Layering modalities: the case of backtracking counterfactuals[1] [2]

Ana Arregui
University of Ottawa

1. Introduction

What are the combinatorial possibilities of modality? This question has not often been addressed in the linguistic literature.[3] Where do we find layered modalities? What restrictions are such combinations subject to? To some degree, of course, limitations will be imposed by syntax. Are there others?

In this paper I investigate certain counterfactual constructions that I characterize as cases of layered modalities. The counterfactuals that interest me have some rather special properties, and are known in the philosophical literature as *backtracking counterfactuals*. I show that the possibility of layering modalities opens up interpretations that are not available to single modality counterfactuals. I introduce backtracking counterfactuals in §2. In §3 I discuss constraints proposed by Lewis to guide the resolution of similarity in counterfactuals, and present a typology of backtracking counterfactuals. In §4 I discuss the exact nature of the embedded modality, and in §5 I spell out details of a compositional analysis of the interaction between modal auxiliaries. Conclusions can be found in §6.

2. Backtracking counterfactuals

Backtracking counterfactuals are counterfactuals that assert that if things had been different at some past, present or future time, they would have been different at an earlier time too. A brief discussion of backtracking counterfactuals made its way into David Lewis's 1979 paper *Time's Arrow*:

(1) *Jim and Jack quarreled yesterday, and Jack is still hopping mad. We conclude that if Jim asked Jack for help today, Jack would not help him. But wait: Jim is a prideful fellow. He never would ask for help after such a quarrel; if Jim were to ask Jack for help today, there would have to have been no quarrel yesterday. In that case Jack would be his usual generous self. So if Jim asked Jack for help today, Jack would help him after all.* (Lewis 1979)

The backtracking conditional in this example has been singled out below:

---

[1] This paper is based on a chapter of my dissertation (Arregui 2004). As well as thanking my dissertation committee (Angelika Kratzer, Philip Bricker and Barbara Partee), I would like to thank Cleo Condoravdi and Frank Veltman for insightful comments, as well as the audience of the Kyoto workshop "Language Under Uncertainty". Remaining errors remain my own.

[2] At the time of writing this paper, I was unaware of Frank (1997)'s discussion of backtracking, which should be addressed in the discussion. More recent investigations into backtracking (such as Schulz (forthcoming)) should also be addressed in future work on this topic (author's note, 2009).

[3] But see Frank Veltman's presentation at "Language Under Uncertainty", Kyoto, January 2005.

(2) *If Jim asked Jack for help today, there would have to have been no quarrel yesterday.*

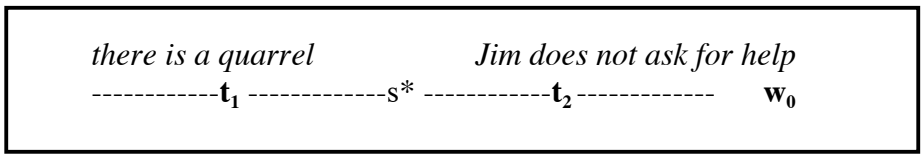
Judgments surrounding backtracking counterfactuals are rather subtle. Backtracking counterfactuals are usually judged false. But not always. As we see in Lewis's example, contextual support can help a lot in shifting our intuitions to the point where we are prepared to judge a backtracking counterfactual as true. Moreover, backtracking counterfactuals are usually helped by a special, characteristic syntactic structure. We find it in the example in (2), which has *would have to have* in the consequent clause. In a very interesting parenthetical comment, Lewis remarks on this 'special syntax':

(3) *(Back-tracking counterfactuals, used in a context that favors their truth, are marked by a syntactic peculiarity. They are the ones in which the usual subjunctive conditional constructions are readily replaced by more complicated constructions: "If it were that….. then it would have to be that…." or the like. A suitable context may make it acceptable to say "If Jim asked Jack for help today, there would have been no quarrel yesterday", but it would be more natural to say "…there would have to have been no quarrel yesterday. (…))* (Lewis 1979)

In this paper, an interpretation for backtracking conditionals is built on top of a Lewis-Stalnaker-style analysis. According to Lewis-Stalnaker, the interpretation of counterfactuals is characterized in terms of quantification over possible worlds: a (counterfactual) conditional is true iff the consequent clause is true in the most similar worlds in which the antecedent clause is true.[4] Making certain assumptions:

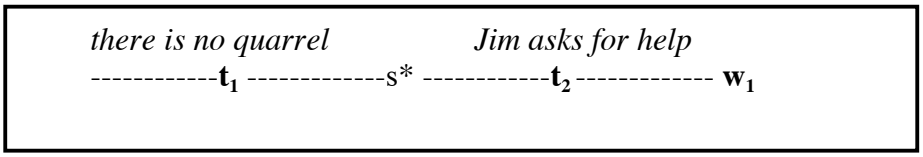
(4) *if A, would B* is true in w iff B is true in R(A, w),
where R is a relation that takes as arguments the antecedent clause proposition and the evaluation world, and returns as value the most similar worlds to the evaluation world in which the antecedent clause is true.

Let's look at (2) more closely. Say the actual world is w_0:



In the actual world there is a quarrel, and Jim does not ask for help. The counterfactual in (2) seems to claim that the worlds quantified over are worlds like w_1, which differ from the actual world at some before Jim's request:

[4] There are many differences between the analysis proposed by Stalnaker and Lewis. I will set them aside here. There are enough similarities to be able to speak of a Stalnaker-Lewis analysis, and the term is widespread in the literature.



This is the pattern that is typical of backtracking counterfactuals. We only judge them true if we accept that the worlds quantified over are worlds that differ from the actual world at some time before the *if*-clause time. For this reason, I will occasionally say that backtracking counterfactuals ask us to 'change the past'.

We will distinguish between two kinds of backtracking conditionals, which I will term *real backtrackers* and *conditionals with backtracking resolution*.

Real backtrackers are conditionals that explicitly claim that if a certain hypothesis held at some time (past, present or future), something different would have happened at some earlier time. Examples are given below:

(5) a. *Real backtrackers with regular syntax*
If Jim asked Jack for help, there would have been no quarrel yesterday.

b. *Real backtrackers with special syntax*
If Jim asked Jack for help, there would have to have been no quarrel yesterday.

