

Fixing de Morgan’s Laws in Counterfactuals*

Jacopo Romoli¹, Paolo Santorio², and Eva Wittenberg³

¹ Ulster University, Belfast, U.K.
j.romoli@ulster.ac.uk

² University of Maryland, College Park, College Park MD, U.S.A.
santorio@umd.edu

³ UC San Diego, La Jolla CA, U.S.A.
ewittenberg@ucsd.edu

Abstract

Classical semantics for counterfactuals are based on a notion of comparative similarity. These semantics are intensional, hence they predict that logically equivalent clauses can be substituted in counterfactuals *salva veritate*. A recent truth-value judgment study by Ciardelli, Zhang, and Champollion ([6]; CZC) appears to challenge both the idea that comparative similarity plays a role in counterfactual semantics and the prediction that logical equivalents are substitutable. CZC account for their data via an inquisitive semantics for disjunction and a semantics for counterfactuals that does not exploit the standard similarity algorithm. We report on a study consisting of two experiments that start from CZC’s general idea, but use a simpler scenario, manipulate negation more systematically, and add an extra task based on the selection of pictures. Our results replicate the differences found by CZC, but they also suggest that the effect is linked to the presence of overt negation rather than disjunction. We conclude that (i) inquisitive disjunction is neither necessary nor sufficient to account for the problem in full generality, and (ii) the evidence does not encourage rejecting a similarity semantics.

1 Introduction

1.1 Classical semantics and disjunctive antecedents

Classical theories of counterfactuals start from a simple idea, which is pithily put by Stalnaker: “Consider a possible world in which A is true, and which otherwise differs minimally from the actual world. “If A , then B ” is true (false) just in case B is true (false) in that possible world.” ([19]) Theories in the tradition of Stalnaker and Lewis ([19], [14], [15], a.o.) model this idea formally via a relation of comparative closeness, represented as ‘ \preceq_w ’. \preceq_w compares worlds with respect to their closeness to a benchmark world w : ‘ $w' \preceq_w w''$ ’ says that w' is closer to w than w'' . Comparative closeness singles out a set of ‘maximally close’ antecedent worlds, which are then used to evaluate the consequent. The schematic truth conditions of a counterfactual are:¹

- (1) $A \Box \rightarrow C$ is true at w iff all A -worlds that are maximally \preceq_w close to w are C -worlds

This paper discusses a challenge raised to classical accounts of counterfactuals by Ciardelli, Zhang, and Champollion ([6]; henceforth, CZC). This challenge has its roots in a classical

*We would like to thank Moysh Bar-Lev, Matthew Mandelkern and Paul Marty for very helpful discussion, and audiences at NELS 50 at MIT, Ulster University, University of Maryland, University of California San Diego, and University of Chicago. Work on this project was partially supported by the Leverhulme trust grant RPG-2018-425 to Jacopo Romoli.

¹Here we strike a compromise between Stalnaker’s and Lewis’s actual theories, using universal quantification like Lewis but making the so-called limit assumption like Stalnaker. For discussion of the latter, see [20], [10].

problem for semantics of counterfactuals, first raised by Kit Fine ([7]). Fine observes that counterfactuals with disjunctive antecedents seem to entail the 'simplified' counterfactuals with the individual disjuncts as antecedents.

- (2) If John had taken Syntax or Semantics, he would solve this exercise.
 \rightsquigarrow If John had taken Syntax, he would solve this exercise.
 \rightsquigarrow If John had taken Semantics, he would solve this exercise.

Examples like (2) seem to suggest that (3) is a valid principle of counterfactual logic.

- (3) **Simplification:** $(A \vee B) \Box \rightarrow C \models A \Box \rightarrow C, B \Box \rightarrow C$

Unfortunately, comparative closeness semantics does not validate Simplification, nor can it be tweaked to validate it without substantial consequences. The reason is that Simplification is inconsistent with two logical principles that standard semantics validates.

- (4) **Substitution:** $A \Box \rightarrow C \models A' \Box \rightarrow C$ (with A and A' logically equivalent)
(5) **Failure of Antecedent Strengthening:** $A \Box \rightarrow C \not\models A^+ \Box \rightarrow C$ (with $A^+ \models A$)

In the face of this problem, theorists have split into two camps. The first camp tries to accommodate Simplification as a broadly pragmatic effect (see e.g. [11], [12], [8], [2], [3]). The second camp tries to account for Simplification by rejecting Substitution (see e.g. [1], [16], [21]). CZC's proposal falls in this second camp.

