Breaking de Morgan's law in counterfactual antecedents[*]

Lucas Champollion   Ivano Ciardelli
New York University   *University of Amsterdam*

Linmin Zhang
New York University

Abstract The main goal of this paper is to investigate the relation between the meaning of a sentence and its truth conditions. We report on a comprehension experiment on counterfactual conditionals, based on a context in which a light is controlled by two switches. Our main finding is that the truth-conditionally equivalent clauses (i) *switch A or switch B is down* and (ii) *switch A and switch B are not both up* make different semantic contributions when embedded in a conditional antecedent. Assuming compositionality, this means that (i) and (ii) differ in meaning, which implies that the meaning of a sentential clause cannot be identified with its truth conditions. We show that our data have a clear explanation in inquisitive semantics: in a conditional antecedent, (i) introduces two distinct assumptions, while (ii) introduces only one. Independently of the complications stemming from disjunctive antecedents, our results also challenge analyses of counterfactuals in terms of minimal change from the actual state of affairs: we show that such analyses cannot account for our findings, regardless of what changes are considered minimal.

1   Introduction

There are two ways to look at the relation between the meaning of a sentence and its truth conditions. The textbook view is that truth conditions completely determine meaning: "To know the meaning of a sentence is to know its truth conditions" (Heim & Kratzer 1998: p. 1). In the standard intensional semantics framework, this

---

view is implemented by representing the meaning of a sentence as a set of possible worlds—the set of those worlds in which the sentence is true.

An alternative view is that the meaning of a sentence carries some extra structure beyond what is needed to capture its truth conditions, and that the notion of sentential meaning is therefore more fine-grained than what is provided by sets of possible worlds. Here, we focus on a particular framework instantiating this view, inquisitive semantics (Ciardelli, Groenendijk & Roelofsen 2013), although other semantic frameworks are similar in this respect (Alonso-Ovalle 2006, 2009; Fine 2014).[1]

In inquisitive semantics, the meaning of a sentence is not represented as a set of worlds, but as a set of such sets, which are referred to as the alternatives for the sentence.[2] A sentence is true at a world w in case w belongs to some alternative for the sentence. Thus, in inquisitive semantics, the meaning of a sentence determines the sentence's truth conditions, but not the other way around: two sentences sharing the same truth conditions may have distinct meanings.

The difference between these two views can be illustrated with a simple example. Imagine two switches, called A and B, for which only two positions, *up* and *down*, are possible. In this scenario, the sentences in (1) have the same truth conditions.

(1)   a.   Switch A or switch B is down.
       b.   Switch A and switch B are not both up.

Whenever (1a) is true, (1b) is true as well, and vice versa. In fact, given our assumption that a switch is down just in case it is not up, the equivalence between (1a) and (1b) can be regarded as an instance of de Morgan's law ($\neg A \vee \neg B \equiv \neg(A \wedge B)$), which is valid in classical logic—the logic arising from truth-conditional semantics. On the textbook view, sentences (1a) and (1b) therefore have the same meaning.

In inquisitive semantics, by contrast, these sentences have distinct meanings. As we will see in Section 3, (1a) is associated with two distinct alternatives, corresponding to the two disjuncts, while (1b) is associated with a unique alternative.

The primary goal of this paper is to compare these two types of approaches, and to determine whether the meaning of a declarative clause is identical to its truth conditions. We did this by conducting a context-based comprehension experiment in which we compared native speakers' truth value judgments for the following counterfactuals, in which (1a) and (1b) are embedded as antecedents:

(2)   a.   If switch A or switch B was down, the light would be off.
       b.   If switch A and switch B were not both up, the light would be off.

---

1 Dynamic theories (Kamp 1981; Heim 1982; Groenendijk & Stokhof 1990) also take this view but are primarily concerned with discourse phenomena and anaphora, which we set aside here.

2 The semantic notion of alternatives at stake here should not be confused with the notion of focus alternatives (Rooth 1985), or with the notion of scalar alternatives (e.g., Horn 1972).

Our experimental results show that in a context where both switches are up, only (2b), but not (2a), is sensitive to what would happen if both switches were down. Given natural assumptions about compositionality, it follows that (1a) and (1b) are not semantically equivalent, even though they have the same truth conditions. We conclude, therefore, that meaning is not completely determined by truth conditions.

Having established this conclusion, we set out to explain how the truth-conditional difference between the two counterfactuals in (2) arises from the non-truth-conditional difference between their antecedents. A natural explanation can be given by combining an inquisitive semantics for the propositional connectives with the proposal by Alonso-Ovalle (2009) for counterfactuals. According to this proposal, an antecedent does not always provide a unique counterfactual assumption. When it is associated with multiple alternatives, as in the case of (1a), each of these alternatives introduces a separate assumption into the compositional process. Since (1a) and (1b) are associated with different alternatives, the consequents of (2a) and (2b) are assessed in different counterfactual scenarios, resulting in different truth conditions.

Furthermore, independently of the complications stemming from disjunctive antecedents, our finding that the interpretation of (2b) depends on what happens when both switches are toggled raises a novel challenge for theories of counterfactuals that are based on the notion of minimal change from actuality (Stalnaker 1968; Lewis 1973). As we will see, regardless of what counts as a minimal change, our intuitions about certain counterfactuals including (2b) are not accounted for.

The paper is structured as follows. Section 2 describes our experiment. Section 3 introduces inquisitive semantics and shows how, in this framework, two clauses sharing the same truth conditions can differ in meaning. Then, Section 4 shows how a non-truth-conditional difference between antecedent clauses can give rise to a truth-conditional difference at the level of the corresponding counterfactuals. Section 5 explains why our data challenge accounts of counterfactuals that are based on the notion of minimal departure from actuality. Section 6 wraps up and concludes.

2 Experiment

2.1 Hypotheses and predictions

Throughout this paper, we assume the principle of semantic compositionality for natural language; that is, the meaning of a complex expression is completely determined by the meaning of its constituents and by the way they are combined. This implies that the meaning of a complex expression does not change when one of its constituents φ is replaced by another expression φ' having the same meaning.

The central question we seek to answer is whether the truth conditions of a natural language sentential clause completely determine its meaning. Our experiment took

advantage of the truth-conditional equivalence between (1a) and (1b).