The examples in (5) explicitly claim that if the future were different, the past would be different too. Some real backtrackers come with regular syntax (5a), and some come with special syntax (5b). As Lewis (1979) points out, it is usually much easier to agree to the truth of real backtrackers with special syntax than to the regular ones.

Conditionals with a backtracking resolution do not make a claim about times preceding the antecedent clause. However, they are special because we only judge them true if we accommodate a difference at some time earlier than the *if*-clause. Lewis's example is given below:

(6) If Jim asked Jack for help, Jack would help him.

In the context of Lewis's story, the example in (6) is initially judged false. But if we follow the reasoning proposed by Lewis, and think that Jim would never ask for help if there had been a quarrel, and so if Jim asked for help there would have to have been no quarrel, and if there had been no quarrel, Jack would be helpful, then we judge (6) to be true. That is, (6) is judged true if we allow ourselves to understand it as the strengthened conditional in (7):

(7) If there had been no quarrel and Jim asked Jack for help, Jack would help him.

In this paper I will be concerned with real backtrackers. I will set aside conditionals that merely have a backtracking resolution. To me, these seem to constitute a

slightly different case. They bear some similarity with examples like (8), topic of von Fintel (2001):

(8) If the US threw its weapons into the sea tomorrow, there would be war; but if all nuclear powers threw their weapons into the sea tomorrow, there would be peace.

The example in (8) is quoted by von Fintel (2001) from a discussion by Lewis as an example illustrating subtle context change. We evaluate the first conditional with respect to minimally different scenarios (the US is the only nuclear power that disposes of its weapons), and judge it true. However, once the set of scenarios has been changed (by the strengthened antecedent), we are happy to conclude that if the US threw its weapons into the sea tomorrow, there would be peace. Conditionals with backtracking resolutions are somewhat similar: we only judge them true after a little story that explicitly modifies the set of scenarios in which we evaluate the antecedent (some kind of strengthening). They are a little different. But even if we set aside conditionals with a backtracking resolution, it is important to keep them in mind. Intuitions in this domain are often slippery, and conditionals with a backtracking resolution show that context can play a powerful role.

3. On the possibility of changing the past

As we will see, restrictions on the similarity relation needed to make sure that the analysis of non-backtracking counterfactuals works, straightforwardly predict that backtracking should lead to falsehood. This makes the backtracking counterfactuals that are judged true rather interesting. I will begin this section by presenting Lewis's discussion of restrictions on similarity (§3.1). I will also present a typology of backtracking counterfactuals (§3.2). Not all backtrackers are born alike, and what needs to be said varies quite significantly depending on the case.

3.1 The beginning of my story: Lewis (1979)

In his 1979 paper *Time's Arrow*, David Lewis defends a temporally asymmetric view of counterfactual dependencies. He argues that the way things are later depends counterfactually on the way things are earlier, and not the other way around. The future depends counterfactually on the present (*If the present were different, the future would be different*), and the present depends counterfactually on the past (*If the past had been different, the present would have been different*). But the past does not depend counterfactually on the present (*If the present had been different, the past would have been different*), and the present does not depend counterfactually on the future (*If the future were different, the present would be different*).

Backtracking conditionals appear to challenge Lewis's asymmetric view. They explicitly claim that something that happens earlier depends counterfactually on something that happened later. Lewis thought that backtracking conditionals are special: usually considered false, and often marked by a special selection of modals and auxiliaries. Not considered counterexamples, he set them aside. Still, of course, we want to know how they work.

Why does Lewis claim that the resolution of counterfactuals is usually asymmetric? A good starting point is the example used by Kit Fine to challenge Lewis's use of similarity to resolve the interpretation of counterfactuals:

(9) If Nixon had pushed the button, there would have been a nuclear holocaust.
(Fine 1975)

Consider a world that differs minimally from the actual world in which somebody disconnected the button before Nixon pushed it, and there was no nuclear holocaust. Intuitively, such a world appears more similar to the actual world than a world that differs minimally from the actual world in which nobody disconnected the cable, Nixon pushed the button, and a nuclear holocaust followed. Yet, we typically judge (9) true, indicating that worlds in which somebody disconnected the cable do not make it into the domain of quantification of the modal. Fine thought that this showed that Lewis's use of similarity was problematic. Lewis thought that it showed that similarity was constrained. He suggested that the notion of similarity relevant to the resolution of counterfactuals was subject to restrictions that, evaluated in the actual world, favored worlds that were similar to the actual world in the past. That is, early divergence from the history of the actual world is not permitted. Worlds that diverge early are not amongst the most similar.

Let us accept this conclusion. The similarity relation associated with *would* is evaluated asymmetrically. What happens then with backtracking counterfactuals? Do they really violate the 'usual rules' for resolving similarity? The answer will be 'no'.

3.2 Towards a typology

Philosophers' discussion of backtracking counterfactuals has not always paid much attention to grammatical features. Lewis himself did not attribute particular meaning to the 'special syntax' found in backtracking counterfactuals (though he did note it). From Downing (1959) to Bennett (2003), philosophers have discussed the problems posed by backtracking conditionals without making a distinction between the ones with special syntax and others.

But not all of them. In an early, 1979 paper, Wayne Davis suggested that the special syntax marks a specially modalized consequent. Defending Lewis's asymmetric view, Davis claimed that counterfactuals are never really 'backward looking'. The ones that appear to be backward looking really are not. He presented the following example:

(10) If the plane had arrived at 2:00, it would have to have departed at 1:00.
(Davis 1979)

According to Davis, the consequent clause in (10) is *the plane has to have departed at 1:00*, which is in the present tense. For Davis, the time corresponding to the consequent clause is the present, the time of utterance, at which the obligation is said to hold. Here are Davis's own words:

(11) *The consequent of* [(10)] *is some sort of tensed modal statement. It entails that the plane could have departed at 1:00, and furthermore, that it actually did.*
(Davis 1979)

I will side with Davis in thinking there is a second layer of modality in the consequent clause. I differ with respect to the kind of modality. Davis thinks that truth in the evaluation world is entailed. I believe otherwise, and will return to this in §4.3.1.