1.2 CZC's challenge

CZC claim that Substitution fails in a specific and striking way: In counterfactual antecedents, classical de Morgan equivalences between disjunctions of negations and negations of conjunctions fail. In particular, CZC argue that counterfactuals where antecedents are disjunctions of negations don't entail counterfactuals where antecedents are negated conjunctions:

- (6) **de Morgan Failure in Counterfactuals:** $(\neg A \vee \neg B) \Box \rightarrow C \not\models \neg(A \wedge B) \Box \rightarrow C$

CZC claim is based on experimental evidence. They run one main experiment, preceded by two pre-tests and followed by three post-hoc tests; here we report their main experiment.

Participants were presented with a brief description of a scenario involving a lightbulb and two switches, together with a picture (Figure 1). The description makes clear that the light is on whenever the two switches, A and B, are in the same position (both up or both down). Subject are asked to provide a truth value judgment (TVJ), choosing between 'true', 'false', or 'indeterminate' on a sentence from one of five conditions:²

- (7) a. If A was down, the light would be off. $\bar{A} \Box \rightarrow C$
b. If B was down the light would be off. $\bar{B} \Box \rightarrow C$
c. If A or B was down, the light would be off. $(\bar{A} \vee \bar{B}) \Box \rightarrow C$
d. If A and B were not both up, the light would be off. $\neg(A \wedge B) \Box \rightarrow C$
e. If A and B were not both up, the light would be on. $\neg(A \wedge B) \Box \rightarrow \bar{C}$

Their result raise two main challenges for the classical theories. First, the counterfactuals with de Morgan-equivalent antecedents in (7-c) and (7-d) are endorsed at very different rates

²Following CZC, we use $\ulcorner \bar{A} \urcorner$ to denote a clause that is equivalent to the negation of A but does not include overt negation, and $\ulcorner \neg A \urcorner$ to denote clauses that involve overt negation.

(69.33% vs 22.04%). This appears to be an obvious violation of **Substitution**, which predicts that (7-c) and (7-d) are equivalent. Second, (7-d) is endorsed at a much lower rate than the corresponding counterfactuals with simple negated antecedents in (7-a) and (7-b) (66.02% and 65.11%, respectively). This seems to show that the following principle is invalid:

$$(8) \quad \text{Negated Conjunction: } \neg A \Box \rightarrow C, \neg B \Box \rightarrow C \vDash \neg(A \wedge B) \Box \rightarrow C$$

But, CZC argue, **Negated Conjunction** is a key principle for any similarity-based semantics.³

$$(9) \quad \text{Negated Conjunction: } \neg A \Box \rightarrow C, \neg B \Box \rightarrow C \vDash \neg(A \wedge B) \Box \rightarrow C$$

We now move to discuss two general approaches to accounting for these results.

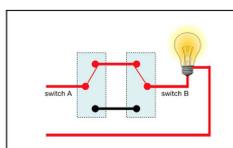


Figure 1: Stimulus of CZC’s main experiment.

2 Two approaches

In this section, we review two solutions to CZC’s challenge. One is due to CZC themselves and exploits inquisitive semantics; the other locates the source of the problem in negation.

2.1 The disjunction approach

The first approach is couched within an inquisitive semantics framework ([5]). Inquisitive semantics aims at unifying clause types across declarative and interrogative sentences: every clause denotes a set of propositions, i.e. a set of sets of worlds. Declarative and interrogative clauses differ in that the denotations of declarative sentences include a single weakest proposition, while the denotations of interrogative clauses don’t. Following [5], we say that declarative clauses include exactly *one alternative*, while interrogative clauses include *multiple alternatives*. Clauses that denote sets of multiple alternatives are said to have *inquisitive meanings*.