Under the hypothesis that truth conditions completely determine meaning, (1a) and (1b) are predicted to have the same meaning no matter what environment they are embedded in. Consequently, given compositionality, two natural language constructions that embed (1a) and (1b) but are otherwise identical should have the same meaning as well, and should be judged true in the same set of situations.

By contrast, under the hypothesis that sentence meanings are not fully determined by their truth conditions, (1a) and (1b) may well differ in their meaning. If so, when embedded in a larger construction, they could make a different semantic contribution, resulting in different truth conditions for the whole sentence, which might be reflected in native speakers' truth value judgments.

To adjudicate between these two hypotheses, we embedded (1a) and (1b) in the antecedent of a counterfactual conditional, as shown in (2), and we conducted a comprehension experiment.

2.2 Experiment design and methods

Our experiment included three parts: (i) two pretests (Section 2.3), (ii) the main experiment (Section 2.4), and (iii) three post-hoc tests (Section 2.5). Pretest I confirmed the truth-conditional equivalence between (1a) and (1b) for native speakers of English, and Pretest II confirmed that the critical sentences used in our main experiment – (2a) and (2b) – are natural to native speakers to the same degree. In the main experiment, we elicited native speakers' truth value judgments for a set of counterfactuals including (1a) and (1b). We used the three post-hoc tests to rule out some alternative accounts for the finding of our main experiment.

We implemented our experiments and tests as web surveys using TurkTools (Erlewine & Kotek 2016), which relies on the platform Amazon Mechanical Turk (http://www.mturk.com). Participants were required to be located in the U.S. and have a Mechanical Turk approval rate (an indication of reliability) of at least 95%.

In all tests, each participant was asked to judge two sentences: one target and one filler sentence. For half of the participants, the target preceded the filler, while for the other half, the order of presentation was reversed. Our fillers were all uncontroversial in terms of naturalness or truth value, and thus the response to them was an indication showing whether participants paid enough attention to stimuli.

In the main experiment, Pretest I, and the three post-hoc tests, participants were shown a pictorial context[3] along with a short descriptive text and were asked to judge whether what the sentences say about the picture is 'true', 'false' or 'indeterminate'.

In Pretest II, there was no pictorial context or descriptive text. Participants were

---

3 Our figures are adapted from multiway switches © Cburnett (https://en.wikipedia.org/wiki/Multiway_switching#/media/File:3-way_switches_position_2.svg) CC BY-SA 3.0.
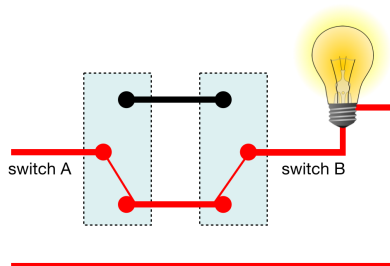
Figure 1   Pretest I. Switch A and B are both down, and the light is on.

asked to judge whether the sentences sound natural on a 7-point scale, where 1 stands for "totally unnatural" and 7 for "perfectly natural".

Before the presentation of our stimuli, we gave examples illustrating the truth value or naturalness judgment task. At the end of the survey, we asked participants whether they were native speakers of English, whether they spoke British or American English or another dialect, and whether they had any comments for us (few did). We stated that their answers to these questions would not affect the payment.

For the truth value judgment task, we paid each participant $0.10. For the naturalness judgment task, we paid each participant $0.02.

We used participants' responses to demographic questions and filler sentences to filter data: responses from those who did not self-identify as native speakers of American English or who failed to judge the filler sentence correctly were ruled out from further analyses. If someone took part in our study more than once, only their first response was included in data analysis. In all our tests, incorrect responses to filler items accounted for the vast majority of rejected data.

2.3 Two pretests

2.3.1 Pretest I

Materials   The goal of Pretest I was to confirm that (1a) and (1b) have identical truth conditions. Since these sentences are both undoubtedly true when exactly one of the two switches is down, and false when the two switches are up, we only elicited truth value judgments of these sentences in a scenario where both switches are down.

To this end, we used the pictorial context in Figure 1 and asked participants to provide truth value judgments for (1a)/(1b). We included the sentence *Switch A is up* as a filler item, and discarded data from participants who failed to judge it false.

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
(1a)	145	118	81.38%	23	15.86%	4	2.76%
(1b)	130	118	90.77%	11	8.46%	1	0.77%

Table 1   Results of Pretest I

Sentence	Label	Number	Mean rating	Standard deviation
(2a)	$\overline{A} \vee \overline{B} >$ OFF	73	5.07	1.63
(2b)	$\neg(A \wedge B) >$ OFF	55	5.16	1.76

Table 2   Results of Pretest II

Results   We collected data from 330 non-repetitive participants and rejected 16.67% of the responses. As shown in Table 1, each sentence was judged true by over 80% of participants. As expected, the results for both sentences did not differ significantly ($\chi^2(2, N = 275) = 5.23$, $p = 0.07$). In the absence of evidence to the contrary, we conclude that (1a) and (1b) are indeed truth-conditionally equivalent.

2.3.2   Pretest II

Materials   Pretest II aimed to verify that the counterfactual sentences (2a) and (2b) sound equally natural to native speakers. We used the sentence *If I were in the hallway, I would turn the light off* as the filler item, and we excluded data from participants who judged the filler lower than 5 (on a 7-point scale).

Results   As shown in Table 2, both sentences were judged acceptable at comparable levels: the *t*-test comparing the scores of these two sentences showed no significant difference ($p = 0.37$). We conclude that any potential differences between the truth value judgments of these two sentences are unlikely to be attributable to one of the sentences being less natural than the other.

2.4   Main experiment

In our main experiment, we presented a context in which a light is controlled by two switches, and asked participants to give truth value judgments for one of five counterfactual sentences including (2a) and (2b).
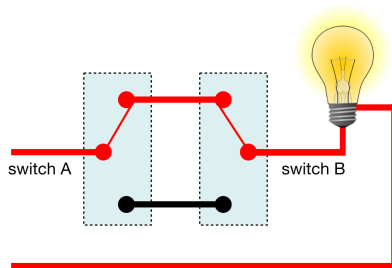
Figure 2   Main experiment. Switch A and B are both up, and the light is on.

Materials   Our context consisted of the picture shown in Figure 2 and of the following descriptive text:[4]

(3)   Imagine a long hallway with a light in the middle and with two switches, one at each end. One switch is called switch A and the other one is called switch B. As this wiring diagram shows, the light is on whenever both switches are in the same position (both up or both down); otherwise, the light is off. Right now, switch A and switch B are both up, and the light is on. But things could be different . . .

