

Counterfactuals and Conditional Questions under Discussion *

Michela Ippolito
University of Toronto

Abstract In this paper, I explore the issue of the context-dependence of counterfactual conditionals and how the context constrains similarity in selecting the correct set of worlds necessary to arrive at the correct truth-conditions. I review the previous literature and make a proposal based on the idea of discourse structure and the concept of a question under discussion.

Keywords: Counterfactuals, similarity, possible worlds, questions under discussion.

1 The puzzle of similarity

According to Lewis (1973), a conditional of the form *If it had been the case that ϕ , it would have been the case that ψ* is true in a world w just in case the consequent is true in all the ϕ -worlds that are (otherwise) maximally similar to w , as shown in (1).

- (1) $\phi \Box \rightarrow \psi$ is true at a world w (according to a given comparative similarity system) if and only if either (a) no ϕ -world belongs to S_w (the set of worlds accessible from w), or (b) there is a ϕ -world w' in S_w such that, for any world w'' , if $w'' \leq_w w'$ then $\phi \rightarrow \psi$ (material implication) holds at w'' .

The notion of comparative similarity, represented formally by the relation \leq_w and spelled out as shown in (2), is crucial to these truth-conditions in that it determines the set of ϕ -worlds in which the consequent ψ is required to be true.

- (2) $w' \leq_w w''$ means the world w' is at least as similar to the world w as the world w'' .

Spelling out exactly which worlds are most similar to the evaluation world turns out to be a very difficult task. In most cases, we have very clear intuitions about the truth or falsity of these conditionals, yet spelling out exactly the measure of similarity that is needed to account for our intuitions has turned out to be one of the most difficult problems for both philosophers and linguists. To appreciate the puzzle,

* For helpful discussions, I would like to thank Donka Farkas, and audiences at the University of Toronto and Università degli Studi di Milano-Bicocca. Mistakes and omissions are mine.

consider Jones and the rain example from Tichy (1976). Jones always wears his hat if the weather is bad. If the weather is good, Jones wears his hat at random. Today the weather is bad and Jones is wearing his hat. In this context, suppose someone were to utter the counterfactual in (3).

(3) If the weather had been fine, Jones would be wearing his hat.

We judge (3) false, which means that in selecting counterfactual worlds in which the weather is fine that are otherwise maximally similar to the actual world, we disregard the fact that Jones is wearing his hat. The list of priorities in Lewis (1979), according to which similarity of laws trumps similarity of particular facts, can account for the judgment in (3): assuming determinism, the worlds we select are those worlds that shared the same history as the actual world up to the divergence time, i.e. the time when (thanks to a “miracle”) the deterministic chain of events broke and these worlds took different paths following the actual laws. That is, when selecting the most similar worlds we need to select those worlds that are just like the actual world up until they diverge from the actual world, but that follow their own course afterwards. Applied to Tichy’s example, these worlds are going to be worlds that are just like the actual world up to the time when some miracle breaks the deterministic chain of events and the weather turns out to be fine but which, after that, follow undisturbed the actual laws. We should not try to make these worlds converge again just for the sake of maximizing the number of particular facts in common with the actual world since this would involve more “miracles” or inexplicable violations of the actual laws. Thus, since the actual laws say that if the weather is fine Jones might or might not wear his hat, the conditional in (3) comes out false.

However, things are not so simple. Consider a variant of Tichy’s example from Veltman (2005). Every morning Jones tosses a coin. If heads comes up and the weather is fine, then he wears his hat. If the weather is bad, Jones always wears his hat (regardless of the outcome of the coin-tossing). Today the weather is bad, heads came up, and Jones is wearing his hat. In this context we judge (3), repeated in (4), true.

(4) If the weather had been fine, Jones would be wearing his hat.

This time, we judge this counterfactual true. Lewis’s system of priorities requires that we select those worlds that after the divergence proceed according to the actual laws. Since the outcome of a coin tossing is probabilistic, there are going to be worlds where heads comes up and worlds where tails comes up. Hence, the conditional in (4) should be true. One might elaborate on Lewis’s story by saying that when selecting the worlds most similar to the actual world, we want to maximize similarity to the actual world even after the divergence: we keep as many facts true in the

actual worlds as possible unless they make the actual laws “vacuous”. This is indeed the proposal in [Arregui 2009](#). Arregui’s example is given in (5).

- (5) Peter presses the button in a completely random coin-tossing device and the coin comes up heads. If Susan had pressed the button, the coin would have come up heads.

This counterfactual is judged false. According to Arregui, what explains this is that we cannot keep the fact that heads came up because if we did, we would “render the laws set in motion by the counterfactual “vacuous”” ([Arregui 2009](#)). What Arregui means is that the outcome of Susan’s coin tossing is up to the probabilistic laws that regulate this type of physical event. Already assuming the outcome of Susan’s coin tossing makes the laws “vacuous”.

