Fixing de Morgan's law in counterfactual antecedents

Jacopo Romoli (Ulster University), Paolo Santorio (University of Maryland), and Eva Wittenberg (UCSD)

Background: Standard semantics for counterfactuals (see a.o. Stalnaker 1968, Lewis 1973, Kratzer 1981) treats them as involving universal quantification over a set of closest worlds, (1). This analysis makes two predictions: (i) SUBSTITUTION OF LOGICAL EQUIVALENTS (SLE): logically equivalent clauses are substitutable in the antecedents of counterfactuals; (ii) SIMPLIFICATION: the inference from $(p \vee q) \mathbin{\Box\!\!\rightarrow} r$ to the 'simplified' $p \mathbin{\Box\!\!\rightarrow} r$ and $q \mathbin{\Box\!\!\rightarrow} r$ is invalid (Fine 1975, Nute 1975 a.o.).

(1) $[\![p \mathbin{\Box\!\!\rightarrow} r]\!]^{w} = 1$ iff every closest p-world w' to w is such that $[\![r]\!]^{w'} = 1$

Previous study: Ciardelli, Zhang & Champollion 2018 (CZC) conducted a series of experiments, whose results challenge this view. They presented the scenario on the right involving two switches and a lightbulb; the light is on iff the switches are both up or both down. Participants were asked to evaluate the counterfactuals with De Morgan-equivalent antecedents (2a) and (2b) in a Truth Value Judgment task (TVJ).

(2)   a.   If switch A or switch B was down, the light would be off. $(\neg p \vee \neg q) \mathbin{\Box\!\!\rightarrow} r$
       b.   If switch A and switch B were not both up, the light would be off. $\neg(p \wedge q) \mathbin{\Box\!\!\rightarrow} r$

(2a) and (2b) are predicted to be equivalent in the standard semantics, given (i). Yet, in CZC's main experiment, (2a) is judged true by 69.33% of the participants, and (2b) by only 22.04% of them.

Two approaches have been proposed to account for CZC's results. The first is the INQUISITIVE approach by CZC themselves, who argue that we should abandon the standard view above for an inquisitive semantics-based approach (a.o. Ciardelli et al. 2018). Their system locates the source of the difference between (2a) and (2b) in the (inquisitive) denotation of disjunction and how it interacts with counterfactuals, effectively invalidating SLE (and validating SIMPLIFICATION). The second is the NEGATION-based approach by Bar-Lev and Fox 2019 (see also Bar-Lev 2018, Shulz 2018), who argue for keeping the standard semantics of counterfactuals coupled with an implicature-account of SIMPLIFICATION, while accounting for the difference in CZC's results by assuming that negation introduces more alternatives for implicature calculation.

Predictions: The two approaches can both account for the difference between (2a) and (2b) but make divergent predictions with respect to the source of their difference: the INQUISITIVE approach predicts that the effect should be linked to the presence of disjunction vs. conjunction, while the NEGATION approach predicts that it should be associated with the presence of negation.

Goals and motivations: First, we wanted to test the predictions above by systematically manipulating the presence of negation and controlling for its scope with respect to disjunction/conjunction. Second, CZC's participant exclusion rates ranged from 38% to 71% (excluded participants gave a 'wrong' judgment on an uncontroversial filler sentence). This suggests substantial comprehension difficulties of their scenario, casting doubts on how generalisable their results are. We therefore wanted to probe judgments about the same type of sentences in similar but considerably simpler scenarios. Finally, given the complexity of these sentences and their possible interpretations, we added a further measure to probe how speakers imagine the scenario and which alternatives they consider, in order to interpret their TVJs.