3.2.1 Backtracking conditionals that do not need special syntax

As we have seen, judgments about backtracking counterfactuals are slippery. Some examples, however, are judged as 'straightforwardly true'. Even without special syntax:

(12) If Stevenson were President in February 1953, he would have been elected in November 1952. (Bennett 1984)

(13) a. If he were a bachelor, he wouldn't have married.
b. If she had a twin sister, her mother would have had at least two children.
c. If she had sold a horse, she would have owned a horse.

The interesting feature of these examples is that the antecedent and the consequent seem to be related by the very meaning of the words. To be a bachelor is to be somebody who hasn't married. The only way in which this could fail to be the case is if 'bachelor' meant something else. To be a twin is to be one of a (special) pair of children. Etc. Similarly, in Bennett's example, to be the president, at any given point, is to be the person who won the last presidential elections. We wouldn't really call somebody 'president' otherwise (unless the political system changed radically, also changing the very meaning of the word 'president').

Examples like (12-13) above seem to me to fall into the rubric Bennett terms *independent conditionals*. Independent conditionals, whether subjunctive or indicative, are ones *in which the route from A to C owes nothing to any particular contingent fact or belief* (Bennett 2003: 174). Bennett illustrates the difference between independent conditionals and others with examples like these:

(14) *Regular subjunctive*
If the river were to rise another two feet, the subway system would be flooded.

(15) *Independent subjunctive*
If the river were to rise another two feet, it would be two feet higher than it is now.
(Bennett 2003)

In the case of independent conditionals, all *A*(ntedecent)-worlds are *C*(onsequent)-worlds. It isn't really necessary to worry about finding the relevant 'closest' A-worlds. Independent conditionals are not 'variably strict', they are simply 'strict'.

Bennett points out that, not subject to the constraints of closeness, independent subjunctives validate the usual laws of logic that the other subjunctives appear to violate. This is expected if it is the case that all A-worlds are C-worlds, and no selection on the basis of similarity takes place. Here are some backtracking examples:

(16) *Contraposition*
a. If he were a bachelor, he wouldn't have married.
If he had married, he wouldn't be a bachelor.
b. If he were president, he would have been elected.
If he hadn't been elected, he wouldn't be president.

The examples above show that *contraposition* works for backtracking conditionals that do not require 'special syntax'. This makes them somewhat special within the domain of subjunctives. Below is one of Kratzer's illustrations of the failure of contraposition in counterfactuals:

(17) (even) if Goethe hadn't died in 1832, he would still be dead now.
If Goethe were alive now, he would have died in 1832.
(attributed to Kratzer in von Fintel (2001))

The case with strengthening of the antecedent is slightly more complex to spell out. Below are some examples, to get us started:

(18) *Strengthening of the antecedent*
a. If he were a bachelor, he wouldn't have married her.
If he were a bachelor and she were the most beautiful woman in the world, he wouldn't have married her.
b. If he were president, he would have been elected.
If he were president and very poor, he would have been elected.

The examples are a little funny because it is very hard to make sense of the conjunction. After all, it is irrelevant to the claim being made! The same idea can be expressed a little more naturally embedding an extra *if*-clause and emphasizing contrast. The examples seems to work better:

(19) If he were actually a bachelor, even if she were the most beautiful woman in the world, he wouldn't have MARRIED her.

According to Bennett, it isn't necessary to postulate ambiguity between independent conditionals and others. It is simply the case that in independent conditionals the requirement that the consequent be true at the closest antecedent worlds is 'idle':

(20) *Even when A entails C and you know it does, your statement A>C* [the subjunctive conditional] *can perfectly well mean that C obtains at all the closest A-worlds (...). This will be true because in fact C obtains at all the A-worlds, close or not; but there is no good reason to suppose that A>C says that this is so. We can quite well take it as meaning something about the closest A-worlds, though the entailment makes the closeness constraint idle.* (Bennett 2003)

I'll accept Bennett's position, and return to this below.

3.2.2 Backtracking conditionals that are helped by special syntax

Many authors have pointed out that backtracking conditionals are usually/often judged much better with 'special syntax'. We have already seen some examples above (Lewis 1979, Davis 1979). Below is an example from Bennett (1984):

(21) If the die had fallen six uppermost, it would (have to) have been thrown differently. (Bennett 1984)

Here are some further illustrations. The first involves a night guard and an alarm button:

(22) A: It's lucky the guard didn't push the alarm button. It wasn't a serious problem after all!
B: a. Well, he is a very cautious man. If he had pushed the button, something serious would have happened.
b. Well, he is a very cautious man. If he had pushed the button, something serious would have to have happened.

The conditional in (22B-a) is rather deviant. We would hesitate to judge it true. But the conditional in (22B-b) is much more straightforward. It makes us think about the fact that the guard would never push the button unless something serious had happened. It is easier to judge this conditional true.

Another example involves eating habits:

(23) A: She is a very strict vegetarian. If she had eaten pudding, she would have broken her diet.
B: a. No, if she had eaten pudding, it would have been made without gelatin.
b. No, if she had eaten pudding, it would have to have been made without gelatin.

Our first reaction is to say that the conditional in (23B-a) is very odd, maybe even false. But we are more flexible with (23B-b). It overtly asks us to consider the antecedent in situations in which she follows her usual dispositions not to eat animal-related products. It is much easier to judge that conditional true.

The Jim and Jack example discussed by Lewis follows the pattern above:

(24) He (Jim) never would ask for help after such a quarrel; if Jim were to ask Jack for help today, there would have to have been no quarrel yesterday.

The special syntax *there would have to have been no quarrel yesterday* emphasizes that people do not usually ask for help after a quarrel.

It is not easy to play the same tricks on these conditionals that we played on the ones in the previous section. Some examples do indicate that *strengthening of the antecedent* does not hold:

(25) a. If Jim asked Jack for help, there would have to have been no quarrel yesterday.
b. If Jim asked Jack for help and he did not worry about rejection, there would have to have been no quarrel yesterday.

The example is clumsy, but it seems to go against Strengthening of the Antecedent. (25a) could well be true and (25b) false. Similarly:

(26) a. If the guard had pushed the button, something serious would have to have happened.
b. If the guard had been drunk and pushed the button, something serious would have to have happened.