Applied to the Jones examples above, this idea would explain why we cannot retain the fact that Jones is wearing his hat in Tichy’s example but we do retain the fact that heads came up in Veltman’s example: in worlds in which the weather is fine, and Jones tosses a coin, assuming the outcome of the coin tossing (heads) would make the probabilistic laws at work in those worlds vacuous, just like in Arregui’s case. Hence, since we cannot assume that heads came up, it does not follow from the supposition that the weather is fine together with the laws of nature that Jones would be wearing his hat.

While interesting, this story is not right. To see this consider the variant of Arregui’s example given in (6), from [Ippolito 2013](#).

- (6) Peter and Susan are taking turns at pressing a button on a completely random coin-tossing device. They both bet each time one presses the button, but (as part of their game) only the one actually pressing the button pays \$10 if he or she loses. It’s Peter’s turn to press the button. Peter bets that the coin will come up heads, Susan bets that it will come up tails. Peter tosses the coin and heads comes up. Peter wins. Susan had bet on tails but since she wasn’t the one pressing the button she does not have to pay \$10. Now I say: Susan, you’re lucky! If it had been your turn to press the button, you would have lost \$10.

Unlike (5), we judge (6) true and the reason is that, unlike in Arregui’s example, we are allowed to keep the fact that the coin came up heads, despite the fact that this does make the probabilistic laws of nature we are assuming “vacuous”. I conclude that the problem of explaining how we measure similarity across worlds remains unsolved.

One influential way of analyzing counterfactuals goes under the name of Premise Semantics and goes back to [Ramsey \(1929\)](#) and [Goodman \(1947\)](#), among others. There are different varieties of Premise Semantics, but the basic idea which is

shared by all of them is that a counterfactual *if* ϕ , *would* ψ is true in w just in case ψ follows from ϕ together with a “suitable” set of premises, where the premises are propositions true in the actual world. Just like in other versions of Premise Semantics, Veltman (2005) proposes that, when accommodating the counterfactual antecedent ϕ , we make the minimal retraction from the premise set. What is special is that in Veltman’s proposal the premise set is a set of facts constituting what he calls a *basis* of a world. A basis of a world w is “a part of w consisting of mutually independent facts which, given the general laws, bring the other facts constituting w in their train.” When a fact is retracted from the basis, all the facts that “it takes in its train” are retracted as well. As Veltman himself acknowledges, his proposal runs into a problem with Kratzer’s King Ludwig of Bavaria example from Kratzer 1989. Here is the context. King Ludwig of Bavaria likes to spend his weekends at Leoni Castle. Whenever the Royal Bavarian flag is up and the lights are on, the king is in the castle. At the moment, the lights are on, the flag is down, and the king is away. Suppose counterfactually that the flag were up.

- (7) a. If the flag were up, the king would be in the castle.
 b. If the flag were up, the lights would be off.

We judge (7a), but not (7b), true. The three propositions in question are: *that the flag is up* (p); *that the lights are on* (q); *that the king is out* (r). Here is what we assume to be true: $\Box((p \wedge q) \rightarrow \neg r)$, $\neg p$, q , r . The basis of the world that is consistent with these assumptions contains the propositions q and r , which are both true. Now, in order to accommodate the counterfactual antecedent p , we have two options: we can remove either q or r from the basis. That is, we could either remove the fact that the lights are on or the fact that the king is away. However, the contrast between (7a) and (7b) tells us that the option of removing the fact that the lights are on (q) is not available.

Kratzer (1989) also proposes a version of Premise Semantics, this time based on the notion of *lumping*. Recall the moral of the dialogue with the lunatic from Kratzer (1989): when Paula paints a still life that depicts apples and bananas, the proposition that she painted a still life lumps the proposition that she painted apples and bananas.¹

Let us see how the lumping machine accounts for the King Ludwig’s example. The propositions involved are: (a) whenever the flag is up and the lights are on,

¹ (i) Here is the technical definition of lumping. S is the set of possible situations and $\wp(S)$ is the power set of S , i.e. the set of propositions.

- (1) For all propositions p and $q \in \wp(S)$ and all $w \in W$: p lumps q in w iff the following conditions hold:
 (i) $w \in p$;
 (ii) for all $s \in S$, if $s \leq w$ and $s \in p$, then $s \in q$.

the King is in the castle; (b) the flag is down; (c) the lights are on; (d) the king is away; (e) the flag is up. The proposition in (e) is the counterfactual antecedent and it must be included in the set of propositions relevant for the truth-conditions of the counterfactual. The question is: which propositions among the ones listed above must be removed in order to accommodate (e)? As we saw above, there are in principle two possibilities.

(8) **Possibility (i):**

- (e) the flag is up
- (a) whenever the flag is up and the lights are on, the king is in the castle
- (c) the lights are on

Possibility (ii):

- (e) the flag is up
- (a) whenever the flag is up and the lights are on, the king is in the castle
- (d) the king is away

According to possibility (ii), we should be able to remove the propositions that the lights are on and keep the proposition that the king is away. If this possibility were available, the counterfactual in (7b), repeated below, would incorrectly be predicted to be true.

(9) If the flag were up, the lights would be off.