Experiments: We ran two experiments. In each experiment, 200 adult English native speakers judged a counterfactual sentence against a simple, intuitive scenario – two children trying to balance a see-saw – in a TVJ task (see Fig.1a). Each participant was randomly assigned to one of the four conditions in each experiment. The only difference between experiments was that Exp.1 used DP conjunction and disjunction (e.g. *if Arthur or Bill were not on the left ...*), and Exp.2 also used clausal disjunction (e.g. *if Arthur were not on the left or Bill were not on the left ...*) to better control for the scope of negation. After the TVJ, participants were asked to select one or more of the pictures that would match the counterfactual supposition corresponding to the antecedent of the sentence they just evaluated, e.g., "What would it look like if Arthur or Bill were on the right?". RESULTS: Bonferroni-corrected Mann-Whitney tests on contrast-coded conditions showed that Exp.1 resulted in significant differences between all comparisons (all $ps <$ .006), except between the conditions that contained negation, regardless of whether they involved disjunction
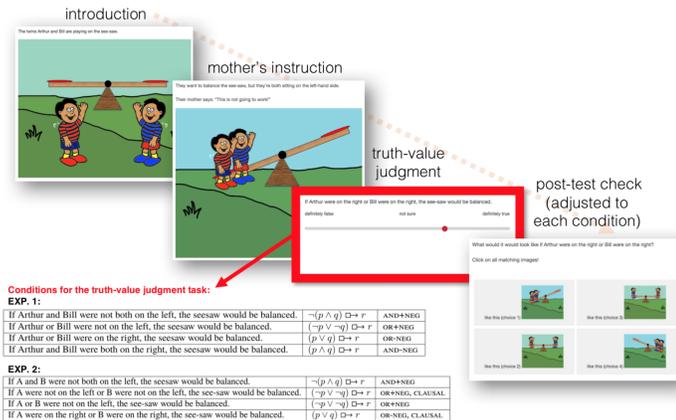
or conjunction ($p > .8$). In Exp.2, the effect of negation was again significant ($p < .02$), but we did not find an effect of clausality of disjunction ($ps > .13$.) Importantly, the picture-matching task revealed that people considered three alternative scenarios (A and B on different sides and A and B on the right, green in Fig.1d & 1e) only in negated sentences, regardless of the connective involved. In the positive disjunction condition, they only considered two alternative scenarios (A and B on different sides, red in Figs.1d & 1e). In addition, considering all three alternative scenarios was correlated with lower TVJs (y-axis in Figs.1d & 1e).

Discussion: In our results, we found a main effect of negation: systematically lower endorsement rates for the conditions with overt negation, regardless of the connective involved. In the picture choices, participants considered all three alternatives if and only if the sentence contained negation, again irrespective of the main connective and considering all three alternatives correlated with lower TVJ endorsement. Overall, our results suggest two conclusions. First, while there is something right in CZC's claim that sometimes disjunctions of negations and negations of conjunctions produce different results, this happens only when overt negation is involved, challenging CZC's idea that the explanation lies with the inquisitiveness of disjunction; That is, in line with the NEGATION-based approach, it was negation which played a crucial role in both tasks, not the type of connective. Moreover, in Exp.2 we found no evidence that this effect was due to the scope of negation.

Conclusion: We tested De Morgan's equivalences in the antecedent of conditionals building on CZC, using a simpler scenario across two experiments and systematically manipulating negation. In addition to participants' TVJs, we also collected picture-matching choices as an additional way to assess their interpretations of the relevant counterfactual sentences. Our results are challenging for CZC's proposal which encodes a difference between conjunction and disjunction, while entirely in line with the alternative account, which attributes the difference to negation. More generally, our results are compatible with a standard view on counterfactuals in which de Morgan's law is not broken, in counterfactual antecedents or elsewhere.
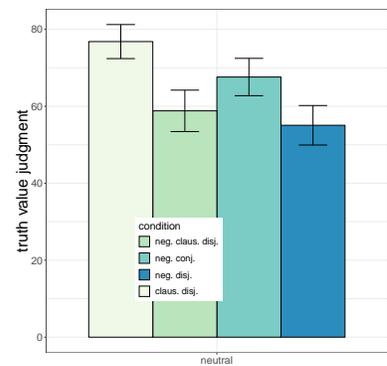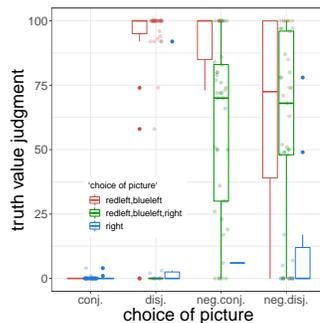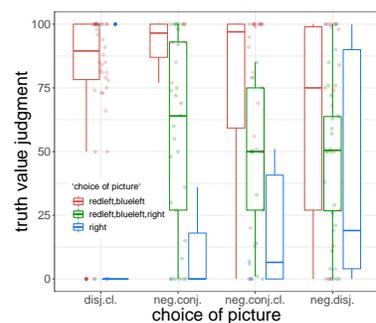

(a) Trial structure and conditions.
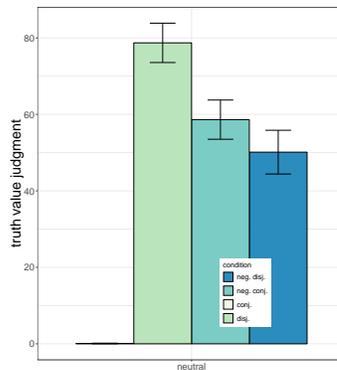
(b) Results of Exp.1 (TVJ).

(c) Results of Exp.2 (TVJ).

(d) Picture Choices Exp.1.

(e) Picture Choices Exp.2.

Figure 1: Design and Results.

Selected references: Bar-Lev & Fox 2019 *Free Choice, Simplification, and Innocent Inclusion* · Ciardelli, Zhang & Champollion 2018 *Two switches in the Theory of Counterfactuals* · Schulz 2018. *The similarity approach strikes back.*