Alternatives in counterfactuals: What is *right* and what is *not*[*]

Jacopo Romoli
University of Bergen

Paolo Santorio
University of Maryland, College Park

Eva Wittenberg
UC San Diego

Abstract Classical semantics for counterfactuals is based on a notion of comparative similarity and minimal change: *If A, would C* says that the most similar A-worlds are C-worlds. This semantics suffers from a well-known difficulty with disjunctive antecedents, which has generated a number of proposals combining the semantics of counterfactuals with alternatives (see e.g. Alonso-Ovalle 2009, Willer 2018, Santorio 2018, a.o.). In a recent study, Ciardelli, Zhang, and Champollion (2018b; henceforth, CZC) present new, related difficulties for the classical approach having to do with unpredicted differences between counterfactuals with De Morgan-equivalent antecedents, and related pattern of inferences. They propose a new semantics for counterfactuals, which builds on inquisitive semantics (see Ciardelli et al. 2018a) and gives up on comparative similarity and minimal change. We report a series of experiments extending their investigation. Our results replicate CZC's main effects, but they also indicate that those effects are linked to the presence of overt negation. We propose a novel account, based on three key assumptions: (i) the semantics for counterfactuals is standard; (ii) the meanings of disjunction and negation are associated with alternatives, which interact with the meaning of counterfactuals; (iii) the alternatives generated by negation are partially determined by the question under discussion (QUD). We compare our account with other existing accounts, including CZC's own proposal, as well as Schulz's (2019) and Bar-Lev & Fox's (2020) ones.

* We would like to thank Maria Aloni, Moysh Bar-Lev, Fabrizio Cariani, Ivano Ciardelli, Lucas Champollion, Julie Gerard, Matthew Mandelkern, Paul Marty, and Yasu Sudo for very helpful discussion, and audiences at NELS 50 at MIT, the Amsterdam Colloquium 2019, Ulster University, University of Maryland, University of California San Diego, University of Chicago, and UCL. Work on this project was partially supported by the Leverhulme trust grant RPG-2018-425 to Jacopo Romoli. The authors equally contributed to the project and are listed in alphabetical order.

Contents

1 Background

1.1 The standard semantics of counterfactuals

Theories of counterfactuals in formal semantics start from a simple idea, pithily put by Stalnaker (1968):

> 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.

Theories in the tradition of Stalnaker and Lewis (Stalnaker 1968, Lewis 1973a, Lewis 1973c, a.o.) formalize this idea via a relation of comparative closeness, represented as '$\leq_w$'. $\leq_w$ compares worlds with respect to their closeness to a benchmark world w: $\ulcorner w' \leq_w w'' \urcorner$ says that w' is closer to w than w''. Comparative closeness singles out a set of 'maximally close' (or 'minimally different') antecedent worlds, which are then used to evaluate the consequent. The schematic truth conditions of a counterfactual are as follows:[1]

(1) A $\Box\!\!\rightarrow$ C is true at w iff all A-worlds that are maximally$_{\leq_w}$ close to w are C-worlds.

The literature has produced plenty of variants of classical Lewis/Stalnaker semantics. In particular, so-called premise semantics frameworks are very popular in linguistic semantics (see e.g. Veltman 1976, 2005, Kratzer 1981, 1986, 2012).[2]

1 Here 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 Stalnaker 1981, Kaufmann 2017.

2 For our purposes, the differences between comparative closeness semantics and premise semantics are irrelevant, so we simply stick to the former. Also, we won't be concerned with how comparative closeness should be interpreted, accounts of which vary: on some accounts (e.g. Lewis 1979), worlds counts as closer the more they overlap in law and history (at least, roughly); on others, closeness tracks causal dependencies (see e.g. Kaufmann 2013, Santorio 2019 for recent accounts).

The focus of this paper is on the logical features of counterfactuals. Let us start by noticing that counterfactuals appear to invalidate antecedent strengthening (see (2)). Discourses that exemplify these failure, so-called ‘Sobel sequences’, are easy to find: (3) is a classical example. (3) shows that counterfactuals are nonmonotonic in the antecedent position: adding information to a counterfactual antecedent doesn't generally preserve truth value.

(2) **Failure of Antecedent Strengthening**: $\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \nvDash \mathsf{A}^{+} \mathrel{\Box\!\!\rightarrow} \mathsf{C}$ (with $\mathsf{A}^{+} \vDash \mathsf{A}$)

(3) If this match was struck, it would light.
If it was struck and it was wet, it wouldn't light.

The failure of Antecedent Strengthening is part of a cluster of related logical features, which set apart natural language counterfactuals from other conditional operators in extensional and modal logics (like e.g. the material conditional and the strict conditional).[3]

Semantics based on a notion of minimal change are designed just to account for these features. For illustration, here is how minimal change semantics explain the consistency of (3). It may be that the closest worlds where the match is struck are all worlds where the match is dry. In this case, the conditionals in (3) quantify over sets of worlds that are entirely distinct. So it is possible that the match lights in all of the former worlds, and doesn't light in all of the latter worlds, making (3) consistent.