Fortunately, Kratzer's lumping can rule out possibility (ii), as follows. Proposition (a) and proposition (d) jointly imply that either the flag is down or the lights are off. Since the premise set is closed under logical consequence, we have to add this disjunctive proposition to the set. The problem is that this disjunctive proposition lumps (b), which is inconsistent with the counterfactual assumption (e) already in our set. The conclusion is that possibility (ii) is ruled out by lumping and the counterfactual in (9) is out.

Kanazawa, Kaufmann & Peters (2005) have raised some objections to the lumping analysis in Kratzer 1989 and I refer the reader to that work for a discussion of their criticisms. Here I would like to point out what I think are some counterintuitive consequences of the lumping semantics. Consider the following example again from Kratzer 1989.

(10) Paula is buying a pound of apples. The Atlantic Ocean is not drying up.
If Paula were not buying a pound of apples, the Atlantic Ocean might be drying up.

The counterfactual in (10) is strange and arguably false. According to Kratzer's story, these are the propositions involved: (a) that Paula is buying a pound of apples;

The set of propositions relevant for the truth-conditions of a counterfactual *if* ϕ , *would* ψ must be (i) consistent, (ii) must include ϕ ; (iii) must be closed under lumping; and (iv) must be closed under logical consequence.

(b) that the Atlantic Ocean is not drying up; (c) that Paula is buying a pound of apples or the Atlantic Ocean is drying up. If we could add the disjunctive proposition in (c) to the premise set, then the set (which contains the counterfactual proposition that Paula is not buying a pound of apples) would entail that the Atlantic Ocean is drying up, and the counterfactual would be predicted to be true. However, Kratzer argues, we cannot add the disjunctive proposition in (c) because (c) lumps (a) in the actual world and (a) is inconsistent with the antecedent. Hence, due to the requirement that the premise set be closed under lumping, we cannot add (c) without creating an inconsistent premise set. Technically, Kratzer's lumping semantics succeeds in ruling out (10), but this technical solution seems dissociated from our intuition that what is wrong with (10) is not just that it is false but that Paula buying or not a pound of apples is irrelevant to whether or not the Atlantic Ocean is drying up.

2 The proposal

The general idea is that what drives the selection of possible worlds is the need to avoid trivial moves in discourse, where a move is to be understood in the sense of Roberts (1996). Here is the outline of the proposal. A counterfactual conditional *if ϕ , would ψ* is understood to be a conditional answer, that is, an answer to a question raised relative to a temporary context in which ϕ is true. Let us call this conditional question the conditional question under discussion (CQUD). In other words, the counterfactual *if ϕ , would ψ* is an answer to the CQUD *if ϕ , $Q?$* , where both $Q?$ and ψ are understood relative to a temporary ϕ -context.

For the question *if ϕ , would $Q?$* to be a non-trivial question, the answer must not be already entailed by the context in which the question is asked. Since the actual question $Q?$ is modally subordinated, the context in which it is asked is the set of ϕ -worlds maximally similar to the actual world. If any of the facts assumed to be true in these ϕ -worlds entail any of the answers to the modally subordinated question $Q?$, then the question is trivial. If the question is trivial and *would ψ* is a relevant answer to the modally subordinated question *would $Q?$* , then the counterfactual *if ϕ , would ψ* will be either vacuously true or vacuously false. Either way, it will be uninformative. If *would ψ* is not a relevant answer, then it will be vacuously false.

In what follows I will say a few words on two notions that will be important for the discussion to follow: the notion of a Question Under Discussion and the notion of a Conditional Question.

2.1 Conditional Questions

I loosely follow the analysis of conditional questions like (11) in Isaacs & Rawlins 2008. For reasons of space, the following discussion will be informal and brief and

will only be concerned with polar questions.

(11) If Alfonso comes to the party, will Joanna leave?

According to [Isaacs & Rawlins \(2008\)](#), the question *will Joanna leave* is modally subordinated to the supposition expressed by the *if*-clause. The framework that these authors adopt is a variant of context change semantics, whereby the meaning of a sentence is its context change potential (cf. [Heim 1992](#)). The interpretation of a conditional question is done in two steps. The first step consists of creating a temporary copy of the context and updating it with the antecedent. The second step consists of interpreting the question relative to this temporary context: this is meant to capture the intuition that the issue of whether Joanna will leave is only raised relative to the temporary context updated with the antecedent proposition that Alfonso will come to the party. Isaacs and Rawlins assume the analysis in [Groenendijk 1999](#): interpreting a question means to update the context inquisitively, that is, only partitioning the context set into two cells. This operation creates an *inquisitive context*, that is, a context with more than one cell.