Our five target sentences are shown in (4) along with their condition labels, by which we refer to them in the following.[5]

(4)   a.   Condition label: $\overline{A} >$ OFF
               If switch A was down, the light would be off.
         b.   Condition label: $\overline{B} >$ OFF
               If switch B was down, the light would be off.
         c.   Condition label: $\overline{A} \vee \overline{B} >$ OFF   = (2a)
               If switch A or switch B was down, the light would be off.

4 The two-switches scenario was originally introduced by Lifschitz (1990) in the context of causal reasoning. Within the literature on counterfactuals, it was first discussed in Schulz 2007, as a counterexample to the theory in Veltman 2005. That discussion is unrelated to our present concerns. The specific text in (3) is our own, and to the best of our knowledge, our paper is the first to discuss the two-switches scenario in connection with negated conjunctive antecedents.

5 These labels were not shown to participants, and are based on the following conventions: we write *A* as an abbreviation for *switch A is up*, and $\overline{A}$ for *switch A is down*, and similarly for switch B. We use the standard logical notations $\neg, \wedge, \vee$ for negation, conjunction, and disjunction, and $>$ for the counterfactual conditional construction. In Section 3, we will assume that these representations correspond to the logical forms of these sentences, at a suitable level of abstraction.

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\overline{A} >$ OFF	256	169	66.02%	6	2.34%	81	31.64%
$\overline{B} >$ OFF	235	153	65.11%	7	2.98%	75	31.91%
$\overline{A} \vee \overline{B} >$ OFF	362	251	69.33%	14	3.87%	97	26.80%
$\neg(A \wedge B) >$ OFF	372	82	22.04%	136	36.56%	154	41.40%
$\neg(A \wedge B) >$ ON	200	43	21.50%	63	31.50%	94	47.00%

Table 3   Results of the main experiment

  d.   Condition label: $\neg(A \wedge B) >$ OFF   = (2b)
        If switch A and switch B were not both up, the light would be off.
  e.   Condition label: $\neg(A \wedge B) >$ ON
        If switch A and switch B were not both up, the light would be on.

Our filler sentence is shown in (5). We ruled out data from those participants who failed to judged it false under the context shown in Figure 2.

(5)   If switch A and switch B were both down, the light would be off.

Results   We collected data from 2299 non-repetitive participants and rejected 38.02% of the responses. The remaining 1425 responses are summarized in Table 3.

Differences across all five sentences were highly significant ($\chi^2(8, N = 1425) = 383.36, p < 0.0001$). Our results fall naturally into two blocks, as indicated by the dashed line in Table 3. The first block consists of $\overline{A} >$ OFF, $\overline{B} >$ OFF, and $\overline{A} \vee \overline{B} >$ OFF, which were all judged true by a wide majority. In the second block, $\neg(A \wedge B) >$ OFF and $\neg(A \wedge B) >$ ON were generally judged false or indeterminate. The frequency difference between $\overline{A} \vee \overline{B} >$ OFF and $\neg(A \wedge B) >$ OFF is significant: $\chi^2(2, N = 734) = 197.84, p < 0.0001$. Differences within each block were not significant (first block: $\chi^2(4, N = 853) = 3.33, p = 0.5042$; second block: $\chi^2(2, N = 572) = 1.92, p = 0.3829$).

Crucially, our results show that $\overline{A} \vee \overline{B} >$ OFF and $\neg(A \wedge B) >$ OFF are judged differently, indicating that these two counterfactuals have different truth conditions.

2.5   Three post-hoc tests

Given our account, the finding of our main experiment suggests that (1a) and (1b) differ in meaning. To further solidify this conclusion, we ran three post-hoc tests that rule out some potential alternative accounts for the drop in 'true' judgments from $\overline{A} \vee \overline{B} >$ OFF and $\neg(A \wedge B) >$ OFF.
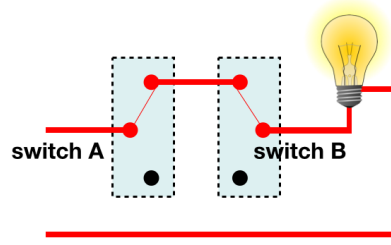
Figure 3   Post-hoc test I. There is no wire between the two "down" positions.

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\overline{A} >$ OFF	52	41	78.85%	5	9.61%	6	11.54%
$\overline{B} >$ OFF	68	60	88.24%	5	7.35%	3	4.41%
$\overline{A} \vee \overline{B} >$ OFF	110	104	94.55%	1	0.91%	5	4.54%
$\neg(A \wedge B) >$ OFF	116	99	85.34%	9	7.76%	8	6.90%
$\neg(A \wedge B) >$ ON	103	19	18.45%	79	76.70%	5	4.85%

Table 4   Results of Post-hoc test I

2.5.1   Post-hoc test I: The light is on only if both switches are up

Materials   Post-hoc test I aimed to test whether the difference in reported truth value judgments between $\overline{A} \vee \overline{B} >$ OFF and $\neg(A \wedge B) >$ OFF might be due to context-independent factors such as differences in complexity or processing load. To this end, we replaced the pictorial context by the one shown in Figure 3 and we replaced the third sentence in our descriptive text by the sentence in (6):

(6)   As the following wiring diagram shows, the light is on whenever both switches are up; otherwise, the light is off.

If in our main experiment, the difference in truth value judgments between $\overline{A} \vee \overline{B} >$ OFF and $\neg(A \wedge B) >$ OFF is mainly due to context-independent factors, we should observe exactly the same difference in this post-hoc test. Alternatively, if it indeed tracks actual differences in truth conditions, then in this new context, we expect that the result pattern for the five counterfactual sentences might change.

We used the filler *If switch A and switch B were both down, the light would be on*, and we rejected data from participants who failed to judge the filler false.

Results   We collected data from 553 non-repetitive participants and rejected 18.81% of the responses. The remaining 449 responses are summarized in Table 4.