It's also helpful to point to a principle that all standard semantics based on comparative closeness *do* vindicate (see Kraus et al. 1990).[4]

(4) **Negated Conjunction**: $\neg\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}, \neg\mathsf{B} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \vDash \neg(\mathsf{A} \wedge \mathsf{B}) \mathrel{\Box\!\!\rightarrow} \mathsf{C}$

This principle is obvious, if we assume a standard Boolean meaning for the connectives, and a comparative closeness semantics for counterfactuals. The closest $\neg(\mathsf{A} \wedge \mathsf{B})$-worlds are a subset of the union of the closest $\neg\mathsf{A}$-worlds and the closest $\neg\mathsf{B}$-worlds. Now, suppose that both the closest A-worlds and the closest B-worlds are C-worlds. It immediately follows that the closest $\neg(\mathsf{A} \wedge \mathsf{B})$-worlds are also C-worlds.

3 For discussion, see Stalnaker 1968, Lewis 1973a. For attempts at reconciling the data with a monotonic semantics, see the dynamic accounts in von Fintel 2001 and Gillies 2007.

4 **Negated Conjunction** is classically equivalent to the following:

(i) **Disjunction**: $\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}, \mathsf{B} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \vDash \mathsf{A} \vee \mathsf{B} \mathrel{\Box\!\!\rightarrow} \mathsf{C}$

Disjunction is often explicitly appealed to in counterfactual logics. For example, it is axiom (A4) in the axiomatization in Burgess 1981.

1.2 A classical problem: Simplification

Many contemporary debates about counterfactuals and alternatives have their roots in a classical problem first raised by Fine (1975). Fine observes that counterfactuals with disjunctive antecedents seem to entail the 'simplified' counterfactuals with the individual disjuncts as antecedents.

(5) 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 (5) seem to suggest that (6) is a valid principle of counterfactual logic.

(6) **Simplification**: $(A \vee B) \mathrel{\Box\!\!\rightarrow} C \vDash A \mathrel{\Box\!\!\rightarrow} C, B \mathrel{\Box\!\!\rightarrow} C$

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.

(7) **Substitution**: $A \mathrel{\Box\!\!\rightarrow} C \mathrel{=\!\!\mid\!\vDash} A' \mathrel{\Box\!\!\rightarrow} C$ (with A and A' logically equivalent)

(8) **Failure of Antecedent Strenghtening**: $A \mathrel{\Box\!\!\rightarrow} C \nvDash A^{+} \mathrel{\Box\!\!\rightarrow} C$ (with $A^{+} \vDash A$)

To illustrate why Simplification triggers a failure of Substitution or a validation of Antecedent Strengthening, notice that A is classically equivalent to $A \vee (A \wedge B)$. For example, *Alice runs* is equivalent to *Alice runs, or Alice and Bob run*. Now, via Substitution the following two counterfactuals are predicted to be equivalent:

(9) If Alice ran, Charlie would run.

(10) If Alice ran, or Alice and Bob ran, Charlie would run.

But from (10), via Simplification, we can infer:

(11) If Alice and Bob ran, Charlie would run.

Which is a strengthening of (9). The same maneuver can be repeated for any strengthening of (9).

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. Klinedinst 2007, Klinedinst 2009, Franke 2011, Bar-Lev 2018, Bar-Lev & Fox 2020). The second camp tries to account for Simplification by weakening or rejecting Substitution altogether (see e.g. Alonso-Ovalle 2009, Santorio 2018, Willer 2018). Ciardelli et al.'s (2018b) proposal falls in this second camp.

1.3 The novel challenges

Ciardelli, Zhang, and Champollion (2018b; henceforth, CZC) present experimental results that call into question two aspects of standard semantics. We first review the experiment, and then explain each of the two challenges. CZC report one main experiment, preceded by two pre-tests and followed by three post-hoc tests. Here we summarize their main experiment, and discuss one of the post-hoc tests in the next subsection.

CZC's participants were presented with a description of a scenario involving a lightbulb and two switches, together with a picture (Figure 1).[5] The description explains that the light is on whenever the two switches, A and B, are in the same position (both up, or both down). Participants are asked to provide a truth-value judgment (truth-value judgment), choosing between 'true', 'false', or 'indeterminate' on a sentence from one of the five conditions in (12).[6] The results are summarised in Table 1.

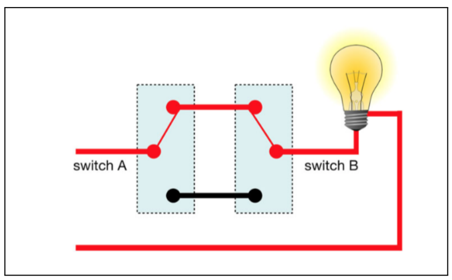


Figure 1 Picture context of CZC's main experiment.