What is important for our purposes is the following. First, the question part in a conditional question is modally subordinated, that is, it is interpreted not relative to the main context but relative to a temporary context which consists of the main context updated with the proposition expressed by the antecedent *if*-clause. Second, the answer to the conditional question is an answer to the modally subordinated question. Hence, the answer will be interpreted as eliminating one of the two cells in which the temporary context was partitioned by the question. In other words, both the question part in a conditional question and its answer are modally subordinated to the same temporary context. Because we are looking at counterfactual conditionals, the temporary context (that is, the set of relevant ϕ -worlds) cannot be the main context merely updated with ϕ . This temporary context will have to undergo some revisions at least so as to accommodate the counterfactual antecedent (see [Heim 1992](#), [Ippolito 2006](#), among others). The goal of this paper is precisely that of proposing a mechanism for selecting the relevant set of antecedent worlds.

2.2 Questions Under Discussion

Following [Roberts \(1996\)](#), let us assume that a discourse is a structure of questions and answers. The QUD-stack is the set of questions under discussion (QUD) at a given point in the conversation. At each point in discourse, the question at the top of the stack is the (immediate) QUD. Once a question is raised and accepted, then the participants in the conversation are committed to answering it (if it is answerable). A discourse is structured coherently if it obeys the principle of Relevance, which informally requires that a given assertion select from the Q-alternative set of the

QUD, where the Q-alternative set of a question is the set of possible answers denoted by the question, as in Hamblin 1973. Following Roberts (1996), we can define Relevance a bit more formally as shown in (12).

- (12) A move m is relevant to the QUD q iff m either introduces a (partial) answer to q (m is an assertion) or is part of a strategy to answer q (m is a question), where a strategy to answer q consists of answering all those subquestions whose answers constitute partial answers to q .

It follows that in a felicitous discourse, each move will be relevant to the current question under discussion.

Roberts (1996) applies these ideas about information structure to the phenomenon of association with focus, and in particular to the analysis of the focus-sensitive particle *only*. Roberts agrees with Rooth (1985) that association with focus is the result of how prosodic focus affects the restriction of the domain of the relevant operator but argues that this can be explained in terms of information structure and the notions introduced above. In what follows, I will argue that something very similar to what Roberts has proposed for focus operators can be used to account for our intuitions about the truth and falsehood of counterfactual conditionals: in order to successfully restrict the domain of the modal operator we need to identify the QUD.

3 Back to counterfactuals

The discourse in which an utterance of *if ϕ , would ψ* is made is felicitous if it obeys Relevance, as defined above. Now, this is the case if the counterfactual is a relevant conditional answer to the CQUD. This means that the answer (ψ) must eliminate one of the cells in which the temporary context was partitioned by the modally subordinated question. For this to be possible, two requirements must be satisfied. First, it must be the case that both the modally subordinated answer ψ and the modally subordinated question are interpreted relative to the same set of worlds, i.e. a temporary context revised and updated with ϕ . In other words, the CQUD to which *if ϕ , would ψ* is a relevant answer must be of the form *if ϕ , would $Q?$* . Second, it must also be the case that the modally subordinated answer ψ selects from the Q-alternative set of the question [$Q?$].

Before proceeding with the arguments, let me spell out those assumptions about the semantics for counterfactuals that will be relevant in constructing my proposal. For ease of exposition, I will combine elements from Lewis 1973, Kratzer 1991, and von Stechow 2001.

A counterfactual of the form *if ϕ , would ψ* is true in the actual world w_c just in case ψ is true in all ϕ -worlds most similar to the actual world w_c . This is

schematically given in (13).

$$(13) \quad \llbracket \text{if } \phi, \text{ would } \psi \rrbracket^c = 1 \text{ in } w_c \text{ iff } \forall w' \in \text{sim}_{\leq A_{w_c}}(\llbracket \phi \rrbracket^c) : \llbracket \psi \rrbracket^c(w')=1$$

The crucial notion here is that of comparative similarity between worlds, $w' \leq_A w''$, which is defined as follows.

$$(14) \quad \text{For all } w, w' \in W, \text{ for any } A \subseteq \wp(W): \\ w \leq_A w' \text{ iff } \{p : p \in A \text{ and } w' \in p\} \subseteq \{p : p \in A \text{ and } w \in p\}$$

Here, comparative similarity is defined relative to a set of propositions A : for any two worlds w and w' , w is ranked as high as w' just in case the number of propositions in A true in w is at least as high as the number of propositions in A true in w' . Since in counterfactuals the ordering is given by a relation of comparative similarity to the actual world, let A be a set of proposition which fully describe the actual world w_c . The similarity function $\text{sim}_{\leq A_{w_c}}$ applies to the antecedent proposition and returns the set of antecedent worlds which are at least as similar to w_c as any other (accessible) antecedent-world.

$$(15) \quad \text{sim}_{\leq A_{w_c}}(p) = \{w' : p(w')=1 \ \& \ \forall w'' : p(w'')=1 \rightarrow w' \leq_{A_{w_c}} w''\}$$

We saw above that $\leq_{A_{w_c}}$ needs to be constrained. In what follows I will propose an algorithm which systematically constrains the ordering source by constraining A .

