterance, pre-evaluation stage of context update—see Stalnaker (2002), von Fintel (2008) for further discussion and motivation for this level of theoretical grain.
antecedent-satisfying \( \neg b \)-worlds to the resulting context prior to evaluating what proposition it expresses. Let \( C^{++} \) be the context that results from adjusting \( C^+ \) to account for the utterance of (8) and accommodating its presuppositions if necessary. The crucial point is that \( C^{++} \) will contain some \( \neg b \)-worlds, unlike \( C^+ \)—and furthermore, since (let’s suppose), the accommodated worlds satisfy all our mutual presuppositions, including the one that a crime was committed, they will all be worlds at which someone else (besides the butler) did it. That is, all the accommodated worlds will be \( (\neg b \land s) \)-worlds. Now, the proposition expressed by (8) in \( C^{++} \) is accepted in \( C^{++} \) iff

\[
(12) \quad \forall w \in C^{++} : f_c(\neg b, w) \in n
\]

and hence rejected in \( C^{++} \) if

\[
(13) \quad \neg \forall w \in C^{++} : f_c(\neg b, w) \in n
\]

And certainly (8) can be rejected in \( C^{++} \) while (9) accepted in \( C^+ \). Our earlier counterexample to the possibility that (8) be rejected while (9) accepted mistakenly assumed, so the thought here goes, that the conditionals are evaluated in the same context. As long as we assume that subjunctive conditionals presuppose the falsity of their antecedents while indicatives presuppose the possibility of their antecedents, we have a way around that earlier problem: we can accept (9) and reject (8), just not in the same context. Proposal: this accounts for our intuition that (8) is false and (9) true.

There are two reasons to resist this response. The first is that it’s not clear it predicts the facts. In the scenario sketched above, you aren’t presupposing that the butler did it (only that he’s the most likely culprit). Yet the subjunctive:

\[\text{Here’s explicit example. Let } \neg b = \{w_2, w_3\}, n = \{w_1, w_3\}, C^+ = \{w_1\}, C^{++} = \{w_1, w_2\} \text{ and suppose that } f(\neg b, w_1) = w_3 \text{ and that } f_{c^{++}}(\neg b, w_1) = w_2. \text{ Then}
\]

(i) \( \forall w \in C^+ : f(\neg b, w) \in n \)

and so (9) is accepted in \( C^+ \). Yet,

(ii) \( \neg \forall w \in C^{++} : f_{c^{++}}(\neg b, w) \in n \)

(the witnessing world is \( w_1 \)) and so (8) is rejected in \( C^{++} \).

\[\text{A similar story can be told about our intuitions about (1) and (2).}\]
If the butler hadn’t done it, no one else would have.

seems true. The proposal we’re now considering aims to predict this judgment by saying that you only accept (9) if you also presuppose that the butler did it—that’s the best the theory can do. But it’s not obvious this is enough to predict our judgment. I hear (9) as true even if I’m not presupposing the butler did it, as long as I focus on my certainty that it was a crime of passion, and hence that if the actual culprit hadn’t done it, no one else would have. But I recognize this intuition might not be widely shared—so I won’t rest my case against this response on this intuition alone.

The second reason to resist this response is that the kind of context shifting it appeals to is not independently motivated. The above line of reasoning relied on the claim that subjunctive conditionals presuppose the falsity of their antecedents, but they don’t—indeed, Stalnaker’s theory is designed exactly to predict this fact. Consider the following example from Anderson (1951):

If Jones had taken arsenic, he would have shown exactly those symptoms he in fact does show.

Intuitively, this conditional serves as evidence that Jones in fact took arsenic. Hence, it doesn’t presuppose that Jones didn’t take arsenic—if it did, it would be self-defeating. Indeed, examples like this are commonplace. Here’s another:

So you think John won the election? Well, if he had won the election, it would have been recorded in the city’s official documents. So, let’s check to see if it is.

Indeed, Stalnaker himself gives the following example:

a. The murderer used an ice pick.
   b. If the butler had done it, he wouldn’t have used an ice pick.
   c. Therefore, the murderer must have been someone else.