(12)
a. If switch A was down, the light would be off. $\bar{A} \mathrel{\Box\!\!\rightarrow} C$
b. If switch B was down, the light would be off. $\bar{B} \mathrel{\Box\!\!\rightarrow} C$
c. If switch A or switch B was down, the light would be off. $(\bar{A} \vee \bar{B}) \mathrel{\Box\!\!\rightarrow} C$
d. If switch A and switch B were not both up, the light would be off. $\neg(A \wedge B) \mathrel{\Box\!\!\rightarrow} C$

5 The figure was adapted by CZC from multiway switches, Colin M.L. Burnett, CC BY-SA 3.0, via Wikimedia Commons.

6 Ciardelli et al. (2018b) also included a filler sentence:

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

Ciardelli et al. (2018b) assume that, given the scenario, the only acceptable truth-value judgment response for (i) is 'false'. 38% of their participants failed to give this response and were consequently excluded from analysis.

e. If switch A and switch B were not both up, the light would be on. $\neg(A \wedge B) \mathrel{\Box\!\!\rightarrow} \bar{C}$

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

Table 1 Results from CZC's main experiment.

Their results, if taken at face value, raise two main challenges. Let's see each of them in turn.

1.3.1 Failure of Substitution for De Morgan equivalents

The first challenge is that the results appear to illustrate a failure of **Substitution**. In standard propositional logic, the disjunction of two negations and the negation of a conjunction are equivalent (this is one of the so-called 'De Morgan equivalences').

(13) **De Morgan Equivalence**: $\neg A \vee \neg B \mathrel{=\!\!|\!\!\models} \neg(A \wedge B)$

For illustration, the two sentences in (14) are predicted to be De Morgan equivalent (assuming that connectives in natural language work as in classical propositional logic).

(14) a. Swith A is not up or Switch B is not up.
b. Switch A and Switch B are not both up.

Hence, if Substitution holds and connectives in natural language have their Boolean meanings, we expect counterfactuals of the form $(\bar{A} \vee \bar{B}) \mathrel{\Box\!\!\rightarrow} C$ and $\neg(A \wedge B) \mathrel{\Box\!\!\rightarrow} C$ ((12)-c and (12)-d in CZC's experiment) to be equivalent (on the assumption that *Switch A is not up* and *Switch A is down* are equivalent in meaning). Yet these counterfactuals are judged to be true at very different rates (69.33% vs 22.04%). So Substitution seems to fail. In particular, it seems that (12)-c does not entail (12)-d.

(15) **De Morgan Failure in Counterfactuals**:
$(\neg A \vee \neg B) \mathrel{\Box\!\!\rightarrow} C \not\models \neg(A \wedge B) \mathrel{\Box\!\!\rightarrow} C$

Sentence	# Responses	True	%	False	%	Indet.	%
$\neg A \boxright C$	36	27	75.00%	1	2.78%	8	22.22%
$\neg B \boxright C$	43	28	65.12%	7	16.28%	8	18.60%
$(\neg A \vee \neg B) \boxright C$	80	48	60.00%	16	20.00%	16	20.00%

Table 2 Results from one of the follow-ups by CZC.

1.3.2 Failure of Negated Conjunction

The second challenge is related to Negated Conjunction, repeated from above:

(16) **Negated Conjunction**: $\neg A \boxright C, \neg B \boxright C \vDash \neg(A \wedge B) \boxright C$

(16) is also called into question by CZC's data. In particular, notice that (12)-d is endorsed at a much lower rate (22%) than the corresponding counterfactuals with simple negated antecedents in (12)-a and (12)-b (66% and 65%, respectively). CZC take this to be a strike against all counterfactual semantics based on a notion of minimal change. They suggest dropping altogether the idea that counterfactuals quantify over minimally different worlds, and switching to a different kind of semantic theory that incorporates causal notions, 'background semantics'. For reasons of space here we won't discuss background semantics. Rather, we will focus on CZC's empirical claim that Negated Conjunction is invalidated.[7]

1.4 Other data, 1/2: CZC's follow up

In addition to their main experiment, CZC ran a number of follow-up experiments. The most relevant for us was intended to control for the effects of overt negation.

The set of sentences that CZC use for this follow-up is:

(17) a. If switch A was not up, the light would be off. $\bar{A} \boxright C$
b. If switch B was not up, the light would be off. $\bar{B} \boxright C$
c. If switch A or switch B was not up, the light would be off. $(\bar{A} \vee \bar{B}) \boxright C$

Their results are summarized in Table 2.

The main result is that 60% of their participants judged the sentence with

7 We should notice that abandoning the requirement of minimal change is technically compatible with preserving a notion of comparative closeness in the semantics. See e.g. the semantics in Santorio 2019, Icard 2017; CZC's semantics can be seen as a semantics in a similar vein. See also Schulz 2019 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.

negated disjuncts true. We should flag that, for this follow-up, the exclusion rate of the participants (based on a judgment about a filler sentence that is uncontroversially false in the scenario) is very high, i.e. 71.66%. The reason behind this is unclear. CZC themselves conjecture that, with the addition of negation the sentences might have been harder to process, and participants might have been confused. We suspect that, in part, this is also due to the fact that their scenario is fairly complex and difficult to understand. Part of the motivation for our project is providing participants with a simpler and more intuitive scenario.

1.5 Other data, 2/2: Schulz's experiment

Schulz 2019 runs a follow up experiment, building on CZC's main experiment. Participants were presented with a modified switches scenario, in which switch A is up, switch B is down, and in addition the electricity is not working (see Figure 2).