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\neg A > \text{OFF}$	36	27	75.00%	1	2.78%	8	22.22%
$\neg B > \text{OFF}$	43	28	65.12%	7	16.28%	8	18.60%
$\neg A \vee \neg B > \text{OFF}$	80	48	60.00%	16	20.00%	16	20.00%
$\neg(A \wedge B) > \text{OFF}$	372	82	22.04%	136	36.56%	154	41.40%
$\neg(A \wedge B) > \text{ON}$	200	43	21.50%	63	31.50%	94	47.00%

Table 5   Results of Post-hoc test II. The last two lines are repeated from Table 3.

This time, there were no significant differences among the truth value judgments of the first four sentences ($\chi^2(6, N = 346) = 11.26, p = 0.08$). Moreover, for both $\overline{A} \vee \overline{B} > \text{OFF}$ and $\neg(A \wedge B) > \text{OFF}$, most ($> 85\%$) participants judged them to be true in this context. These results suggest that the differences in truth value judgments between $\overline{A} \vee \overline{B} > \text{OFF}$ and $\neg(A \wedge B) > \text{OFF}$ that we observed in our main experiment is unlikely to be due to context-independent factors.

2.5.2   Post-hoc test II: Replacing *down* by *not up*

Materials   Post-hoc test II was designed to test whether the presence or absence of explicit negation affects the result pattern of the main experiment. To this end, we replaced the word *down* by *not up* in the target sentences that used it. We did not replace *down* by *not up* in the filler sentence.

Results   For $\neg A > \text{OFF}$, $\neg B > \text{OFF}$, and $\neg A \vee \neg B > \text{OFF}$, we collected data from 561 non-repetitive participants and rejected 71.66% of the responses. The remaining 159 responses are summarized in Table 5 along with the results of $\neg(A \wedge B) > \text{OFF}$ and $\neg(A \wedge B) > \text{ON}$ from the main experiment.

Table 5 shows that substituting *not up* for *down* did not change the pattern in the observed results: differences across all five sentences were highly significant ($\chi^2(8, N = 743) = 129.26, p < 0.0001$). The results shown in Table 5 also fall naturally into two blocks, as indicated by the dashed line. Sentences of the first block were all judged true by a majority, and differences within the first block were not significant ($\chi^2(4, N = 159) = 5.93, p = 0.2044$). The difference between $\overline{A} \vee \overline{B} > \text{OFF}$ in this test and $\neg(A \wedge B) > \text{OFF}$ in the main experiment is still significant: $\chi^2(2, N = 452) = 46.37, p < 0.0001$. Therefore, we can exclude the presence or absence of the word *not* as a potential confounding factor. This also confirms our background assumption that few participants, if any, would consider the possibility that a switch might be in an intermediate position (that is, neither up nor down).

Sentence	Number	True	(%)	False	(%)	Indet.	(%)
$\overline{A}>$ OFF	57	46	80.70%	0	0%	11	19.30%
$\overline{B}>$ OFF	42	35	83.33%	2	4.76%	5	11.90%
$\overline{A}\vee\overline{B}>$ OFF	83	61	73.49%	13	15.66%	9	10.84%

Table 6   Results of Post-hoc test III

2.5.3   Post-hoc test III: Replacing *was* by *were*

Materials   Post-hoc test III was designed to rule out the possibility that the choice of auxiliary affected the truth value judgments we found in our main experiment. To this end, we replaced the word *was* by *were* in the target sentences that used it ($\overline{A}>$ OFF, $\overline{B}>$ OFF, and $\overline{A}\vee\overline{B}>$ OFF).

Results   We collected data from 556 non-repetitive participants and rejected 67.27% of the responses. The remaining 182 responses are summarized in Table 6.

Overall, the results of Post-hoc test III yielded the same pattern as in the main experiment. Each of the sentences in this test was judged true by most ($>70\%$) of the participants. Moreover, the difference between $\overline{A}\vee\overline{B}>$ OFF in this test and $\neg(A\wedge B)>$ OFF in the main experiment is still significant: $\chi^2(2,N=455)=83.89, p<0.0001$. Therefore, we can exclude the choice of auxiliary as a potential factor affecting our main finding.[6]

2.6   Discussion and conclusions

As shown in the results of our main experiment (Table 3), $\overline{A}>$ OFF and $\overline{B}>$ OFF were generally judged 'true'. Given the way the switches are wired, this suggests that most participants interpreted $\overline{A}>$ OFF and $\overline{B}>$ OFF by considering what would be the case if just the switch in question was toggled, leaving the other one in place. Similarly, it seems that most participants interpreted $\overline{A}\vee\overline{B}>$ OFF by considering one switch at a time, while ignoring the option that both switches might be toggled simultaneously.[7]

As for $\neg(A\wedge B)>$ OFF and $\neg(A\wedge B)>$ ON, most participants judged them 'indeterminate' or 'false'. This suggests that the predominant strategy for these sentences

---

6 This time, the comparison among the three sentences $\overline{A}>$ OFF, $\overline{B}>$ OFF and $\overline{A}\vee\overline{B}>$ OFF showed a significant difference: $\chi^2(4,N=182)=13.18, p=0.01$. We have no explanation for this fact.

7 The filler sentence queried that option; by discarding data from participants who judged it incorrectly, we guarded against the possibility that participants were unaware of the fact that the light remains on when both switches are toggled.

is to consider all three possibilities: only switch A is toggled; only switch B is toggled; both switches are toggled. These possibilities do not all agree on the state of the light, leading to the lack of 'true' judgments.

Crucially, $\overline{A} \vee \overline{B} >$ OFF and $\neg(A \wedge B) >$ OFF are judged in a different way. Having excluded various confounds, we take this difference in judgments to reflect an actual difference in truth values: in the given context, $\overline{A} \vee \overline{B} >$ OFF is true, while $\neg(A \wedge B) >$ OFF is not. Thus, these two counterfactuals must have different meanings.

By compositionality, their antecedents, corresponding to (1a) and (1b), must then have different meanings as well. However, these antecedents have the same truth conditions, as confirmed by Pretest I. Hence, it is possible for two sentential clauses to have the same truth conditions and different meanings—which shows that meaning is not completely determined by truth conditions.

In fact, our results are problematic not only for truth-conditional semantics, but for any semantic theory that validates de Morgan's law $\neg(A \wedge B) \equiv \neg A \vee \neg B$. One example is the truth-maker semantics of Fine 2012: while taking sentence meanings to be more structured than truth conditions, this theory validates de Morgan's law, and therefore it leads to the problematic prediction that (1a) and (1b) are equivalent.