Intuitively, (16-b) serves as a premise in an informative argument for the conclusion (16-c). But if (16-b) presupposed that the butler didn’t do it, the argument would simply beg the question of (16-c). Since presuppositions don’t come and go like this—their triggers are
lexically encoded—we may conclude that subjunctives don’t presuppose the falsity of their antecedents.\textsuperscript{16}

So then why are examples like (10) infelicitous?

(10) #It’s possible that the butler didn’t do it, and if the butler hadn’t done it, no one else would have.

One standard thought is that by asserting a subjunctive conditional, the speaker conversationally implicates that she believes that its antecedent is false. Given that this belief is not a mere belief but rather the belief that one knows, then we might expect the kind of infelicity exhibited in (10). For in (10), the speaker asserts that it’s epistemically possible that the butler didn’t do it, and hence that the speaker does not know that the butler did it, and then asserts (9), thus implicating that she believes that she knows the butler did it. Hence, she ends up asserting $p$ and then implicating that she believes that $\neg p$, which should be infelicitous for the same reason that Moore-paradoxical sentences like (17) are:

(17) #It’s raining but I believe it’s not raining.

A feature of conversational implicatures that distinguishes them from conventional implicatures and presuppositions is that they can be canceled—indeed, we might think that this is so in cases like (14), (15), and (16-b) above, since by asserting them in their respective contexts one does not implicate the falsity of their antecedents.\textsuperscript{17} Furthermore, when the implicature doesn’t arise, neither does the pattern of infelicity exhibited by (10):

(18) It’s possible that Jones took arsenic. And if Jones had taken arsenic, he would have shown exactly those symptoms he in fact does show.

\textsuperscript{16}This is the standard view in the recent literature—examples like (14) have convinced most theorists that subjunctives implicate (either conversationally or conventionally, c.f., Grice (1989)) that their antecedents are false. In addition to Stalnaker (1975), see Karttunen & Peters (1979), von Fintel (1997), Iatridou (2000), Ippolito (2003), Arregui (2007) for further discussion.

\textsuperscript{17}Sketching my account of why these particular subjunctives do not carry the counterfactual implicature will take us too far afield, so I’ll omit it for now, and if people want we can discuss this in Q&A. The important point here is that a presupposition account of the infelicity of (10) won’t do, and that’s the kind of account the context-shifting Stalnakerian theory seems to require.
Back to our main point. Recall that the version of the Stalnakerian theory under consideration aims to explain how we can reject (8) and accept (9) by holding that the two are not evaluated in the same context. All this was in service of predicting our judgment that (8) seems false and (9) true in scenarios like Mystery!. But for this explanation to be successful, we want some independent motivation for this kind of context shifting, which is why we looked for evidence that subjunctives carry counterfactual presuppositions. If subjunctive conditionals presupposed the falsity of their antecedents (while indicatives presupposed the possibility of their antecedents), we’d have independent motivation that (8) and (9) are evaluated in different contexts. But we’ve just seen, first, that there is reason to think that subjunctive conditionals do not presuppose the falsity of their antecedents, and second, the data that might seem to motivate such counterfactual presuppositions (i.e., the infelicity of (10)) can be explained by the fact that by asserting a subjunctive conditional, one implicates that one believes its antecedent is false. In spite of all this, the Stalnakerian could simply stick to his or her guns and hold that (8) and (9) are evaluated in different contexts. But without any independent evidence why this is so, this isn’t a very satisfying position to be in.

This isn’t to say that there is no way of extending the theory in some way so as to account for the difference in our intuitions. Indeed, since the theory is silent on what other constraints might hold on the selection function \( f \) of conditionals evaluated at worlds outside of the context set, it is an open question whether more can be said to extend the theory’s account of the semantic differences between indicative and subjunctive conditionals.\(^{18}\) However, I am not sure that proceeding in this way will ultimately prove fruitful—that is, it might turn out that the semantic differences between indicative and subjunctive conditionals can be explained without recourse to the pragmatic constraint at all. I happen to think this is actually the case, but telling that story I leave for another day.\(^{19}\)

---

\(^{18}\)I thank Bob Stalnaker for emphasizing this point to me.

\(^{19}\)Many thanks to Dan Greco, Aaron Norby, Bob Stalnaker, Zoltán Szabó, Steve Yablo, and Igor Yanovich for helpful discussion and comments on earlier drafts.
References