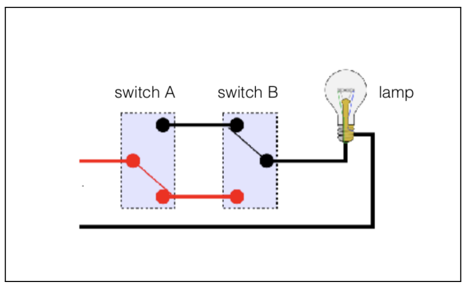


Figure 2 Picture context of Schulz's (2019) main experiment.

The task for participants was to evaluate a few sentences via a slider with five positions, the first of which is labeled 'true' and the last of which is labeled 'false'. The key target sentence is:

(18) If the electricity was working and switch A and switch B were not both up, then the light would (still) be off.

Schulz reports that 59% of her participants rate the sentence as false (26% rate as true, and 15% as indeterminate).

We won't have space to discuss extensively the predictions for Schulz's target sentence here. (In particular, we are not going to discuss in detail the predictions that CZC's account makes for (18), since these predictions crucially depend on a part of their theory, the 'background semantics', which we don't examine in detail.) But we will add (18) as one of the data points to explain.[8]

8 Another relevant experiment is in McHugh & Cremers 2019. McHugh and Cremers use a further

2 The current study

This paper aims at pushing forward the debate, both on the experimental and the theoretical side. On the experimental side, we use a new and more intuitive scenario, and supplement the truth-value judgment task with a new task, on which participants are asked to select a picture matching a counterfactual antecedent. On the theoretical side, we investigate the role of overt negation in the generation of alternatives. In particular, our experiments aim at distinguishing a view on which negation does not introduce any new alternatives (call this the '**Classical Negation**' view) and a view on which negation introduces alternatives, on a par with disjunction (call this the '**Alternative Negation**' view).

Let us give a quick overview of what's to come. The first key data point we investigate concerns the comparison between two conditionals with disjunctive antecedents. One involves a simple disjunction of two sentences $\bar{\mathsf{A}}$ and $\bar{\mathsf{B}}$ (as in (19-a)), the other a disjunction of two sentences that involve overt negation, i.e. sentences of the form $\neg\mathsf{A}$ and $\neg\mathsf{B}$ (as in (19-b)). The two disjunctions are equivalent in a standard truth conditional framework, and differ merely because of the presence of overt negation.

(19) a. $(\bar{\mathsf{A}} \vee \bar{\mathsf{B}}) \mathrel{\Box\!\!\rightarrow} \mathsf{C}$
b. $(\neg\mathsf{A} \vee \neg\mathsf{B}) \mathrel{\Box\!\!\rightarrow} \mathsf{C}$

The Classical Negation hypothesis predicts that the sentences in (19) should be equivalent, while the Alternative Negation view is compatible with the existence of a difference, arising from the alternatives introduced by overt negation in (19-b). Our Exp. I and III test these predictions.

The second data point we investigate concerns counterfactuals whose antecedents involve no binary connectives, and which differ only with respect to the presence of overt negation. I.e., we investigate pairs of sentences like the following:

(20) a. $\bar{\mathsf{A}} \mathrel{\Box\!\!\rightarrow} \mathsf{C}$
b. $\neg\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}$

Again, the Classical Negation approach predicts no difference between the two, while the Alternative Negation approach is compatible with a difference in meaning between (20-a) and (20-b). We test these predictions in Exp. II.

Here are our results, in summary. Exp. I shows that counterfactuals like (19-b) are evaluated differently from their counterpart involving no overt negation, in

elaboration of the CZC scenario to test for cases involving double negation. They take their results to support a version of Schulz's (2019) account. For reason of space, we do not discuss their experiment in detail; we leave it to the reader to check that the theory in §4, modified as they suggest, can account for their predictions as well.

(19-a). This finding is replicated and refined in Exp. III, where the scope of negation with respect to disjunction is explicitly controlled for. In Exp. II, we find again that there is a difference between counterfactuals that involve overt negation and those that don't, even in absence of binary connectives. Again, this is in line with the predictions of the Alternative Negation approach, but conflicts with the Classical Negation view. Finally, in Exp. IV, we assess the general effect of overt negation on judgments about this type of sentences. While we find an effect of overt negation, this effect is too small to account for the full differences we find between positive and negative sentences involving counterfactuals.

As we discuss below, these findings partially confirm and partially conflict with CZC's conclusions. On the one hand, our results, like CZC's, suggest that Substitution fails in counterfactual antecedents. On the other hand, our results are consistent with the validity of the Negated Conjunction inference. On the basis of these results, we argue that a semantics based on minimal change should be maintained after all, coupled with a version of the alternative negation approach.

After discussing our experimental findings, we propose a semantics for counterfactuals that accounts for all experimental results in the literature, as well as our own. The key move is that negation introduces alternatives, in a similar way to disjunction. Differently from what happens in standard inquisitive systems, alternatives are partially determined by the question under discussion (QUD). This builds on the idea, frequently encountered in the literature, that negation interacts with various alternative-sensitive mechanisms in natural language.