We could well judge (26a) true and (26b) false. So strengthening of the antecedent fails.

The case of contraposition is difficult to test, as special syntax in the antecedent clause is hard to make sense of:

(27) If it were false that there would have to have been no quarrel yesterday, it would be false that Jim asked Jack for help.

Intuitions are not readily accessible, but it seems to be the case that (25a) can be true while (27) is false. So contraposition appears not to work.

According to my descriptions, the presence of special syntax in the examples above invokes generalizations or rules. I will elaborate on this in §4 and §5.

3.2.3 Backtracking conditionals that are not helped by special syntax

In many cases, we are willing to judge a backtracking conditional with special syntax true where we would have judged the simple backtracking conditional false (or 'undecided'). But special syntax does not deliver truth in all cases. Intuitions, of course, are slippery. Here is an example in which special syntax does not seem to get the job done:

(28) The bridge wasn't completed, and the distracted driver came to a sudden stop.
a. If the driver had kept going, the bridge would have been completed.
b. If the driver had kept going, the bridge would have to have been completed.

Both conditionals in (28) are odd, and not straightforwardly true. This is not surprising in the case of (28a), since this is clearly not an 'independent' conditional. But it is interesting to see that the special syntax in (28b) is not immediately convincing. The distracted driver could well have kept going and fallen off the half-finished bridge. We judge (28b) false.

One could imagine that the difference between the cases in which special syntax delivers truth and the others has to do with the strength of quantification over worlds. Could it be that in the cases in which special syntax does not result in truth, universal

quantification over worlds is too strong? Here is an example that seems to support this view:

(29) She loves desserts in general, but she doesn't like chocolate. She didn't even touch the chocolate mousse.
  a. If she had eaten dessert today, the cook would have made a peach pie yesterday.
  b. If she had eaten dessert today, the cook would have to have made a peach pie yesterday.

We judge both conditionals in (29) false. The special syntax in (29b) does not help us to arrive to a past in which the cook has made a peach pie. And, arguably, this is because the peach pie is one option amongst many. So, conceivably, it could be that (29b) is false because universal quantification over worlds is too strong. However, I do not think that this is what is going on. We fail to arrive to truth even if we change *would* for a weaker quantifier:

(30) She loves desserts in general, but she doesn't like chocolate. She didn't even touch the chocolate mousse. If she had eaten dessert today, the cook might have to have made a peach pie.

The conditional in (30) does not have more chances of being true than the on in (29b). In (30) we actually seem to be dealing with a different type of modality in the consequent clause and whatever difference in quantificational strength there may be doesn't seem to help.

3.2.4 Summary

Our observations lead us to two generalizations. One has to do with backtracking *would*-conditionals that are judged to be straightforwardly true without special syntax:

(31) **Generalization 1:** Backtracking *would*-conditionals without special syntax can be judged straightforwardly true only if they are independent.

I hesitate to talk about these examples as backtracking 'counterfactuals'. They are counterfactuals in the sense that their antecedents are not true in the actual world. But their truth does not depend on actual world facts. They are not contingent. Do they really count as counterexamples to Lewis's idea that when we evaluate similarity we favor worlds that are as much as possible like the actual world in the past? I don't think so. In these cases, we don't have any choice.

The other generalization has to do with the conditions in which special syntax can lead us to judge a backtracking *would*-conditional true. We have seen that special syntax does not help in all cases:

(32) **Generalization 2:** Backtracking *would*-conditionals with special syntax can be judged straightforwardly true only if there is some

salient generalization/ law relating antecedent and consequent.

In (32) I've set out a hypothesis about the nature of the connection between the antecedent and consequent. I will elaborate on that in the following sections.

4. On the kind of modality

What kind of linking between antecedent and consequent clause helps backtracking conditionals with special syntax? We have hypothesized that the extra layer of auxiliaries introduces another modality in the consequent clause. What kind of modality is it?

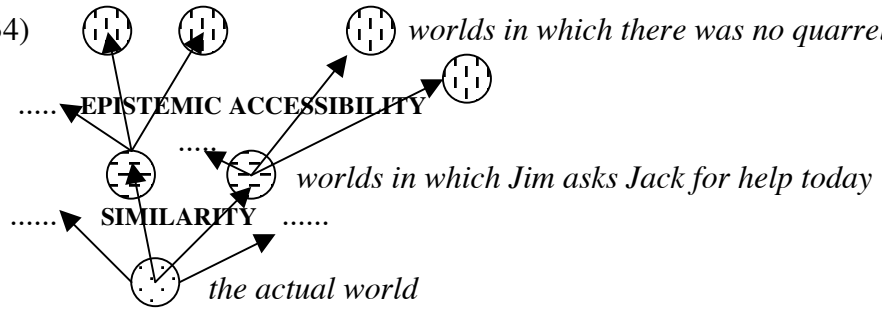
4.1 It is not epistemic modality

As a first-attempt, let us consider epistemic modality. Could *have to* in (33) be epistemic? What would that mean?:

(33) If Jim had asked Jack for help today, there would have to have been no quarrel yesterday.

The antecedent clause takes us to the most similar worlds in which Jim asked Jack for help today. Under the hypothesis that the modality introduced by the consequent is epistemic, the conditional would then claim that in the worlds that are epistemically accessible from such antecedent worlds it would be the case that there was no quarrel yesterday.

(34)



The idea that *have to* can be interpreted with respect to an epistemic accessibility relation is not itself astonishing. There are 'evidential' uses of *have to* that plausibly could be characterized as epistemic:

(35) Her briefcase is on the table. She has to be in the building somewhere.

Is the *have to* that we find in backtracking conditionals interpreted in this way? Or in some other 'epistemic' way? [5] The answer, I think, is 'no'. *Have to* in backtracking counterfactuals is not interpreted in relation to our knowledge (or evidence). In an

[5] I am going to ignore here the fact that it is usually accepted that evidential or epistemic modals are found rather 'high', and so there may be syntactic reasons why they are not embedded.

epistemic reading of *have to* in (33), the conditional would claim that in the most similar worlds in which Jim asks Jack for help today, the worlds that are made accessible, given our knowledge now, are worlds in which there was no quarrel yesterday. But of course, this doesn't really make sense. Our knowledge now includes that Jim did not ask Jack for help today and that there was a quarrel yesterday. If we wonder about which worlds would be accessible from the most similar worlds in which Jim asked Jack for help today given our knowledge now, we would not end up with worlds in which there was no quarrel yesterday.