Crucially, in the inquisitive framework disjunctive clauses have an inquisitive meaning: $A \vee B$ denotes a set with two alternatives, namely $\{a, b\}$. Conversely, negated sentences have a declarative meaning and denote a set with a single alternative. This is at the basis of the different treatment of $(\neg A \vee \neg B) \Box \rightarrow C$ and $\neg(A \wedge B) \Box \rightarrow C$. CZC show how to import counterfactuals from classical truth-conditional semantics to an inquisitive framework. Let ‘ \Rightarrow ’ be a binary connective between propositions encoding any truth-conditional meaning for counterfactuals. We can ‘lift’ the basic meaning of \Rightarrow to an inquisitive framework in the following way:

³Standard axiomatizations of counterfactuals (see e.g. [4]) involve or entail the variant of Negated Conjunction that we get by swapping $\neg(A \wedge B) \Box \rightarrow C$ for the DeMorgan-equivalent $(\neg A \vee \neg B) \Box \rightarrow C$. Even if we reject **Negated Conjunction**, however, this doesn’t necessarily mean that comparative similarity should simply drop out of consideration. In fact, there are a number of semantics for counterfactuals that are still based on similarity and closeness and that invalidate principles logically related to Negated Conjunction: see e.g. [17], [9]; see also [18] for discussion. What the failure of **Negated Conjunction** establishes is that we cannot evaluate counterfactuals against a *fixed* closeness ordering, i.e. an ordering that is unaffected by counterfactual suppositions.

(10) $A \Box \rightarrow C$ is true at w iff all p in $Alt(A)$ there is some q in $Alt(B)$ s.t. $p \Rightarrow q$ is true in w

When an antecedent A has an inquisitive meaning, each of the alternatives denoted by A combines separately with the rest of the counterfactual. Hence (11-a) is equivalent to (11-b).

- (11) a. If A or B was down, the light would be off
 b. If A was down, the light would be off and if B was down, the light would be off

On the other hand, the antecedent of (12) only denotes one alternative, hence it differs from (11-a). This accounts for the asymmetry in endorsement between the two.

(12) If A and B were not both up, the light would be off

CZC also define a meaning for \Rightarrow that allows for failures of **Negated conjunction**. Surveying the details of their account would take us too far afield. Suffice it to say that they employ a nonstandard premise semantics (see [13]) where antecedents can selectively remove propositions from the premise set (not dissimilarly from [17]). This allows for the weakening of the logic required to invalidate **Negated conjunction**.

2.2 The negation approach

The second approach explains CZC's results by appealing to negation. Two accounts following this approach have been proposed. The first is due to Bar-Lev and Fox ([2, 3]; henceforth BLF) and is based on an implicature account of Simplification. The second is due to Kathrin Schulz ([18]) and also exploits an inquisitive framework, though in crucially different ways from CZC. For reasons of space, here we limit ourselves to considering the implicature account. So far as we can see, the two accounts make the same predictions for the cases we're interested in.

BLF's account exploits an exhaustification-based approach to implicatures. This approach postulates a syntactically realized exhaustivity operator EXH, which is appended to a sentence and returns the meaning of that sentence together with its implicatures. For example:

(13) EXH[some of the kids are playing] = some and not all of the kids are playing

BLF show how, given certain assumptions about EXH, we can derive Simplification as an implicature.⁴ (14-a) is predicted to be equivalent to the conjunction of (14-b)-(14-c).

- (14) a. EXH[If A or B was down, the light would be off]
 b. if A was down, the light would be off
 c. if B was down the light would be off

By itself, this mechanism doesn't explain the asymmetry between disjunction and negated conjunction. But BLF suggest a key extra assumption, i.e. that EXH is an alternative of negation. This predicts that antecedents with negated conjunctions give rise to a different (and stronger) Simplification implicature: (15-a) is equivalent to the conjunction of (15-b)-(15-d). Given that (15-d) is false in the relevant context, this accounts for CZC's results.

- (15) EXH[If A and B not both up, the light would be off] =
 a. If A was not up, the light would be off
 b. If B was not up, the light would be off
 c. If A was not up and B was not up, the light would be off

⁴Though see [16] for some arguments against an implicature approach to simplification.

- (17) a. If A or B were on the right, the see-saw would be balanced. DISJ
 b. If A and B were not both on the left, the see-saw would be balanced. NEG CONJ
 c. If A and B were both on the right, the see-saw would be balanced. CONJ
 d. If A or B were not on the left, the see-saw would be balanced. NEG DISJ

(17-a) and (17-b) are equivalent to CZC's positive disjunction and negated conjunction conditions. (17-c) is a sanity check predicted to be *false* by both approaches, corresponding to CZC's filler sentence. The crucial novel condition is (17-d), which involves both disjunction and overt negation. As we discuss below, comparing the latter to (17-a) and (17-b) is what allows us to test the potential role of connective type vs. the presence of overt negation. Each participant was randomly assigned to one condition and saw one item only.