The rest of the paper is organised as follow. In §3, we report the experiments and discuss their results. In §4 we propose our account, and in §5 we compare it to all the others. We conclude the paper in section §6.

3 Experiments

3.1 Overview of the Experiments

Our experiments follow up on Ciardelli et al.'s (2018b) study, but with three key changes: (i) we use a simpler, easy-to-understand scenario involving intuitive physics,[9] (ii) we manipulate negation more systematically; and (iii) we add a picture selection task to the truth-value judgment task to probe the conceptualization of the antecedents, as a proxy to further understand which alternatives are considered in evaluating the counterfactuals we investigate.

9 The intuitiveness of the scenario is particularly important here, because performance on tasks judging conditional scenarios has been shown to vary systematically depending on how intuitively accessible their content is (e.g., Sperber et al. 1995, Fiddick et al. 2000).

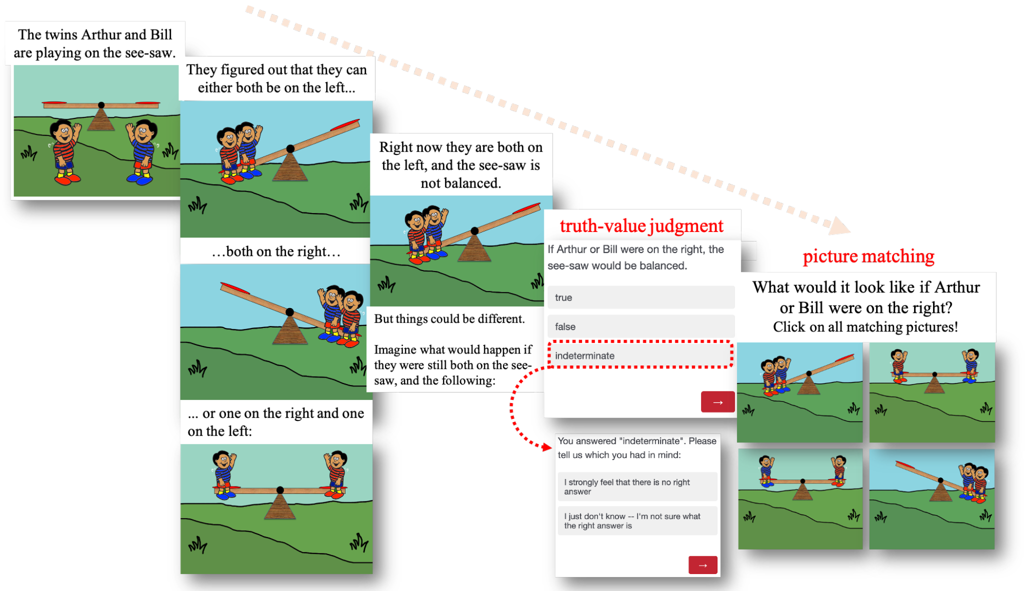


Figure 3 Structure of the experiments below. The basic scenario depicts two twins who are trying to balance a see-saw. Instructions varied only slightly between studies and conditions (see Methods sections below).

In all of the experiments described below, we use the scenario in Fig. 3. In this scenario, two twins named Arthur and Bill are trying to balance a see-saw. The introduction explicitly mentions the relevant counterfactual possibilities (i.e. 'They figured they can either both be on the left, both on the right ...' etc; see Fig. 3).

The critical sentences contain combinations of disjunction, conjunction, and negation in the antecedent of the counterfactuals. Our studies probe participants' intuitions on the the same conditions as in Ciardelli et al.'s (2018b) study, but in addition, explicitly and separately addressing the roles of negation (Exp. I, e.g. (28)), the presence and absence of connectives (Exp. II, e.g. (22)), wide and narrow scope (Exp. III, e.g. (23)), and the presence and absence of counterfactuality itself (Exp. IV, e.g. (24)).

(21) If Arthur or Bill were not on the left, the see-saw would be balanced.
NEGATIVE DISJUNCTION $(\neg A \vee \neg B) \mathrel{\Box\!\!\rightarrow} C$

(22) If Arthur was on the right, the see-saw would be balanced.
SIMPLE POSITIVE $\bar{A} \mathrel{\Box\!\!\rightarrow} C$

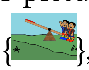
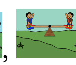
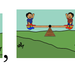
(23) If Arthur was not on the right or Bill was not on the right, the see-saw would be balanced.

CLAUSAL NEGATIVE DISJUNCTION $(\neg A \vee \neg B) \Box\!\!\rightarrow C$

(24) Arthur or Bill is not on the left.
NO COUNTERFACTUAL $\neg A \vee \neg B$

Our main goal is to test the predictions of the Classical Negation and Alternative Negation approaches, outlined above. That is, we investigate whether negation behaves classically, or whether it contributes to generating alternatives.