You might say, but clearly we do not want to consider the worlds compatible with our knowledge now! We need to consider the worlds compatible with what our knowledge would have been if Jim had asked Jack for help today. If we shift the anchoring of the epistemic accessibility relation in this way, the conditional in (33) tells us that if Jim had asked Jack for help today, we would have known that there had been no quarrel yesterday. One possible objection to this view is that, even if Jim had asked Jack for help today, nothing guarantees that we would have known about the request itself (remember that the epistemic modal is embedded in the consequent clause of the conditional, it is not the main operator). And if we hadn't known that Jim asked for help, we wouldn't have concluded that there had been no quarrel yesterday. However, for reasons that will be clear later, I will set aside this possible objection. In my opinion, the main objection to this view is the following: consider the most similar worlds in which Jim has asked Jack for help today as we have been conceiving them (accepting that we know that Jim asked Jack for help). In such (similar) worlds, there was a quarrel yesterday. Why would we not know about it? The development of our knowledge is a fairly accidental (and accidented) thing. It is quite plausible that in some of the minimally different worlds in which Jim asked Jack for help today (and we know about it), we would also know that there was a quarrel yesterday. In such a situation, we would not conclude that it must have been the case that there wasn't a quarrel yesterday. Our knowledge would be inconsistent.

I conclude that *have to* in (33) is not interpreted epistemically, neither with respect to our actual knowledge nor with respect to our hypothetical knowledge in the minimally different circumstances in which the antecedent is true.

4.2 The 'best explanation' account

Bennett (2003) dismissed Davis's ideas about the role of *have to*, and proposed instead a 'best explanation' view. I will turn to Bennett's objections to Davis in §4.3.1. Here I will briefly review the 'best explanation' view:

(36) *This account (Davis's) misunderstands the force of the modal element in the consequent of a backward subjunctive. Its real point is to register the thought that (C) the consequent is acceptable as the best explanation for (A) the antecedent. If the plane arrived at 2 p.m., that would have to have been because it left at noon - the modal 'have to' expresses the compulsion in our being forced to that explanation for lack of any as good.* (Bennett 2003)

Even though he adopts the proposal also for indicatives and subjunctives, Bennett's discussion of the 'best explanation' idea is centered solely on indicatives. He provides illustrations like:

(37) If my umbrella is not in the coat-closet, I must have taken it to campus this morning. (Bennett 2003)

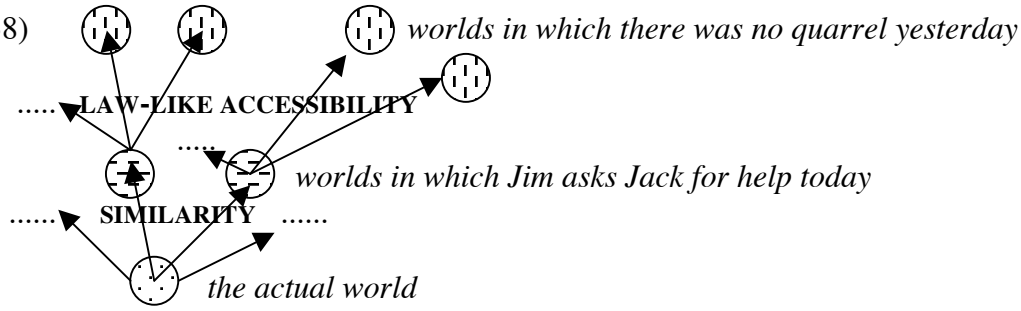
According to Bennett, *must* in examples like (37) expresses a sense of being forced to accept the consequent as the best explanation for the antecedent. He suggests that *have to* in (33) has the same kind of interpretation as *must* in (37).

Bennett doesn't present a formal discussion of the 'best explanation' idea. Indeed, he doesn't say much about it with respect to backtracking subjunctives. The notion of 'explanation' is not elaborated upon, so it is difficult to evaluate. If the 'best explanation' idea were to be reduced to epistemic modality of some kind, in the context of backtracking subjunctives it would run into the problems we already discussed in §4.1. On the other hand, it is also unclear that the notion of 'explanation' that is rather intuitive in discussing indicative conditionals can simply be transposed to counterfactuals. There is a sense in which one could intuitively say that the consequent in (37) explains the antecedent. But can we say the same thing in (33)? Is the antecedent of a counterfactual something that is 'explained' in the same way? In my mind, this is not clear.

4.3 Laws and regularities

There are many things that can be taken into account when evaluating counterfactuals. A broad distinction can be made between similarity with respect to matters of fact and with respect to matters of law (Bennett 2003). Lewis (1979) pointed out that matters of law are not necessarily given highest priority when evaluating counterfactuals. The standard way of resolving similarity gives more importance to similarities of fact (at times before the antecedent clause time) than to strict adherence to the laws. In Lewis's terms, it is more important to ensure *spatio-temporal match* than to avoid *small, localized miracles*. In the story here, the role of the special syntax is to shift the emphasis back to the laws. When there is a second layer of auxiliaries in the consequent clause, the claim made by the conditional is about what the laws would have predicted under the circumstances described by the antecedent. The picture now looks as follows:

(38)



A backtracking counterfactual with special syntax claims that in the most similar worlds in which the antecedent is true, the worlds made accessible by (salient) law/s or generalizations are worlds in which the consequent is true.