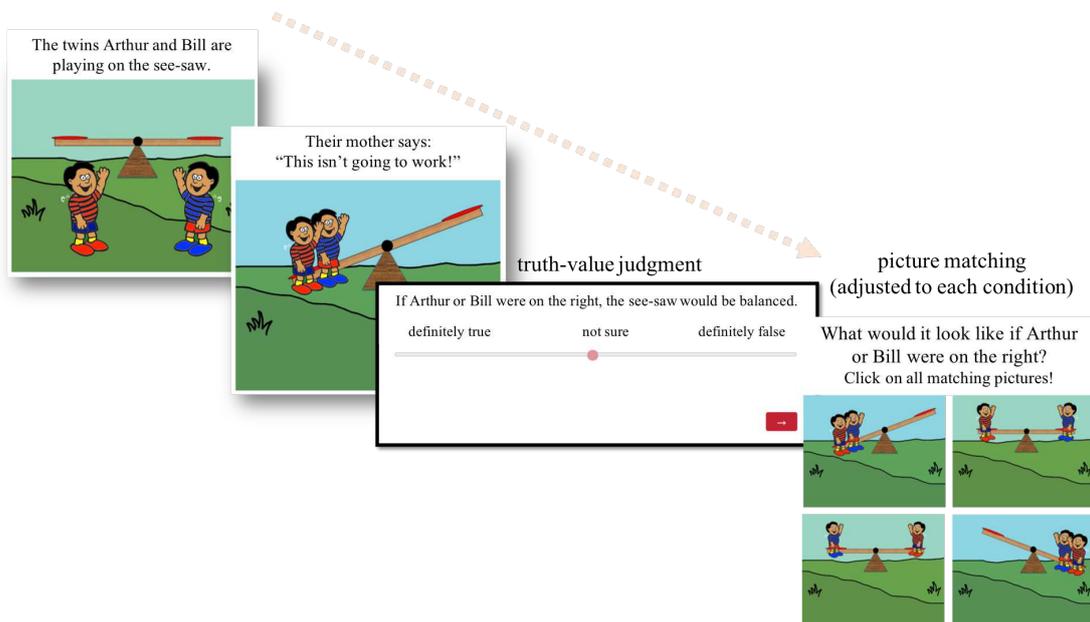


Figure 2: Trial structure in both our experiments.

Participants were presented with two tasks. The first was a TVJ task, as in CZC's study. The only difference was that, rather than giving participants the three options true/false/indeterminate, we presented them with a slider ranging from *definitely false* to *definitely true* (see Fig. 2) and asked them to adjust it according to their judgments. After the TVJ, participants were presented with a picture selection task. They were asked to select pictures that matched a linguistic description, via a question that involved the same antecedent they had seen for the TVJ task (e.g.: *What would it look like if Arthur and Bill were not both on the left?*). Subjects were explicitly told that they could select more than one sentence.

The crucial aspect of this task was in what conditions participants would select the bottom right picture in Fig. 2 ('BOTH RIGHT' picture) corresponding to both twins being on the right and the see-saw not being balanced. We take the selection of the BOTH RIGHT picture to suggest that participants imagined a scenario in which the two twins are on the right (and the see-saw is thus still imbalanced). This would, in turn, correlate with lower endorsement rates of the relevant sentences in the TVJ task.

Let us review the predictions of the two approaches. As for the TVJ task, the disjunction approach predicts higher endorsement rate whenever disjunction is present, regardless of whether negation is also involved. Therefore DISJUNCTION and NEGATED DISJUNCTION should receive higher endorsement rates than NEGATED CONJUNCTION. On the other hand, the negation approach predicts that the conditions involving negation will receive lower endorsement rates, so DISJUNCTION will score higher than both NEGATED DISJUNCTION and NEGATED CONJUNCTION. Both approaches predicts that the sanity check with positive conjunction will be judged definitely false. The predictions are summarized in (18) and (19) ($x > y$ indicates that x is predicted to score higher than y).