2.6.1   Ruling out alternative accounts for our findings

Some previous studies (e.g. Fox 2007; Spector 2007) have suggested that a silent exhaustivity operator *Exh* might strengthen the meaning of natural language *or* and cause it to be interpreted as an exclusive rather than inclusive disjunction (see also Nute 1975). Thus, the use of *Exh* can break de Morgan's law. However, since Pretest I has shown that (1a) is generally judged true even when both switches are down, exhaustive strengthening generally does not take place in (1a). We know of no evidence that strengthening happens in counterfactual antecedents more often than in main clauses.[8] Therefore, it seems unlikely that the observed difference between $\overline{A} \vee \overline{B} >$ OFF and $\neg(A \wedge B) >$ OFF can be attributed to the fact that a silent exhaustivity operator strengthens the antecedent of $\overline{A} \vee \overline{B} >$ OFF.

Since Pretest II has shown that $\neg(A \wedge B) >$ OFF and $\overline{A} \vee \overline{B} >$ OFF are judged equally natural, the drop in 'true' judgments between these two sentences cannot be due to differences in sentence naturalness. Relatedly, Post-hoc tests II and III showed that the drop in 'true' judgments between these two sentences cannot be attributed to the use of *were* or explicit negation in $\neg(A \wedge B) >$ OFF, either.

Finally, it is conceivable that in $\neg(A \wedge B) >$ OFF, the string *not both up* might have been interpreted as *both not up*, either through misreading or as a result of interpreting *up* as focused (Rooth 1996). However, we separately tested the sentence

---

8 In fact, many accounts of exhaustive strengthening assume that if it can occur at all, it can only occur in main clauses; for theoretical issues, see Schlenker to appear.

Switch A and switch B are not both up in a pictorial context that shows switch A up and switch B down: 76.9% of 290 participants judged it true, showing that interpreting *not both up* as *both not up* is at best unlikely. Moreover, if participants really interpreted *not both up* as *both down*, we would expect a spike in 'true' judgments for (4e); however, only 21.5% of participants judged this sentence true.
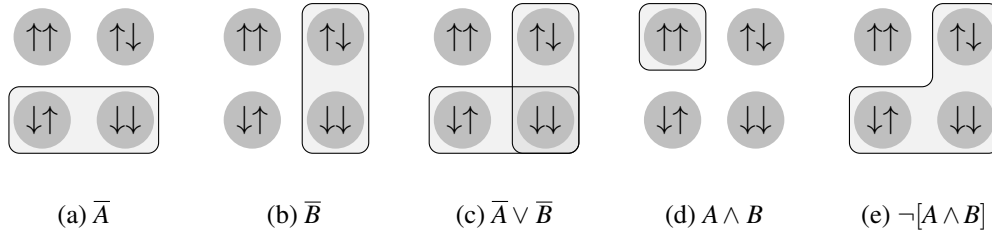
3 Breaking de Morgan's law

The difference between $\overline{A} \vee \overline{B} > \text{OFF}$ and $\neg(A \wedge B) > \text{OFF}$ finds a natural explanation once we move from a purely truth-conditional notion of meaning to a more fine-grained one, such as that provided by inquisitive semantics (Ciardelli et al. 2013). In this framework, the meaning of a sentence φ is given not in terms of truth conditions with respect to possible worlds, but in terms of support conditions with respect to information states, where an information state is modeled as a subset of the set W of possible worlds. The maximal information states supporting a sentence φ are called the alternatives for φ, and the set of alternatives is denoted $\text{Alt}(\varphi)$. A sentence is called inquisitive if it has two or more alternatives, and non-inquisitive if it has only one. The set of worlds where φ is true, denoted $|\varphi|$, is defined as the union of the alternatives for φ. Thus, the inquisitive meaning of a sentence still determines its truth conditions, but the converse is no longer the case: two sentences may very well have the same truth conditions while being associated with different sets of alternatives. This is the case for our counterfactual antecedents (1a) and (1b). To see why, we need to consider how basic clauses are interpreted in inquisitive semantics, and how disjunction, conjunction, and negation operate in this framework.

First, consider the statement *switch A is down*, which we abbreviate as $\overline{A}$. As shown in (7a), this is supported by an information state s in case it follows from the information available in s that switch A is down, that is, in case A is down at each world in s. This in turn means that this statement has a unique alternative, as shown in (7b). The same goes for the other basic clauses *switch B is down*, *switch A is up*, and *switch B is up*, abbreviated here as $\overline{B}$, A, and B. This is illustrated in Figure 4.

(7)   a. $s \models \overline{A}$ iff $s \subseteq \{w \in W \,|\, \text{switch A is down in } w\}$
      b. $\text{Alt}(\overline{A}) = \{\{w \in W \,|\, \text{switch A is down in } w\}\} = \{|\overline{A}|\}$

Inquisitive semantics comes with a natural treatment of propositional connectives, obtained by associating these connectives with the natural algebraic operations on the space of inquisitive meanings (see Roelofsen 2013). In particular, disjunction, conjunction, and negation are interpreted by means of the following support clauses:

(8)   a. $s \models \varphi \wedge \psi$ iff $s \models \varphi$ and $s \models \psi$
      b. $s \models \varphi \vee \psi$ iff $s \models \varphi$ or $s \models \psi$

(a) $\overline{A}$ (b) $\overline{B}$ (c) $\overline{A} \vee \overline{B}$ (d) $A \wedge B$ (e) $\neg[A \wedge B]$

Figure 4   Inquisitive meanings of some simple sentences. ↑↑ represents a world where both switches are up, ↑↓ a world where A is up but B is down, etc. To avoid clutter, only alternatives are depicted.

c. $s \models \neg\varphi$ iff $s \cap t = \varnothing$ for all $t \models \varphi$

We can now verify that in inquisitive semantics, just as in truth-conditional semantics, the sentence *switch A is down* is equivalent with *switch A is not up*, that is, $\overline{A} \equiv \neg A$.[9]

(9)   a. $s \models \neg A$ iff $s \cap t = \varnothing$ for all $t \subseteq \{w \in W \,|\, \text{switch A is up in } w\}$
            iff $s \cap \{w \in W \,|\, \text{switch A is up in } w\} = \varnothing$
            iff $s \subseteq \{w \in W \,|\, \text{switch A is down in } w\}$ iff $s \models \overline{A}$

For our first complex antecedent, (1a), analyzed as $\overline{A} \vee \overline{B}$, inquisitive semantics yields two distinct alternatives: the set $|\overline{A}|$ consisting of those worlds where A is down, and the set $|\overline{B}|$ consisting of those worlds where B is down. These alternatives are depicted in Figure 4(c).