4.3.1 Properties of law-like accessibility

Edgington (1997) and Bennett (2003) have raised objections to accounts that propose that the consequent clause in backtrackers with special syntax makes a modalized claim and does not require real backtracking: Davis (1979) and Woods (1997). In their (brief) discussion of backtracking conditionals with special syntax, both Davis and Woods claimed that they were not really backtracking conditionals. Here are Woods's words:

(39) *But the effect of such forms of words is to make the conditional not a back-tracking conditional at all. Such sentences do not say that if something had been the case at one time, something else would have been the case earlier; they say that if something had been the case, it would (then) have had a certain explanation, or that (then) something would have needed to have happened earlier*. (Woods 1997)

Edgington (1997) objects to Woods's characterization, pointing out that if something is to be explained by P, then P had better be true. Similarly, objecting to Davis, Bennett (2003) says:

(40) *A world at which*
It is the case at T2 that: it must have been the case that P at T1
is a world at which
It is the case at T1 that: P (Bennett 2003)

Both Edgington and Bennett worry that a 'modalized consequent' account of counterfactuals with special syntax predicts that the equivalent counterfactuals without special syntax should also be true. Given their description of the effects of the special syntax, the prediction is that in the nearest worlds in which the antecedent is true, the non-modal version of the consequent clause claim should also be true. The past would have to be changed, contra Lewis (1979).

However, I believe Bennett's and Edgington's concern is not well founded with respect to my proposal. I claim that *have to* in the consequent clause does indeed make a modalized claim: it invokes what would have been the case according to laws/ generalizations. One of the peculiarities of law-like modality is that it is not 'reflexive'. That is, something may be necessary in a world, given laws or generalizations, without being true. Law-like modality differs from logical modality and epistemic modality, for example, in this respect. If something is necessary given what we know, it is also true. And if something is logically necessary, it is also true. The claim that *have to* in the consequent invokes law-like modality, and not other kinds of modality, has important consequences. When we claim that the consequent clause proposition is true in the law-like worlds accessible from the most similar worlds in which the antecedent clause proposition is true, we do not commit ourselves to the claim that the consequent clause proposition is true in the most similar worlds in which the antecedent clause proposition is true. The worlds quantified over by *would* (those made accessible by similarity) need not be worlds where the laws/ regularities are obeyed. The past doesn't have to be changed.

4.2 On the relation between antecedent and consequent

I have proposed an analysis of backtracking conditionals in which two modals interact with each other. In this section I show that the antecedent clause affects the interpretation of both modals independently, providing yet another argument that the relationship between modals and restrictors should be captured 'dynamically', and not (simply) on the basis of syntactic positioning.

Consider one of our old examples:

(41) If she had eaten the pudding, it would have to have been made without gelatin.

It seems pretty obvious that the *if*-clause in this example restricts the domain of quantification of *would*. Still, can we actually show that it does? What would happen if it didn't? It has been noted that *would* is always interpreted with respect to restrictions, somehow recovered from the context (see e.g. Kaspers (1992), and more recently, Veltman (2005)). So even if the *if*-clause didn't restrict *would*, we'd need another restriction. Consider (42):

(42) (Suppose she didn't come home early, she didn't have dinner with us, and she didn't have dessert. We hypothesize:) If she had come home early, she would have eaten dinner with us. But if she had eaten dessert, it would have to have been made without gelatin.

Could we interpret (42) as in (43)?

(43) [would/ $_{\text{if she had come home early and had eaten dinner with us}}$ [have-to/ $_{\text{if she had eaten dessert}}$ [the dessert had been made without gelatin]]]

In this interpretation, *would* is restricted by the extra-sentential context, and the *if*-clause *if she had eaten dessert* only provides a restriction to the modal *have to*. It seems perverse. But is it possible?

Intuitions are tricky. According to (43), *would* takes us to the nearest worlds in which she comes home early and has dinner (in which she presumably did not have dessert), and *have to* takes us to the nearest law-like worlds in which she also had dessert. But this can't be right. *Have to* is not happy with 'counterfactual' restrictions. Look at (44):

(44) a. She is not smart. If she were smart, she would be happy.
b. I'm not sure about her smartness. If she was smart, she had to be happy.
c. She is not smart. #If she were/was smart, she had to be happy.

We can easily understand (44a) as a counterfactual. She is not smart, but if she were, she would be happy. The example in (44b) shows that it is possible to link her happiness to her smartness in *have*-to indicative conditionals (note the epistemic flavor). But, as (44c) shows, it is not possible to 'counterfactually' link her happiness to her smartness with *have-to* conditionals.

The examples in (44) teach us that we cannot use modal *have-to* to make a counterfactual hypothesis.[6] This means that the *if*-clause in (41) has to modify *would*: *would* is responsible for taking us to 'counterfactual' worlds in which she eats pudding. The conditional then asserts that in all such worlds that obey the law, pudding was made without gelatin. The compositional interaction between the modals is crucial to understand the interpretation of the conditional.

The point made laboriously above was rather intuitive. But what about the idea that the *if*-clause restricts *have-to*? This would be unexpected if the restriction relation was necessarily mediated by syntax. After all, the *if*-clause is not sister to the phrase headed by *have-to*. Do we interpret (41) as in (45)?

(45) [would$_{\text{/she ate pudding}}$ [have-to$_{\text{/laws}}$ [have$_{\text{perf}}$ [the pudding been made without gelatin]]]]

According to (45), the *if*-clause associates with *would*, taking us to the most similar worlds in which she ate pudding. The consequent clause then tells us that in the law-like worlds accessible from there, the pudding had been made without gelatin.

Our intuitions indicate that (41) is not simply interpreted as (45). There are two sorts of law-like worlds: worlds in which she ate the pudding and it had been made without gelatin, and worlds in which the pudding had been made with gelatin but she did not eat it. The fact that we do not consider the latter case, and ignore law-like worlds in which she did not eat the gelatinous pudding indicates that the *if*-clause restriction also affects the interpretation of the modal *have-to*.

What conclusion should we draw from this? In figuring out how the interpretation works, we need to allow for a mechanism that allows the interaction between antecedent and consequent to be 'dynamic'. There are several options here, ranging from the inclusion of a free variable in the syntactic representation of the restrictor of the modal (inspired by von Fintel (1994)), to fully dynamic systems (a.o. Kratzer (1981), von Fintel (2001), and Veltman (2005)'s recent proposal about making counterfactual assumptions). I will not make a commitment, but simply note the issue.

5. On matters of internal composition

We finally turn to more fine-grained details, and address the issue of compositionality. The relation between the *if*-clause and the main clause in conditionals, as well as the internal syntactic make up, have been the subject of much debate (see Iatridou (1991) for an early syntactic analysis, and Bhatt and Pancheva (2001) for an overview of syntactic issues, as well as comparisons with correlative structures). The compositionality issues have also been explored in the semantic literature, where the tension between static and dynamic aspects of interpretation have often surfaced (see references above). I won't attempt to deal with all the issues here, and will make several simplifying assumptions, driven by the need to focus on the interaction between modals.