3.2 Analyses

In all studies, we analyze two Dependent Variables. The first is the proportion of each choice (True, False, or Indeterminate) in each condition of the truth-value judgment task. The second is an index of the selection of pictures including the choice of the picture with both twins on the right (either on its own or together with the selection of other pictures). That is, the index included the following combinations of pictures: { {}, {, }, {, }, {, , } }. The rationale for this index was that any choice including the picture in which both boys are on the right (i.e. ) would be showing consideration of this counterfactual possibility.[10] Each study was replicated at least once with minimal changes in the design, with consistent results throughout.[11]

The statistical analysis of the truth-value judgment task was conducted in two steps. The first was a multinomial regression, which is an extension of binomial logistic regression, used when the Dependent Variable consists of unordered categorical variables, like our TRUE, FALSE, and INDETERMINATE answers. We built an omnibus model, predicting the probability of each categorical answer by condition. We always used the INDETERMINATE option as reference level, and adjusted the reference condition for each experiment. The second step in the statistical analysis was a set of planned comparisons, using likelihood-ratio tests for comparing nested multinomial models against intercept-only multinomial models. The reason for this latter set of planned tests is that the strength of evidence for an effect, e.g. negation, is relevant to assessing the overall support of the data for our main hypotheses, and thus, each effect needed to be directly statistically examined (Wittenberg & Levy 2017).

It is important to note that the factors in ours and Ciardelli et al.'s (2018b)

10 These analyses were pre-registered on the Open Science Framework (https://osf.io/47wrp), where all data and analyses scripts can be found as well (https://osf.io/2etx7). Minimal adjustments were made, compared to the preregistration, to avoid what turned out to be redundant analyses.

11 Data and analyses for those replications can be found on in our Open Science Repository in the folder "Additional Experiments"; https://osf.io/2etx7/.

experiments are not always fully crossed. For instance, in Exp. I, the positive conjunction (*If A and B were on the right…*) has different truth conditions from the rest (according to all theoretical approaches). This condition is therefore used as sanity check and not included in the analyses; as a consequence of this design, we could not calculate interactions. Rather, we identified the relevant pairwise comparisons *a priori*, and applied Bonferroni-corrections.

For each trial in which the INDETERMINATE option was chosen, we obtained an additional measure, namely the proportion of ignorance answers (e.g., "I just don't know") over ignorance and nonexistence ("There is no right answer") answers. This measure was collected to investigate whether the indeterminate selection indicates uncertainty on the part of participants, and was analyzed using a binomial intercept-only regression.

For the picture matching task, we created an index of the selection of pictures containing the picture with both twins on the right, either by itself or together with other pictures. We performed a binomial regression, predicting the proportion of this index to all other choices by condition.

None of the models includes any random effects structure, since all manipulations were distributed across participants, and there were no repeated measures; all predictors were contrast-coded when feasible and justified by the analysis.

3.3 Experiment I

3.3.1 Methods

Participants. 200 self-declared native speakers of English participated in this experiment. Here and in all other experiments, we restricted the participant pool to Amazon Mechanical Turk users with an IP address in the United States, with a completed task acceptance rate of 95% or higher, and with at least 100 tasks completed. In addition, we used the services of CloudResearch to recruit only workers who had passed strict attention tests (Litman et al. 2017). Participants were randomly assigned to one of the four conditions below. Five workers were excluded from the analysis because of a coding error.

Procedure and Materials. We used the scenario in Fig. 3, and manipulated connective type and polarity resulting in the following four conditions:

(25) If Arthur or Bill were on the right, the see-saw would be balanced.
DISJUNCTION $(\bar{A} \vee \bar{B}) \,\Box\!\!\rightarrow C$

(26) If Arthur and Bill were not both on the left, the see-saw would be balanced.
NEGATED CONJUNCTION $\neg(A \wedge B) \,\Box\!\!\rightarrow C$

(27) If Arthur and Bill were both on the right, the see-saw would be balanced.
CONJUNCTION $(A \wedge B) \mathrel{\Box\!\!\rightarrow} C$

(28) If Arthur or Bill were not on the left, the see-saw would be balanced.
NEGATIVE DISJUNCTION $(\neg A \vee \neg B) \mathrel{\Box\!\!\rightarrow} C$

The conditions in (25) and (26) are equivalent to Ciardelli et al.'s (2018b) positive disjunction and negated conjunction conditions. (27) is a "sanity check", which all participants who understand the scenario should judge as unequivocally false. The crucial novel condition is (28), which involves both disjunction and overt negation. As discussed, comparing (28) to (25) and (26) allows us to test the predictions of the two approaches to negation.

In this and all other experiments except Exp. IV, participants were presented with two tasks. The first was a truth-value judgment task, as in Ciardelli et al. (2018b), with three answer options: True, False, and Indeterminate (see Fig. 3). In order to further understand our participants' interpretation of the counterfactual sentences, we added a follow-up question to any 'indeterminate' answer, asking whether by 'indeterminate' they meant that they strongly felt that there was no right answer, or whether they were just not sure what the right answer was.