(10)   a. $s \models \overline{A} \vee \overline{B}$ iff $s \subseteq |\overline{A}|$ or $s \subseteq |\overline{B}|$
       b. $\mathsf{Alt}(\overline{A} \vee \overline{B}) = \{|\overline{A}|, |\overline{B}|\}$

Now consider the conjunction *switch A and switch B are both up*, analyzed as $A \wedge B$. Inquisitive semantics yields a unique alternative, consisting of those worlds where both switches are up. This is shown in Figure 4(d).

(11)   a. $s \models A \wedge B$ iff $s \subseteq |A|$ and $s \subseteq |B|$
                 iff $s \subseteq |A \wedge B|$
       b. $\mathsf{Alt}(A \wedge B) = \{|A \wedge B|\}$

Finally, for the clause (1b), analyzed as $\neg(A \wedge B)$ inquisitive semantics yields a unique alternative, consisting of all worlds where the switches are not both up. This

---

9 Recall that we are assuming that *up* and *down* are the only possible positions for our switches. The results we obtained in Post-hoc test II justify this assumption.

is depicted in Figure 4(e).

(12)   a. $s \models \neg(A \wedge B)$ iff $s \cap t = \emptyset$ for all $t \subseteq |A \wedge B|$
                          iff $s \subseteq (W - |A \wedge B|)$
        b. $\mathsf{Alt}(\neg(A \wedge B)) = \{W - |A \wedge B|\}$

Since $|\overline{A}| \cup |\overline{B}| = W - |A \wedge B|$, inquisitive semantics predicts that (1a) and (1b) are true at the same worlds, namely, at those worlds in which one or both switches are down. This is in line with classical logic, and also with the result of Pretest I, as reported in Section 2.3.1. However, these two clauses are assigned different meanings: (1a) has two distinct alternatives, whereas (1b) has only one.

4   Explaining the contrast: Two assumptions for one antecedent

What still remains to be explained is how the non-truth-conditional difference between $\overline{A} \vee \overline{B}$ and $\neg(A \wedge B)$ ends up affecting the truth conditions of the counterfactuals $\overline{A} \vee \overline{B} > \text{OFF}$ and $\neg(A \wedge B) > \text{OFF}$ in which they are embedded. For this, we adopt an idea due to Alonso-Ovalle (2006, 2009) (see also van Rooij 2006; Fine 2012). We assume that a counterfactual antecedent need not always specify a single counterfactual assumption; rather, when an antecedent provides multiple semantic alternatives, as in the case of (1a), each of them constitutes a distinct counterfactual assumption. In order for the counterfactual to be true, the consequent must follow on each of these assumptions. Thus, $\overline{A} \vee \overline{B} > \text{OFF}$ is interpreted in effect as the conjunction of $\overline{A} > \text{OFF}$ and $\overline{B} > \text{OFF}$, and differently from $\neg(A \wedge B) > \text{OFF}$. This explains the strong similarity between the response patterns of $\overline{A} \vee \overline{B} > \text{OFF}$, $\overline{A} > \text{OFF}$ and $\overline{B} > \text{OFF}$.

To implement this idea in our setting, we will apply the general recipe for lifting accounts of counterfactuals into inquisitive semantics described in Ciardelli 2016.[10] The starting point is an arbitrary truth-conditional account of counterfactuals, given in the form of a binary operation $\Rightarrow$ which maps any two propositions p and q to a corresponding conditional proposition $p \Rightarrow q$. Most existing accounts of counterfactuals, including minimal change accounts (Stalnaker 1968; Lewis 1973), premise semantics (Kratzer 1981), and causal accounts (Pearl 2000), can be seen as providing such a map.[11] The lifting recipe interprets a counterfactual statement,

---

10 The theory of Alonso-Ovalle 2009, which inspires the inquisitive lifting recipe we adopt, would account for the semantics of $\overline{A} \vee \overline{B} > \text{OFF}$ in a similar way. However, this theory builds on minimal change semantics, which Section 5 argues is problematic. By contrast, inquisitive lifting is modular, and can be combined with various accounts of conditionals. In addition, whereas conjunction is problematic in alternative semantics (Ciardelli & Roelofsen 2015), inquisitive semantics comes with a principled theory of propositional connectives (Roelofsen 2013), which we put to use in this section.

11 In each of these accounts, the definition of the conditional proposition $p \Rightarrow q$ makes use of some additional piece of structure: a similarity metrics on worlds in the case of minimal change accounts;

denoted by $\varphi > \psi$, by means of the following support clause.

Definition 1 (Inquisitive lifting of an account of counterfactuals)**.**
$s \models \varphi > \psi$ *iff* $\forall p \in Alt(\varphi) \; \exists q \in Alt(\psi)$ *such that* $s \subseteq (p \Rightarrow q)$

According to this clause, when φ and ψ are non-inquisitive, that is, $Alt(\varphi) = \{|\varphi|\}$ and $Alt(\psi) = \{|\psi|\}$, the conditional $\varphi > \psi$ has a unique alternative as well, which coincides with the counterfactual proposition delivered by the given base account: $Alt(\varphi > \psi) = \{|\varphi| \Rightarrow |\psi|\}$.