I assume that conditionals are headed by modals. I will ignore the possibility that there is syntactic structure above the modal (but see a.o. Condoravdi (2002) for an account that ties the interpretation of modals to their relative position with respect to tense; also, in Arregui (2004) I argued that modals are embedded under tense, and linked

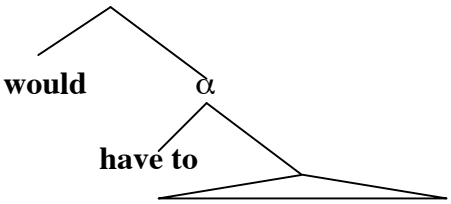
[6] This observation leads to many questions, which I cannot address here. Let me simply note that the restriction on *have to* does not follow from the proposed semantics.

this to the semantics of counterfactuals, but I simplify here). In the case of backtracking conditionals with special syntax, the main modal *would* embeds the modal *have-to*. I make no commitments about the syntactic relation between the *if*-clause and the rest of the structure. In fact, I will leave the *if*-clause out of the representations all together, and take for granted its effect on the interpretation of the modals (see below).

Following a very general convention, I treat modals as restricted quantifiers over possible worlds. I will assume that the *if*-clause, indirectly, restricts the quantificational domain of the modals. The idea that *if*-clauses restrict operators can be found in Lewis (1975) and Kratzer (1977), and has been widely assumed since (though concerns have been raised, eg. Higginbotham (2003)).
With these pieces in place, we see in (47) the consequent clause of (46):

(46) If she had eaten pudding, it would have to have been made without gelatin.

(47)



The modals in (47) are both universal, but differ with respect to the kinds of restrictions they permit. The accessibility relations that determine their domain of quantification are different. Following a Lewis-Stalnaker view, *would* is restricted by a similarity-based accessibility relation. *Have to* is restricted by a law-like accessibility relation. The interpretation of both modals is affected by the antecedent clause proposition. Details of the composition are spelled out below:

(48) [[**would** α]] =
λq λw ∀w' (SIM (w)(w')(λw". she had eaten pudding in w") → q(w') ([[α]])

Where SIM(w)(w')(λw".she had eaten pudding in w") = 1 iff she had eaten pudding in w' and there is no w"' in which she had eaten pudding such that w"' is more similar to w than w' is similar to w.

According to (48) the similarity relation that restricts the modal identifies the worlds in the quantificational domain as the most similar worlds in which the antecedent is true (There are several assumptions in place here: the limit assumptions, the possibility that worlds may tie for similarity, etc. They are not crucial to what is being said, but they simplify the presentation.).

The denotation of the clause headed by the law-like modal is given below:

(49) [[α]] = λw ∀w' (LAW(w)(w')(λw". she had eaten pudding in w") →
the pudding was made without gelatin in w')

Where LAW(w)(w')(λw".she had eaten pudding in w") = 1 iff she had eaten pudding in w' and there is no w'" in which she had eaten pudding such that more of the laws of w are satisfied in w'" than in w'.

According to (49) the law-like accessibility relation that restricts the modal identifies the worlds in the quantificational domain as the most law-like worlds in which the antecedent is true. This lexical entry clearly follows Kratzer's ideas that in the case of law-like modality, the laws serve to order the worlds in which the antecedent is true, and the modal quantifies over the 'best' (most law-like) such worlds (a.o. Kratzer (1981), (1991)).

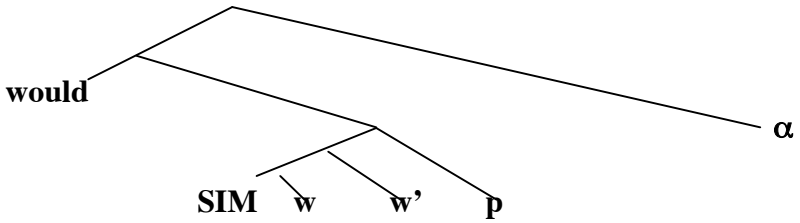
Once *would* combines with its sister, the denotation of (47) is as in (50):

(50) λw ∀w' [SIM (w)(w')(λw". she ate pudding in w") →
∀w'" (LAW (w')(w'")(λw". she ate pudding in w") →
the pudding was made without gelatin in w'")]

This is the proposition that is true in a world only if the most similar worlds in which she ate pudding are worlds in which the law requires that it have been made without gelatin.

In the proposal I have presented here, there is no syntactic representation of accessibility relations, world parameters or antecedent clauses. Yet this could have been done differently. We could, for example, have taken the syntactic structure associated with *would* to be (51):

(51)



Recent work in modality at times makes the assumption that accessibility relations are syntactically represented (see, for example, Ippolito (2003)). I have simplified the structures only because there is nothing in my current topic that bears directly on that issue. One of the things that seems relatively clear, however, is that the world parameter corresponding to the evaluation world should remain 'available':

(52) Every boy believed that if he had married a girl, his mother would have been happy.

We can interpret (52) as the claim that every boy self-ascribes the property of being in a world *w* such that the most similar worlds to *w* in which he marries a girl are worlds in which his mother is happy. Whichever way we set up the syntax-semantics interface, we need to allow for similarity to be evaluated with respect to different worlds.

One of the advantages of having syntactic representations of the components of the restrictor on the modal is that they provide a straightforward way of capturing the relation that both modals have with the antecedent clause proposition. Given the syntactic

representation of the restrictor, we can include a free variable that has as contextually assigned denotation the proposition corresponding to the antecedent clause. Both modals can include such a variable in the syntactic representation of their restriction, providing us with an easy way of accounting for the fact that the antecedent clause proposition restricts both modals. However, the 'free variable' mechanism is not the only one available to model the effects of context on the interpretation of modals. A fully dynamic account of the interpretation of modals could also provide a suitable explanation for the interaction without necessarily including a syntactic representation of the restrictor. For this reason, I have made no commitments.