(18) **Predictions of the Disjunction approach:**
 $\{\text{DISJUNCTION, NEGATED DISJUNCTION}\} > \text{NEGATED CONJUNCTION}$

(19) **Predictions of the Negation approach:**
 $\text{DISJUNCTION} > \{\text{NEGATED DISJUNCTION, NEGATED CONJUNCTION}\}$

As for the picture selection task, the disjunction approach predicts that the BOTH RIGHT picture will be selected for the conditions not involving disjunction, while the negation approach predicts that it will be selected whenever negation is present. The key case is again the NEGATED DISJUNCTION condition. The disjunction approach predicts that the BOTH RIGHT picture *will not* be selected in this case, while the negation approach predicts that it *will* be selected.

3.1.2 Results

The results of the TVJ are plotted in the left half of Fig. 3. Bonferroni-corrected Mann-Whitney tests on contrast-coded conditions revealed significant differences between all comparisons (all $ps < .006$), except between the conditions that contained negation, regardless of whether they involved disjunction or conjunction ($p > .8$). The sanity check POSITIVE CONJUNCTION worked as expected, with participants overwhelmingly judging it as “definitely false”. We also replicated CZC’s difference between the POSITIVE DISJUNCTION and the NEGATED CONJUNCTION condition, with the former being endorsed at a lower rate than the latter. Crucially, however, we found equally low level of endorsement for the NEGATED DISJUNCTION condition.

The results of the picture-selection task are in Fig. 3. The y-axis corresponds to the TVJs, and participants are divided in three groups, depending on which picture(s) they selected. Most importantly for us is the green group, which includes subjects who selected the BOTH RIGHT picture in addition to the pictures where the twins are on different sides. This picture was selected most often when overt negation was present, regardless of the connective involved. That is, subjects selected all three pictures in the NEGATED CONJUNCTION and the NEGATED DISJUNCTION conditions. In the positive disjunction condition, on the other hand, they only considered the pictures where the twins are on different sides (red group). In sum, the results from both tasks reveal an effect of overt negation and are therefore in line with the negation approach, while they are challenging for the disjunction approach.

3.2 Experiment 2

In Experiment 2, we consider antecedents without binary connectives and investigate whether negation plays a role also in this case. As we will show, the results also suggest that **Negated conjunction** holds after all, once we control for the presence of negation.

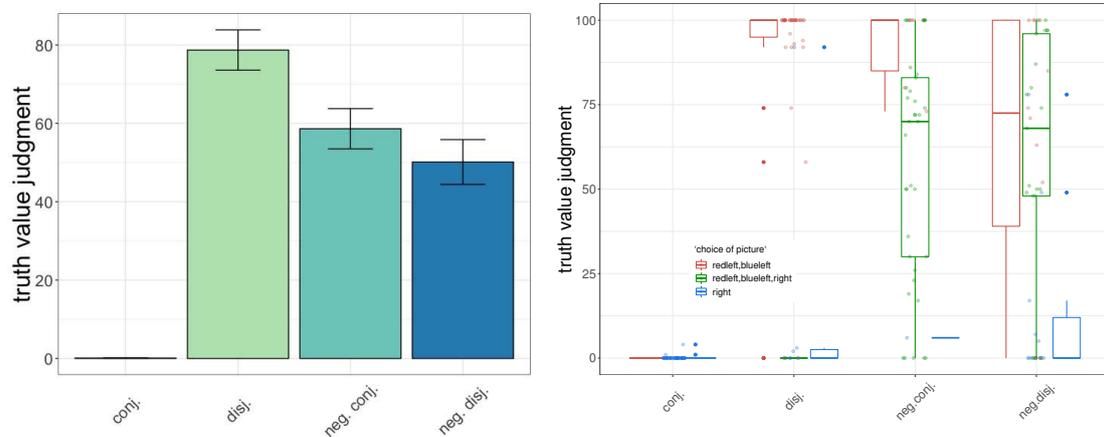


Figure 3: Results from the TVJ task (left) and picture choice task (right) of Experiment 1. Both plots show the TVJs across the four conditions. In the left plot we also divided participants in three groups, depending on their selection choices: participants who only selected the scenario in which both twins are on the right (blue group), participants who only selected the two scenarios in which each twin is one one side (red group), and participant who selected both the latter two scenarios, and in the addition, the one in which both twins were on the right (green group).