As mentioned above, the crucial idea is that an utterance of a counterfactual conditional *if* ϕ , *would* ψ in a context c is always interpreted relative to the CQUD in c at the time the utterance is made. I propose that, when evaluating a counterfactual of the form *if* ϕ , *would* ψ , constraining the similarity ordering $\leq_{A_{w_c}}$ is guided not only by the need to avoid inconsistencies, but also by the need to avoid trivial moves (both question-moves and answer-moves) in the conversational context. More specifically, when selecting the set of ϕ -worlds maximally similar to w_c to be quantified over by the necessity modal, we will keep all propositions true in w_c except (i) those propositions that are inconsistent with the counterfactual antecedent ϕ and (ii) those propositions that entail a member of the Q-alternative set of the question under discussion, that is, those propositions that entail an answer to the question under discussion.

(16) **Constraining similarity (final):**

When evaluating a counterfactual *if ϕ , would ψ* in a context c , where *if ϕ , would Q ?* is the current CQUD, relative to the similarity ordering $\leq_{A_{w_c}}$:

$$\boxed{sim_{\leq_{A_{w_c}}}(\phi) = ntr_Q(con(\phi)(A_{w_c}))}$$

where:

- (i) for every $X \subseteq \wp(W)$ and $\phi \in \wp(W)$: $con(\phi)(X) = \{p \in X : \phi \cap p \neq \emptyset\} \cup \{\phi\}$
- (ii) for every $X \subseteq \wp(W)$ and question Q : $ntr_Q(X) = \{p \in X : \forall q \in Q\text{-alt}, p \not\subseteq q\}$

In (16), the function con captures what I will call the Consistency constraint: $con(\phi)(A_{w_c})$ is going to deliver the union of $\{\phi\}$ with the set of propositions that are true in the actual world and consistent with the antecedent. The ntr function is designed to capture the basic idea behind what I will call the Non-Triviality constraint: ntr will constrain $con(\phi)(A_{w_c})$ by ruling out those propositions entailing a possible answer to the question under discussion.

In what follows I will return to all the examples that we considered above and that raised problems for the theories of counterfactuals we reviewed above, and I will show how the proposal I make in this paper provides a natural way to account for these and other cases that we will introduce later.

4 Jones and the weather

Recall Tichy's example in (3) and Veltman's variant in (4). The observation is that we judge (3) false and (4) true. This means that, in selecting the set of worlds most similar to w_c in which the weather is fine, we do not keep the proposition that Jones is wearing his hat, we do keep the proposition that heads came up and thus require all antecedent worlds to be worlds in which heads came up. In other words, we treat the proposition that Jones is wearing his hat and the proposition that heads came up differently, despite the fact that they are both consistent with the counterfactual antecedent. Why?

Counterfactuals are interpreted as answers to the conditional question under discussion at the time of utterance. In the case of both (3) and (4), the CQUD is a question about what would be the case with respect to a certain issue, if the weather had been fine. What issue? The context has made salient non-accidental generalizations involving the weather and Jones' wearing his hat. It is this connection that is under discussion in the context of utterance. Hence the CQUD here is (17).

- (17) If the weather had been fine, would Jones be wearing his hat?

The Q-alternative set for the modally subordinated question *would Jones be wearing his hat?* is {that Jones is wearing his hat; that Jones is not wearing his hat}.

Let us start with (3). The relevant propositions true in the actual world are given in (18).

- (18) a. that the weather is bad
 b. that Jones is wearing his hat
 c. that whenever the weather is bad, Jones wears his hat.

Now, Consistency requires that we do not keep (18a) because it is inconsistent with the antecedent proposition (that the weather is fine). The proposition in (18b), on the other hand, is ruled out by Non-Triviality as it entails one member of the question's Q-alternative set. Nothing rules out (18c), so we keep it. Since we do not retain (18b), the counterfactual in (3) is false. This is because, as a result of removing (18b), the set of antecedent-worlds maximally similar to the actual world (i.e. $sim_{\leq A_{w_c}}(\lambda w. \text{the weather is fine in } w)$ in the truth-conditions below) is going to include worlds where Jones is wearing his hat and worlds where Jones is not wearing his hat.

- (19) **[[If the weather had been fine, Jones would be wearing his hat]]^c = 1 in w_c iff $\forall w' \in sim_{\leq A_{w_c}}(\lambda w. \text{the weather is fine in } w)$, Jones is wearing his hat in w'**

Let us turn to (4). The relevant propositions this time are shown in (20).

- (20) a. that the weather is bad
 b. that Jones is wearing his hat
 c. that heads came up
 d. that whenever the weather is bad, Jones wears his hat
 e. that whenever the weather is fine and heads comes up, Jones wears his hat

Since the CQUD for (4) is also (17), the Q-alternative set will also be {that Jones is wearing his hat; that Jones is not wearing his hat}. Now, of the propositions listed above, (20a) is ruled out by Consistency, and (20b) is ruled out by Non-Triviality. Crucially, though, the proposition in (20c) is neither ruled out by Consistency nor is it ruled out by Non-Triviality. Therefore, we keep it. As a result, all worlds in which the weather is fine and that are otherwise maximally similar to the actual world (i.e. $sim_{\leq A_{w_c}}(\lambda w. \text{the weather is fine in } w)$ in the truth-conditions above), are going to be worlds where heads came up. Hence, the truth of the counterfactual in (4).