Except for $\overline{A} \vee \overline{B} > \text{OFF}$, all of the counterfactuals in our experiment have non-inquisitive antecedents and consequents, so they will be interpreted just as they are interpreted by any basic account we may choose. As for $\overline{A} \vee \overline{B} > \text{OFF}$, the clause interprets it as follows (where OFF stands for "the light is off"):

$$
\begin{aligned}
s \models \overline{A} \vee \overline{B} > \text{OFF} \quad &\text{iff} \quad \forall p \in \{|\overline{A}|, |\overline{B}|\} \; \exists q \in \{|\text{OFF}|\} \text{ such that } s \subseteq (p \Rightarrow q) \\
&\text{iff} \quad s \subseteq |\overline{A}| \Rightarrow |\text{OFF}| \text{ and } s \subseteq |\overline{B}| \Rightarrow |\text{OFF}| \\
&\text{iff} \quad s \subseteq (|\overline{A}| \Rightarrow |\text{OFF}|) \cap (|\overline{B}| \Rightarrow |\text{OFF}|)
\end{aligned}
$$

As in the previous cases, the counterfactual as a whole has a unique alternative, namely, the proposition $(|\overline{A}| \Rightarrow |\text{OFF}|) \cap (|\overline{B}| \Rightarrow |\text{OFF}|)$. However, this alternative is not the same proposition $|\overline{A} \vee \overline{B}| \Rightarrow |\text{OFF}|$ that would be delivered by applying the basic truth-conditional account to the sentence. Rather, the basic account is applied twice, once for each disjunct of the antecedent, and the resulting propositions are then intersected. Thus, disjunctive antecedents are interpreted as providing multiple counterfactual assumptions, and $\overline{A} \vee \overline{B} > \text{OFF}$ is predicted to be true just in case both $\overline{A} > \text{OFF}$ and $\overline{B} > \text{OFF}$ are true.

This means that, at least insofar as truth is concerned, our data will be fully explained if we can find a truth-conditional account of counterfactuals according to which $\overline{A} > \text{OFF}$ and $\overline{B} > \text{OFF}$ are true, but $\neg(A \wedge B) > \text{OFF}$ and $\neg(A \wedge B) > \text{ON}$ are not. The inquisitive lifting of this account will still make the same predictions about these cases; moreover, it will predict $\overline{A} \vee \overline{B} > \text{OFF}$ to be true—something that no purely truth-conditional account could do without also making $\neg(A \wedge B) > \text{OFF}$ true.

In sum, inquisitive semantics combined with Alonso-Ovalle's idea allows us to explain the observed difference between $\overline{A} \vee \overline{B} > \text{OFF}$ and $\neg(A \wedge B) > \text{OFF}$. $\overline{A} \vee \overline{B} > \text{OFF}$ requires us to consider two assumptions: the assumption that switch A was down, and the assumption that switch B was down; as witnessed by the truth of the counterfactuals $\overline{A} > \text{OFF}$ and $\overline{B} > \text{OFF}$, in our scenario each of these assumptions leads to the conclusion that the light is off; hence, $\overline{A} \vee \overline{B} > \text{OFF}$ is true. By contrast,

---

premise sets in the case of premise semantics; and a causal network in the case of causal accounts. However, our lifting recipe only needs access to the resulting operation on propositions—not to this additional structure.

$\neg(A \wedge B) > $ OFF asks us to make just one assumption, namely, that the position of the switches was different than it is; as witnessed by the fact that $\neg(A \wedge B) > $ OFF is not judged true, this assumption does not lead to the conclusion that the light is off.

5   The challenge for minimal change theories of counterfactuals

Besides providing evidence against the truth-conditional view on meaning, our data also challenge an idea at the heart of many accounts of counterfactuals, namely, the idea that making a counterfactual assumption requires minimizing the departure from the actual state of affairs. We call this the *minimal change requirement*.

In the canonical account of counterfactuals (Stalnaker 1968; Lewis 1973), this requirement is implemented in a straightforward way: a given similarity ordering on worlds is assumed, and a counterfactual $\varphi > \psi$ is declared to be true at a world w in case ψ is true at all the worlds where φ is true and which are otherwise minimally different from w. Other approaches, such as premise semantics and causal accounts, while different in important ways, also incorporate the minimal change requirement.

The role of the minimal change requirement is to make sure that enough information about the actual state of affairs is retained when making a counterfactual assumption. For instance, in the setting of our main experiment, the assumption that switch A is down by itself does not lead to the conclusion that the light is off, because switch B might be down as well, and in that case the light would still be on. Nevertheless, (13) was judged true by a wide majority of our participants.

(13)   If switch A was down, the light would be off.

Intuitively, this is because in making the counterfactual assumption that A is down, the fact that switch B is actually up is retained, which by the laws of the circuit leads to the conclusion that the light is off. The fact that the position of switch B is retained is commonly attributed to a requirement that departure from actuality be as small as needed to accommodate the counterfactual assumption.

However, our data raise a serious challenge for the minimal change requirement —a challenge which is independent of the complications that arise from disjunctive antecedents. To see why, consider the Stalnaker-Lewis semantics.

Let us first give an intuitive argument, which assumes that the causal laws of our circuit are held fixed. Now, either toggling both switches counts as a bigger change than toggling just one, or it does not. If it counts as a bigger change, then the minimally different worlds where the switches are not both up are those worlds where only one of the switches is down. By the causal laws, the light is off in all these worlds. Therefore, $\neg(A \wedge B) > $ OFF is predicted to be true, in contrast to our findings. On the other hand, if toggling both switches does not count as a bigger change than toggling just one, then among the minimally different worlds in which

switch A is down we have worlds in which both switches are down. By the laws of our circuit, the light is on in these worlds. Therefore, $\overline{A} >$ OFF is not predicted to be true, again in contrast to our findings. In sum, whether or not toggling both switches count as a non-minimal change, our data are not accounted for.

Let us now give a purely logical argument, which relies only on the semantics of our sentences, and not on the causal laws of our scenario. The argument goes as follows. In order for $\overline{A} >$ OFF to be true, the light must be off in all minimally different worlds where switch A is down. Similarly, in order for $\overline{B} >$ OFF to be true, the light must be off in all minimally different worlds where switch B is down. Now consider a minimally different world in which the switches are not both up. This must be either a minimally different world in which A is down, or a minimally different world in which B is down. In either case, the light must be off in this world. Therefore, $\neg(A \wedge B) >$ OFF is predicted to be true. This shows that regardless of the specific notion of relative similarity that is at play, it is impossible for $\neg(A \wedge B) >$ OFF not to be true, given that $\overline{A} >$ OFF and $\overline{B} >$ OFF are true. Thus, our data are incompatible with a Stalnaker-Lewis semantics, regardless of the specific assumptions one makes about world similarity. Similar issues stemming from the minimal change assumption arise if we analyze our counterfactuals using existing accounts based on premise sets or on causal networks.