3.2.1 Methods

Participants 200 participants, recruited through AMT, participated in our experiment and were randomly assigned to one of the four conditions described below. They were all self-declared native speakers of English. No participant was excluded from the analysis.

Procedure and Material The procedure and tasks were the same as Experiment 1. The conditions involved again the NEGATED CONJUNCTION, which was now compared to counterfactuals with simple antecedents, either involving negation or not, all summarised in (20-a)-(20-e).

- (20)
- | | | |
|----|--|------------|
| a. | If A was on the right, the see-saw would be balanced. | POSITIVE A |
| b. | If B was on the right, the see-saw would be balanced. | POSITIVE B |
| c. | If A was not on the left, the see-saw would be balanced. | NEGATIVE A |
| d. | If B was not on the left, the see-saw would be balanced. | NEGATIVE B |
| e. | If A and B were not both on the left, the see-saw would be balanced. | NEG CONJ |

The predictions of the two approaches are as follows. In the TVJ task, the disjunction approach predicts that each of the simple counterfactuals should be endorsed at a higher rate than the NEGATED CONJUNCTION condition, while the negation approach predicts that all conditions involving negation, thus also NEGATIVE A and NEGATIVE B, will be endorsed at lower rate than the positive ones (at least, on the assumption that *A is on the right* and *B is on the right* work as alternative to each other in the relevant context). The predictions are summarized in (21) and (22).

- (21) **Predictions of the Disjunction approach:**
 $\{\text{POSITIVE A, POSITIVE B, NEGATIVE A, NEGATIVE B}\} > \text{NEGATED CONJUNCTION}$

- (22) **Predictions of the Negation approach:**
 $\{\text{POSITIVE A, POSITIVE B}\} > \{\text{NEGATIVE A, NEGATIVE B, NEGATED CONJUNCTION}\}$

In the picture selection task, both approaches expects participants to select the BOTH RIGHT picture in the NEGATED CONJUNCTION condition but not in the positive conditions.

3.2.2 Results

The results of the TVJ are plotted in Fig. 4. A Mann-Whitney tests with Bonferroni correction on contrast-coded conditions revealed significant differences between the positive and negative conditions ($p < .0001$), but no difference among the negative conditions ($p = .2$). We found lower endorsement of the NEGATED CONJUNCTION condition than in the POSITIVE simple antecedents conditions. Crucially, however, the NEGATIVE simple conditions were as low as the NEGATED CONJUNCTION one, showing that the key factor is the presence of overt negation.

The results of the picture-selection task are in Fig. 4 on the right. As before, the y-axis corresponds to the TVJs, and participants are divided in three groups, depending on the selection they made. As in Experiment 1, participants selected the BOTH RIGHT picture in addition to the pictures where the twins are on different sides whenever overt negation was present.

These results again lend support to the negation approach, and are challenging for the disjunction approach. In addition, they suggest that the **Negated conjunction** inference holds after all, once we control for the presence of overt negation. This is sufficient to fend off the second challenge raised by CZC. (This said, no approach predicts the results of the picture selection task, so more work is in order for a full account of the data.)