5 Coin-tossing

Suppose that this morning, I bet on tails, I tossed a coin, heads came up, and I lost \$10. Now, suppose counterfactually that I had bet on heads.

(21) If I had bet on heads, I wouldn't have lost \$10.

We judge this counterfactual true. Now, consider Arregui's variant from [Arregui 2009](#). Assume that Peter presses a button in a random coin-tossing device and heads comes up. Now, suppose counterfactually that Susan had pressed the button.

(22) If Susan had pressed the button, the coin would have come up heads.

We judge this counterfactual false. In the remaining part of this section I will show how our proposal about constraining similarity on the basis of the CQUD explains our judgments in these cases.

Let us start with (21). The relevant propositions are given in (23).

- (23)
- a. that I bet on tails
 - b. that I tossed a coin
 - c. that heads came up
 - d. that I lost \$10

The crucial step is to identify the CQUD at the time (21) is uttered. Assuming that (21) is an answer to the conditional question under discussion at the time of utterance means that both the consequent in (21) and the question part of the CQUD are modally subordinated to the same temporary context. Hence, the question is going to be of the form *If I had bet on heads, ... ?*. World-knowledge tells us that there is a non-accidental correlation between betting on something and either winning or losing. Hence, the CQUD will be (24).

(24) If I had bet on heads, would I have lost \$10?

Consistency will rule out (23a). Because the Q-alternative set for the modally subordinated question will be {that I lost \$10, that I didn't lose \$10}, Non-Triviality will then rule out (23d). Both (23b) and (23c) will stay. Hence, all worlds in $sim_{\leq_{Aw_c}}(\lambda w. \text{I bet on heads in } w)$ will be worlds where I tossed a coin and heads came up and in none of those worlds did I lose \$10.

What is different in (22)? What issue does the (counterfactual) supposition that Susan pressed the button raise? Given our world-knowledge (that each coin tossing event has a random outcome) and given the context of utterance, the CQUD is (25).

(25) If Susan had pressed the button, what would have come up?

The Q-alternative set of the modally subordinated question is {that it came up heads; that it came up tails}. The relevant propositions are given in (26).

- (26) a. that Peter pressed the button
b. that heads came up

This time Consistency rules out (26a); Non-Triviality rules out (26b). It follows that $sim_{\leq_{Aw_c}}(\lambda w. \text{Susan pressed the button in } w) \not\subseteq \{w : \text{the coin came up heads in } w'\}$ and the counterfactual is correctly predicted to be false.

6 King Ludwig of Bavaria

Recall Kratzer's King Ludwig of Bavaria example given above in (7a). Whenever the lights are on and the flag is up, the king is in the castle. Now, the lights are on, the flag is down, and the king is away. Now, suppose counterfactually that the flag were up. The relevant examples are repeated below.

- (27) a. If the flag were up, the king would be in the castle.
b. If the flag were up, the lights would be off.

In this context, we accept (27a) but not (27b). Let us begin with (27a). What is the issue that the counterfactual supposition that the flag is up open up? What is the CQUD? The context has established that the position of the flag is one factor in determining the location of the king, and there are two possibilities: either the king is in the castle or the king is away (that is, not in the castle). Hence, the CQUD in the context at the time of utterance is (28).

- (28) If the flag were up, where would the king be?

The Q-alternative set for the modally subordinated question is {that the king is in the castle, that the king is away}. The relevant propositions are given in (29).

- (29) a. that the lights are on
b. that the flag is down
c. that the king is away
d. that whenever the flag is up and the lights are on, the king is in the castle.

Proposition (29b) is ruled out by Consistency and, crucially, proposition (29c) is ruled out by Non-Triviality since it entails a member of the Q-alternative set of the question under discussion. Both (29a) and (29d) are ruled out neither by Consistency nor by Non-Triviality. We keep them both and as a result all worlds in which the flag is up and that are otherwise maximally similar to the actual world are worlds where the lights are on (as shown below) and where the nonaccidental generalization holds.

$$(30) \quad \text{sim}_{\leq A_{w_c}}(\lambda w. \text{the flag is up in } w) \subseteq \{w : \text{the king is in the castle in } w'\}$$

Compare (27a) to (27b). Because the QUD is determined by the context of utterance, and since (27b) is uttered in the same context as (27a), the CQUD is going to be the same as before, i.e. (28). Recall that according to (16), how we constrain similarity is determined by the counterfactual antecedent and the CQUD. This means that, just like in the case of (27a), Consistency will rule out (29b) and Non-Triviality will rule out (29c). As a result, all worlds selected by the similarity function will be worlds where the lights are on, making the counterfactual in (27b) false, as shown below.

$$(31) \quad \text{sim}_{\leq A_{w_c}}(\lambda w. \text{the flag is up in } w) \cap \{w : \text{the lights are off in } w'\} = \emptyset$$

If we manipulate our contextual assumptions, our intuitions about the truth of these conditionals will change too. This is because by changing our contextual assumptions we change what is under discussion in the context and therefore what the CQUD is. Suppose that the king of Bavaria no longer goes to Leoni Castle as he prefers to spend his days at his other residences. However, someone still takes care of the castle and they play with the flag and the lights always making sure, though, that the lights are never on when the flag is up and that the flag is never up when the lights are on (since the king is always away). In this context, I think our judgments would be reversed: we would judge (27a) false and (27b) true. This is because, given the current assumptions, counterfactually assuming that the flag is up does not raise the issue of where the king is (we assume he is always away), but does raise the issue of the status of the lights.