The second task followed the truth-value judgments: Participants were asked to select, among four pictures, all those that represented a possible scenario verifying the antecedent of that same counterfactual. That is, they were asked "What would things look like if A?", where A is the antecedent of the counterfactual they had just evaluated. Participants were explicitly told that they could select more than one picture. This additional task allowed us to probe more directly what kind of scenarios they took to be relevant for evaluating the counterfactual, rather than relying on truth-value judgment data alone.

3.3.2 Results

First, we note that participants understood the task very well: 98% of all participants in this condition correctly judged the sanity check (POSITIVE CONJUNCTION) as false; second bar in Fig. 4. In addition, 98% of the participants in this condition correctly picked the picture in which both boys were on the right (either by itself or at least among others, Fig. 3.3.2). This gives us confidence that participants understood the scenario and the task correctly.

As can be seen from the plot (left in Fig. 4), among the remaining critical conditions of interest (NEGATED CONJUNCTION, NEGATIVE DISJUNCTION, and POSITIVE DISJUNCTION), those involving negation were associated with lower true choices than POSITIVE DISJUNCTION. That is, we replicated Ciardelli et al.'s (2018b) difference in judgments between POSITIVE DISJUNCTION and NEGATED CONJUNCTION,

omnibus multinomial analysis

		(intercept)	disjunction, negated	conjunction, negated
Coefficients:	FALSE	0.956	-1.65	-2.66
	TRUE	1.856	-2.55	-2.95
Std. Errors:	FALSE	0.526	0.634	0.688
	TRUE	0.481	0.597	0.594

pairwise comparisons

contrast	LR statistics	*p-value*
disjunction, negative | conjunction, negated	3.47	0.18
disjunction, positive | conjunction, negated	36.43	<0.0001***
disjunction, positive | disjunction, negative	22.97	<0.0001***

Table 3 Results of the statistical analyses conducted for Experiment I: Parameters of the omnibus model above; likelihood-ratio (LR) statistics and *p*-values below. Bonferroni-corrected significance is indicated by *.

but in addition, we also found the NEGATIVE DISJUNCTION condition to be endorsed less than the POSITIVE one.

To examine the pattern statistically, we fitted a multinomial regression model predicting the proportion of choices as a function of CONDITION with indeterminate answers and positive disjunction as reference levels. We report the coefficients of the model in Tab. 3.3.2. To obtain *p*-values and likelihood-ratio statistics for the three pairwise comparisons of interest, we restricted CONDITION to two levels of interest at a time and conducted nested-models comparisons between models with and without the given conditions of interest as fixed effect.

The results, reported in (Tab. 3.3.2, showed a significant difference between POSITIVE DISJUNCTION and each of the negative conditions. On the other hand, we found no significant difference in truth-value judgments between NEGATIVE DISJUNCTION and NEGATED CONJUNCTION.

As Fig. 4 (right) shows, the vast majority (77%) of participants who chose the indeterminate answer said they selected it because there genuinely was no right answer (rather than them being uncertain about it). An intercept-only binomial regression confirmed that the difference between the two choices was statistically significant (β=1.23, z=4.05, p<.0001).

The picture choices (Fig. 3.3.2) were analyzed in an analogous way to the truth-value judgment, but instead of building a multinomial model, we used a binary

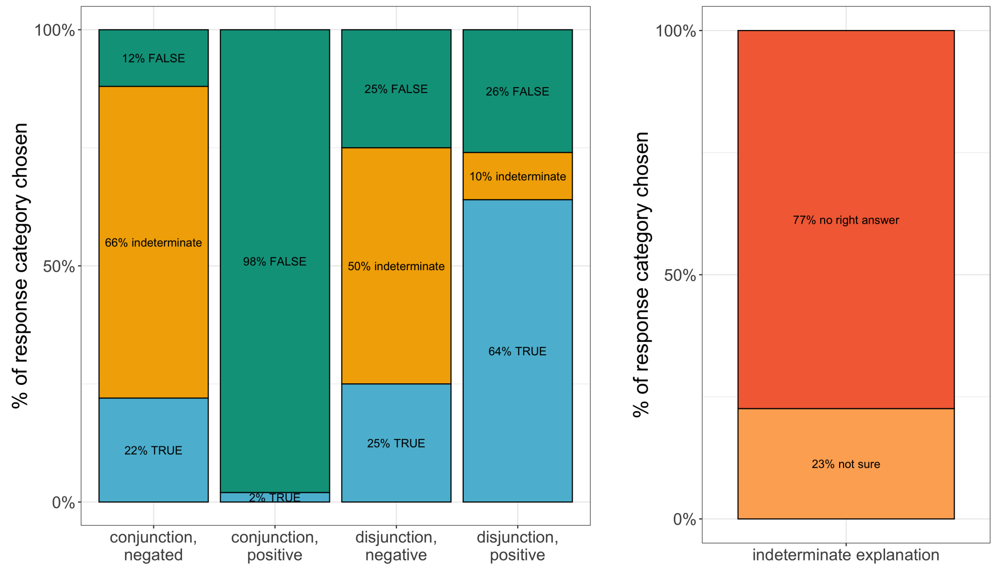


Figure 4 Results of the truth-value judgment task in Experiment I (left), and a detailed look at the Indeterminate choices in Experiment I (right).

coding scheme for the picture choices, since we were interested in whether or not a participant's choice included the picture depicting both boys on the right. Like in the truth-value judgment task, we excluded answers to the sanity check POSITIVE CONJUNCTION, and then fitted a binomial model with positive conjunction as reference level, resulting in β-Estimates of >.39, and t-values of >4.27. To obtain p-values and likelihood-ratio statistics for the three pairwise comparisons of interest, we restricted CONDITION to two levels of interest at a time (e.g. disjunction, positive vs disjunction, negative) and conducted nested-models comparisons.