6 Conclusion

In this paper we reported on a web survey that we conducted to test the truth conditions of certain counterfactual conditionals. The results of this survey indicate that truth-conditionally equivalent antecedents can make different semantic contributions to the interpretation of conditionals they are part of. Assuming compositionality, this leads to the conclusion that the meaning of these antecedents—and sentential clauses more generally—should not be identified with their truth conditions. Instead, we argued that a natural explanation of our data is available in the more fine-grained framework provided by inquisitive semantics, on the basis of the inquisitive treatment of propositional connectives and of a general recipe for making accounts of conditionals sensitive to inquisitive content.

Our data also challenge existing theories of conditionals in a different way. Contrary to the predictions of the minimal change view (e.g. Stalnaker 1968; Lewis 1973), evaluating some counterfactuals involves looking at scenarios that depart in a non-minimal way from the actual world. Due to space limitations, in this paper we have not presented an alternative to the minimal-change view, as would be needed to account for our dataset. In future work we plan to provide such an alternative, relaxing the minimal change requirement so as to reconcile the observations that led to its adoption with our experimental findings.

References

Alonso-Ovalle, Luis. 2006. *Disjunction in alternative semantics*. Amherst, MA: University of Massachusetts Amherst PhD dissertation. http://scholarworks.umass.edu/dissertations/AAI3242324/.

Alonso-Ovalle, Luis. 2009. Counterfactuals, correlatives, and disjunction. *Linguistics and Philosophy* 32(2). 207–244. doi:10.1007/s10988-009-9059-0.

Ciardelli, Ivano. 2016. Lifting conditionals to inquisitive semantics. In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds.), *Semantics and Linguistic Theory (SALT) 26*, Ithaca, NY: LSA and CLC Publications.

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2013. Inquisitive semantics: A new notion of meaning. *Language and Linguistics Compass* 7(9). 459–476. doi:10.1111/lnc3.12037.

Ciardelli, Ivano & Floris Roelofsen. 2015. Alternatives in Montague grammar. In Eva Csipak & Hedde Zeijlstra (eds.), *Sinn und Bedeutung 20*, 161–178. Göttingen, Germany: University of Göttingen. http://semanticsarchive.net/Archive/jQzODVhO/.

Erlewine, Michael Yoshitaka & Hadas Kotek. 2016. A streamlined approach to online linguistic surveys. *Natural Language and Linguistic Theory* 34(2). 481–495. doi:10.1007/s11049-015-9305-9.

Fine, Kit. 2012. Counterfactuals without possible worlds. *The Journal of Philosophy* 109(3). 221–246. doi:10.5840/jphil201210938.

Fine, Kit. 2014. Truth-maker semantics for intuitionistic logic. *Journal of Philosophical Logic* 43(2-3). 549–577. doi:10.1007/s10992-013-9281-7.

Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*, 71–120. London, UK: Palgrave Macmillan. doi:10.1057/9780230210752.

Groenendijk, Jeroen & Martin Stokhof. 1990. Dynamic Montague grammar. In László Kálmán & László Pólos (eds.), *Papers from the 2nd Symposium on Logic and Language*, 3–48. Budapest, Hungary: Akadémiai Kiadó.

Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. Amherst, MA: University of Massachusetts PhD dissertation. http://semanticsarchive.net/Archive/jA2YTJmN.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Oxford, UK: Blackwell Publishing.

Horn, Laurence. 1972. *On the semantic properties of logical operators in English*. Los Angeles, CA: University of California PhD dissertation.

Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Formal Methods in the Study of Language* 135 Mathematical Center Tracts, 277–322. Amsterdam,

Netherlands.

Kratzer, Angelika. 1981. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic* 10(2). 201–216. doi:10.1007/bf00248849.

Lewis, David. 1973. *Counterfactuals*. Oxford, UK: Blackwell.

Lifschitz, Vladimir. 1990. Frames in the space of situations. *Artificial Intelligence* 46(3). 365–376. doi:10.1016/0004-3702(90)90021-q.

Nute, Donald. 1975. Counterfactuals and the similarity of words. *The Journal of Philosophy* 72(21). 773–778. doi:10.2307/2025340.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press. doi:10.1017/cbo9780511803161.

Roelofsen, Floris. 2013. Algebraic foundations for the semantic treatment of inquisitive content. *Synthese* 190(1). 79–102. doi:10.1007/s11229-013-0282-4.

van Rooij, Robert. 2006. Free choice counterfactual donkeys. *Journal of Semantics* 23(4). 383–402. doi:10.1093/jos/ffl004.

Rooth, Mats. 1985. *Association with focus*. Amherst, MA: University of Massachusetts Amherst PhD dissertation. http://scholarworks.umass.edu/dissertations/AAI8509599/.

Rooth, Mats. 1996. Focus. In Shalom Lappin (ed.), *Handbook of Contemporary Semantic Theory*, 271–297. Oxford, UK: Blackwell Publishing. doi:10.1111/b.9780631207498.1997.00013.x.

Schlenker, Philippe. to appear. The semantics/pragmatics interface. In Paul Dekker & Maria Aloni (eds.), *Cambridge Handbook of Formal Semantics*, Cambridge, UK: Cambridge University Press. http://ling.auf.net/lingbuzz/002545.

Schulz, Katrin. 2007. *Minimal models in semantics and pragmatics: Free choice, exhaustivity, and conditionals*. Amsterdam, Netherlands: University of Amsterdam PhD dissertation. http://hdl.handle.net/11245/1.272471.

Spector, Benjamin. 2007. Aspects of the pragmatics of plural morphology: On higher-order implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presuppositions and Implicature in Compositional Semantics*, 243–281. London, UK: Palgrave. doi:10.1057/9780230210752.

Stalnaker, Robert C. 1968. A theory of conditionals. In Nicholas Rescher (ed.), *Studies in Logical Theory*, 98–113. Oxford, UK: Blackwell. doi:10.1007/978-94-009-9117-0_2.

Veltman, Frank. 2005. Making counterfactual assumptions. *Journal of Semantics* 22(2). 159–180. doi:10.1093/jos/ffh022.

Breaking de Morgan's law in counterfactual antecedents

Lucas Champollion
Department of Linguistics
New York University
10 Washington Place
New York, NY 10003
champollion@nyu.edu

Ivano Ciardelli
FNWI / ILLC
Universiteit van Amsterdam
P.O. Box 94242
1090 GE Amsterdam
i.a.ciardelli@uva.nl

Linmin Zhang
Department of Linguistics
New York University
10 Washington Place
New York, NY 10003
linmin.zhang@nyu.edu