Let us go back to (27b). (27b) is not merely false. There is something wrong with it: it is not a felicitous (relevant) answer to the CQUD in (28) because the consequent is not a possible answer to the modally subordinated question *where would the king be?* As a result, (27b) is *trivially false*. We elaborate more on this point in the next section.

7 Relevancy

The mechanism that I have proposed to constrain similarity in the interpretation of counterfactual conditionals raises a crucial question: how do we know what the CQUD is? To answer this question, consider again the counterfactual in (10) *If Paula were not buying a pound of apples, the Atlantic Ocean might be drying up*: Paula is buying a pound of apples, the Atlantic Ocean is not drying up, and the counterfactual is false. More importantly, it is strange. Why? For (10) to be a felicitous (relevant) answer, the CQUD at the time of utterance would have to be something like (32).

$$(32) \quad \text{If Paula were not buying a pound of apples, would the Atlantic Ocean be drying up?}$$

The problem with (32) is that, given our world-knowledge, this conditional question cannot plausibly be under discussion. Supposing counterfactually that Paula is not buying a pound of apples does not raise the issue of whether the Atlantic Ocean will dry up. Of course, it might if our assumptions about the world were to change, in which case our judgment of (10) might change as well.

Similarly for (33), also from Kratzer 1989. Paula and Otto are the only people in this room. They are both painters. Clara is not in the room and she is a sculptor (not a painter).

(33) If Clara were also in this room, she might be a painter.

Just like (10), (33) is not only false but strange. For (33) to be a relevant answer, the CQUD at the time of utterance would have to be (34).

(34) If Clara were also in this room, what would she be?

Just like in the previous example, what is wrong with this is that, given our world-knowledge, entertaining the counterfactual supposition that Clara is in this room does not raise the issue of what she is professionally. Examples like (10) and (33) show that identifying the correct CQUD is crucial not only in selecting the right set of worlds (by selecting the correct premise set) but also in explaining the independent observation that counterfactual conditionals where the truth of the consequent has no connection with the antecedent are odd: counterfactually supposing the truth of the antecedent fails to raise the issue that the consequent is supposed to be an answer to. In other words, asserting *if ϕ , would ψ* is an infelicitous move since the conditional question it is supposed to answer, i.e. *if ϕ , would ψ ?*, was not under discussion in the context of utterance.

Now, what is under discussion (relative to the counterfactual assumption that we are making) is determined by contextual assumptions together with our world-knowledge. Why do the suppositions that Paula is not buying a pound of apples and that Clara is in this room not raise the issues of whether the Atlantic Ocean is drying up and whether Clara is a painter respectively? The answer is that in our world-knowledge there is no “connection” between these facts.

What kind of connection do we need? There are (at least) two ways to connect two propositions. First, two propositions can be *logically* related, as in (35).

(35) If Paula had eaten candy and cake, she would have eaten cake.

The proposition that Paula ate cake is clearly connected to the proposition that Paula ate candy and cake since the latter entails the former. But notice that, because of the logical connection between these two propositions, entertaining the supposition that Paula ate candy and cake fails to raise the issue whether Paula ate cake. Of course, she did! The counterfactual in (35) is true, but trivially so.

Given (16), cases like (35) turn out to be violations of the Non-Triviality constraint: since $con(\phi)(A_{w_c})$ always include ϕ , every time ϕ itself entails a possible answer to the question under discussion, Non-Triviality will force the removal of the antecedent and as a result the relevant set of ϕ -worlds cannot be constructed.²

The conclusion is that a logical connection between the antecedent and the consequent gives rise to violations of our Non-Triviality constraint. The question then is: how can two propositions be connected while avoiding a trivial counterfactual? Two propositions can be connected by a *nonaccidental generalization* that is part of our world-knowledge or that we assume in the context of utterance. A known nonaccidental generalization or law will provide us with the right kind of information to determine what issues are raised by entertaining a counterfactual supposition. In other words, it is because of these nonaccidental generalizations that we are able to figure out in conversation what the CQUD is when interpreting a counterfactual conditional. This, I claim, is the reason why nonaccidental generalizations are central in the interpretation of counterfactuals, as observed by many (Pollock 1976, Lewis 1979, Kratzer 1989, just to cite a few). But crucially, the effort of figuring out what issue is under discussion is not specific to counterfactual or conditional sentences. Following Roberts (1996), it is what we do every time we make a conversational move.