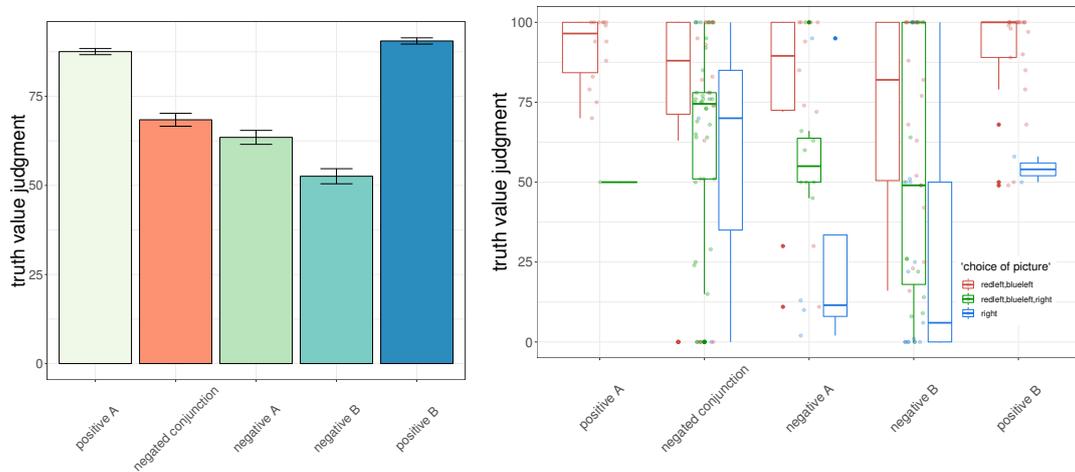


Figure 4: Results from the TVJ task (left) and picture selection task (right) of Experiment 2.

4 Discussion and conclusion

We tested counterfactuals with complex antecedents in a simple scenario across two experiments. Our subjects were asked to perform a TVJ and a picture selection task. Overall, we found a main effect of overt negation. In the TVJ task, subjects endorse a counterfactual with a

complex antecedent at a lower rate iff overt negation is present. In the picture selection task, participants consider the scenario in which both twins are on the right if and only if the sentence contains negation. Overall, our results lend support to an account that blames the phenomena observed by CZC on negation rather than disjunction. To be sure, we confirmed CZC's main result: sometimes disjunctions of negations and negations of conjunctions produce different TVJ. But this happens only when overt negation is involved, challenging the idea that the explanation lies with the inquisitiveness of disjunction. Moreover, our result in Experiment 2 suggests that, once we control for the presence of overt negation, **Negated conjunction** holds, addressing the challenge that CZC raise for comparative similarity semantics.

References

- [1] Luis Alonso-Ovalle. Counterfactuals, correlatives, and disjunction. *Linguistics and Philosophy*, 32(2):207–244, 2009.
- [2] Moshe Bar-Lev. *Free choice, homogeneity and innocent inclusion*. PhD thesis, The Hebrew University of Jerusalem, 2018.
- [3] Moshe Bar-Lev and Danny Fox. Free choice, simplification, and innocent inclusion. Ms Ecole Normal Supérieure and MIT, 2019.
- [4] John P Burgess. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22(1):76–84, 1981.
- [5] Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. *Inquisitive Semantics*. Oxford University Press, 2018.
- [6] Ivano Ciardelli, Zhang Linmin, and Lucas Champollion. Two switches in the theory of counterfactuals: A study of truth conditionality and minimal change. *Linguistic and Philosophy*, 2018.
- [7] Kit Fine. Truth, vagueness and logic. *Synthese*, 1975.
- [8] Michael Franke. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, 4(1):1–82, June 2011.
- [9] Thomas F. Icard. From programs to causal models. In *Proceedings of the 21st Amsterdam Colloquium*, pages 35–44, 2017.
- [10] Stefan Kaufmann. The limit assumption. *Semantics & Pragmatics*, 18(10), 2017.
- [11] Nathan Klinedinst. *Plurality and possibility*. PhD thesis, UCLA, 2007.
- [12] Nathan Klinedinst. (simplification of)² disjunctive antecedents. *MIT Working Papers in Linguistics*, 60, 2009.
- [13] Angelika Kratzer. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, 10(2):201–216, 1981.
- [14] David K. Lewis. *Counterfactuals*. Harvard University Press, Cambridge, MA, 1973.
- [15] David K. Lewis. Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, 2(4):418–446, 1973.
- [16] Paolo Santorio. Alternatives and truthmakers in conditional semantics. *Journal of Philosophy*, 115(10):513–549, 2018.
- [17] Paolo Santorio. Interventions in premise semantics. *Philosophers' Imprint*, 19, 2019.
- [18] Katrin Schulz. The similarity approach strikes back: Negation in counterfactuals. *Proceedings of Sinn und Bedeutung*, 22(2):343–360, May 2019.
- [19] Robert Stalnaker. A theory of conditional. In N. Rescher, editor, *Studies in Logical Theory*. Oxford, 1968.
- [20] Robert C. Stalnaker. A defense of conditional excluded middle. In William Harper, Robert C. Stalnaker, and Glenn Pearce, editors, *Ifs*, pages 87–104. Reidel, 1981.
- [21] Malte Willer. Simplifying with free choice. *Topoi*, 37(3):379–392, 2018.