6. Conclusion

The main contribution made by the discussion of backtracking counterfactuals has been to understand the interaction between the modal auxiliaries in the consequent clause. I have proposed an analysis of backtracking counterfactuals with special syntax that treats them as a case of layered modalities. The interpretation of the first modal (*would*) follows a Lewis-Stalnaker-style analysis for counterfactuals. The second modal (*have to*) introduces law-like modality in the consequent clause. In understanding the relationships between modals, we have to ask ourselves which combinations of modalities are possible. As we have seen, it is possible to make sense of law-like modality embedded in counterfactuals, but it is harder to make sense of epistemic modality embedded in counterfactuals. The idea that there are restrictions on the combinatorics of modalities that come from plausibility seems rather interesting.

Much of our discussion has been aimed at distinguishing the properties of the different accessibility relations associated with the modals. Lewis (1979) placed a series of restrictions on the similarity relation that identifies the quantificational domain of counterfactuals, resulting in a relation that favored worlds that were, as much as possible, like the actual world in the past. Whether we agree with his implementation or not, there were good reasons for those restrictions. As we have seen, however, not all kinds of modality grant the past special status. Law-like modality, which is not based on similarity relations, gives the modal access to worlds that differ from the evaluation world in the past.

Our conclusions about backtracking counterfactuals have not contradicted Lewis's views on how we evaluate similarity in counterfactuals. In the case of backtrackers with special syntax, the modals that lead us to worlds that differ from the actual world in the past are not similarity-based modals. And the straightforward cases of backtracking counterfactuals with regular syntax do not really argue against Lewis either. Lewis's restrictions give us a recipe for choosing, amongst worlds in which the antecedent is true, those that are most similar to the actual world. Preference is given to worlds that are like the actual world in the past as much as possible. In the case of backtrackers with regular syntax that are judged straightforwardly true, all worlds that satisfy the antecedent are worlds that differ from the actual world in the past. There is no choice to be made. We cannot really say that Lewis's recipe has gone wrong. The idea that the past is important remains firmly in place.

References

Arregui, Ana. (2004) *On the accessibility of possible worlds: the role of tense and aspect*, PhD dissertation, University of Massachusetts, Amherst.

Bennett, Jonathan. (1984) 'Counterfactuals and temporal direction', *Philosophical Review* 93: 57-91.

Bennett, Jonathan. (2003) *A philosophical guide to conditionals*, Clarendon Press, Oxford.

Bhatt, Rajesh and Roumyana Pancheva: 2001, 'Conditionals', ms. University of Texas, Austin and USC.

Condoravdi, Cleo. (2002) 'Temporal interpretation of modals', in D. Beaver, L. Martinez, B. Clark & S. Kaufmann (eds.), *The Construction of Meaning*. CSLI Publications. Palo Alto, 59–88.

Davis, Wayne, 1979, 'Indicative and subjunctive conditionals', *Philosophical Review* 88: 544-564.

Downing, P.B. (1959) 'Subjunctive Conditionals, Time Order, and Causation', *Proceedings of the Aristotelian Society* 59: 125-140.

Edgington, Dorothy. (1997) 'Commentary', in Woods (1997): 95-137.

Fine, Kit. (1975) 'Review of Lewis 1973', *Mind* 84, 451-458.

Fintel, Kai von. (1994) *Restrictions on quantifier domains*, PhD dissertation, University of Massachusetts, Amherst.

Fintel, Kai von. (2001) 'Counterfactuals in a Dynamic Context', in Kenstowicz, M. (ed.), *Ken Hale: A Life in Language*, Cambridge, MIT Press, 132–152.

Higginbotham, James. (2003) 'Conditionals and compositionality', in Hawthorne, J. (ed.) *Philosophical Perspectives* 17, *Language and Philosophical Linguistics*, 181-193.

Iatridou, Sabine. (1991) *Topics in Conditionals*, PhD dissertation, MIT.

Ippolito, Michela. (2003) 'Presuppositions and Implicatures in Counterfactuals', *Natural Language Semantics* 11, 145-186.

Kasper, Walter. (1992) 'Presupposition, Composition, and Simple Subjunctives', *Journal of Semantics* 9: 307-331.

Kratzer, Angelika. (1977) 'What *must* and *can* must and can mean', *Linguistics and Philosophy* 1, 337-355.

Kratzer, Angelika. (1981) 'The Notional Category of Modality', in Hans-Jürgen Eikmeyer and Hannes Rieser (eds.), *Words, Worlds, and Contexts*, De Gruyter, Berlin, 38–74.

Kratzer, Angelika. (1991) 'Modality', in Arnim von Stechow and Dieter Wunderlich (eds.), *An International Handbook of Contemporary Research*, De Gruyter, Berlin, 639–650.

Lewis, David. (1973) *Counterfactuals*. Harvard University Press, Cambridge, Mass.

Lewis (1975) 'Adverbs of quantification', in Keenan, E. (ed.), *Formal Semantics of Natural Language*, Cambridge University Press, Cambridge.

Lewis, David. (1979) 'Counterfactual Dependence and Time's Arrow', *Noûs* 13, 455–476. Reprinted in D. Lewis (1986), *Philosophical Papers*, Vol. 2, Oxford University Press, Oxford.

Ogihara, Toshiyuki. (2000) 'Counterfactuals, Temporal Adverbs, and Association with Focus', in *Proceedings of SALT 10*, CLS Publications, Cornell University, Ithaca, N.Y. 115–131.

Schulz, Katrin (forthcoming). “If you wiggle A, the B will change”: Causality and the Meaning of Conditional Sentences. (Synthese).
Stalnaker, Robert. (1968) ‘A Theory of Conditionals’, in Nicholas Rescher (ed.), *Studies in Logical Theory 2*, Blackwell, Oxford, 98-122.
Stowell, Tim. (2004) 'Tense and Modals' in Jacqueline Guéron and Julie Lecarme, (eds.) *The Syntax of Time*, MIT Press, Cambridge.
Veltman, Frank. (2005) 'Making Counterfactual Assumptions', *Journal of Semantics* 22:2, 159-180.
Woods, Michel. (1997) *Conditionals*, Wiggins, D. (ed.), Oxford University Press.

Ana Arregui
Linguistics Department
University of Ottawa
70 Laurier Ave. East
Arts 423
Ottawa, ON K1N 6N5
Canada