In the previous sections we saw that identifying the correct CQUD is essential in being able to rule out the right facts and select the correct set of antecedent-worlds for the truth-conditions of the counterfactual conditional. Odd examples like (10) on the one hand, and (35) on the other, show that identifying the correct CQUD is crucial in explaining the observation that counterfactual conditionals where either (i) the truth of the consequent has no connection with the antecedent or (ii) the truth of the consequent is entailed by the antecedent alone, are odd. This is because if there is no connection between the two propositions in the context of utterance (cf. (10)) or if the connection between the two is too “tight” (cf. (35)), then counterfactually supposing the truth of the antecedent fails to raise the issue that the consequent is supposed to be an answer to. In all these cases, asserting *if ϕ , would ψ* is an infelicitous move since the conditional question that the conditional assertion is supposed to answer, i.e. *if ϕ , would ψ ?*, was not under discussion in the context of utterance.

8 Conclusion

The goal of this paper was to provide an answer to the question about how we resolve the context dependence of counterfactuals. The starting point of our proposal was

² Note that the constraint in (16) captures what Condoravdi (2002) calls the “Diversity Condition” without actually stipulating it.

the idea proposed and defended by Roberts (1996) that all conversational moves (questions and assertions) are answers to often implicit questions under discussion. When selecting the set of worlds maximally similar to the actual world in the evaluation of an utterance of a counterfactual conditional *if ϕ , would ψ* , the context-dependence of similarity is reduced to the context-dependence of what we called the conditional question under discussion, that is the conditional question under discussion at the time of utterance. Besides Consistency, which every theory must adopt in some form or other, I proposed that the selection of the relevant antecedent-worlds is constrained by what I called Non-Triviality: any proposition that entails a possible answer to the CQUD cannot be part of the set of propositions A that determines the similarity ordering \leq_A . I have proposed that identifying the CQUD depends on features of the context and world-knowledge.

In the proposal defended in this paper the mechanisms responsible for the selection of the relevant antecedent worlds in giving the truth-conditions of counterfactuals are not specific to counterfactuals but are part of the general strategies that structure the discourse and regulate the flow of information among participants in a conversation.

References

- Arregui, Ana. 2009. On similarity in counterfactuals. *Linguistics and Philosophy* 245–278.
- Condoravdi, Cleo. 2002. Temporal interpretation of modals. In D. Beaver, S. Kaufmann & L. Casillas (eds.), *Stanford papers on semantics*, CSLI Publishers.
- von Stechow, Kai. 2001. Counterfactuals in a dynamic context. In M. Kenstowicz (ed.), *Ken Hale: A life in language*, 123–152. Cambridge, MA: MIT Press.
- Goodman, Nelson. 1947. The problem of counterfactual conditionals. *Journal of Philosophy* 44. 113–128.
- Groenendijk, Jeroen. 1999. The logic of interrogation. In T. Matthews & D. L. Strohmann (eds.), *Proceedings of salt ix*, 109–126. CLC Publications.
- Hamblin, C. L. 1973. Questions in montague english. *Foundations of Language* 10. 41–53.
- Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9. 183–221.
- Ippolito, Michela. 2006. Semantic composition and presupposition projection in subjunctive conditionals. *Linguistics and Philosophy* 29(6). 631–672.
- Ippolito, Michela. 2013. *Subjunctive conditionals: a linguistic analysis*. Linguistic Inquiry Monograph. MIT Press. To appear.
- Isaacs, James & Kyle Rawlins. 2008. Conditional questions. *Journal of Semantics* 25. 269–319.

Counterfactuals

- Kanazawa, Makoto, Stefan Kaufmann & Stanley Peters. 2005. On the lumping semantics of counterfactuals. *Journal of Semantics* 22. 129–151.
- Kratzer, Angelika. 1989. An investigation of the lumps of thought. *Linguistics and Philosophy* 12. 607–653.
- Kratzer, Angelika. 1991. Modality. In A. v. Stechow & D. Wunderlich (eds.), *Semantics. an international handbook of contemporary research*, 639–650. Berlin: De Gruyter.
- Lewis, David. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Lewis, David. 1979. Counterfactual dependence and time's arrow. *Noûs* 13. 455–476.
- Pollock, J. 1976. *Subjunctive reasoning*. Reidel.
- Ramsey, F. P. 1929. General propositions and causality. In R. B. Braithwaite (ed.), *The foundations of mathematics and other logical essays*, 237–255. Routledge & Kegan Paul, London.
- Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. The Ohio State University.
- Rooth, Mats. 1985. *Association with focus*. GLSA Publications.
- Tichy, Pavel. 1976. A counterexample to the stalnaker-lewis analysis of counterfactuals. *Philosophical Studies* 29. 271–273.
- Veltman, Frank. 2005. Making counterfactual assumptions. *Journal of Semantics* 22. 159–180.

Michela Ippolito
Sidney Smith Hall
100 St. George Street
Toronto, ON M5S 3G3 Canada
michela.ippolito@utoronto.ca