These comparisons revealed that the picture choices were in line with the truth-value judgments: POSITIVE DISJUNCTION led to significantly fewer choices including both kids on the right than both NEGATIVE DISJUNCTION (p<.0001), and NEGATED CONJUNCTION (p<.0001), but there was no difference in picture selection between NEGATIVE DISJUNCTION and NEGATED CONJUNCTION (p>.9).

3.3.3 Discussion of Experiment I

As discussed above, the Classical Negation approach predicts that the NEGATIVE DISJUNCTION condition should have clustered together with the POSITIVE one and both should differ from the NEGATED CONJUNCTION condition, both in truth-value

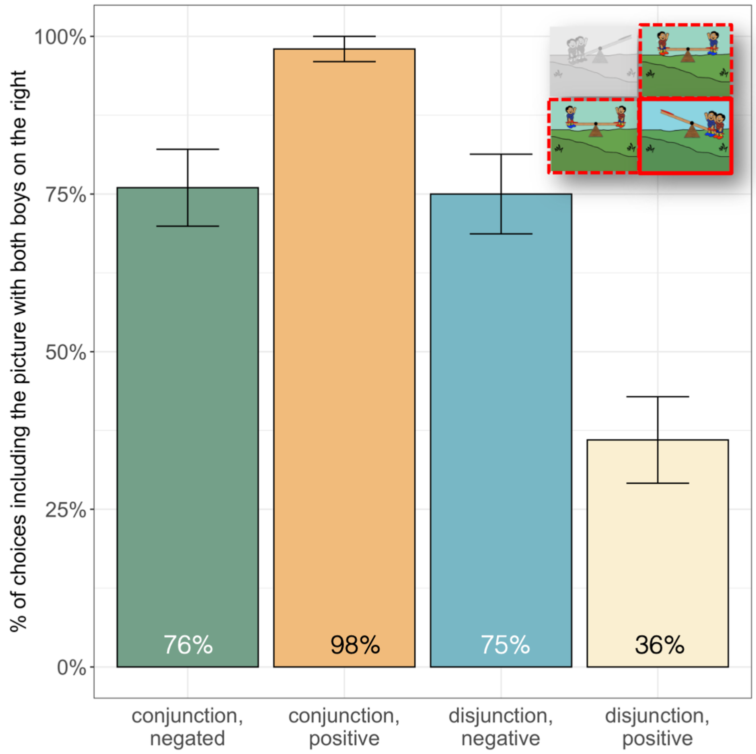


Figure 5 Picture choices in Experiment I.

judgments and in picture choices. In other words, under this approach, we expect counterfactuals of the form $(\bar{A} \vee \bar{B}) \mathrel{\Box\!\!\rightarrow} C$ (POSITIVE DISJUNCTION) and those of the form $(\neg A \vee \neg B) \mathrel{\Box\!\!\rightarrow} C$ (NEGATIVE DISJUNCTION) to behave similarly to each other and both differently from the corresponding one of the form $\neg(A \wedge B) \mathrel{\Box\!\!\rightarrow} C$ (NEGATED CONJUNCTION). Note that this is true only under the intended parse where negation scopes below disjunction, a potential issue we come back to in Experiment III below.

Conversely, under the Alternative Negation approach, negation introduces alternatives, which can have an effect on the overall interpretation of the sentence. Therefore, this approach is compatible with the two conditions involving overt negation (i.e. the NEGATIVE DISJUNCTION and the NEGATED CONJUNCTION conditions) clustering together, in truth-value judgments and in picture choices. (Of course, whether we get a *prediction* that they cluster together depends on how alternatives are generated; see below.)

The results from both tasks of Exp. I reveal an effect of overt negation. In the truth-value judgment task, the negated conjunction condition was endorsed less than the positive disjunction one, like in Ciardelli et al.'s (2018b) results. In addition, crucially, we observed the same drop in endorsement rate for the negative disjunction condition as well. Similarly, the pictures where both boys are on the right were selected more often in those same two conditions than in the positive disjunction one. These results are therefore challenging for the Classical Negation hypothesis and in line with the predictions of the Alternative Negation hypothesis.

In Experiment II, we move to consider counterfactuals with simple antecedents, without binary connectives, and test the predictions of the two approaches in those cases. In Experiment III, we go back to the sentences of Experiment I in order to control for the scope of negation more systematically.

3.4 Experiment II

3.4.1 Methods

Participants. 200 self-declared native speakers of English participated in this experiment, like in Exp. I. Participants were randomly assigned to one of the five conditions below. One worker was excluded from the analysis because of a coding error.

Procedure and Materials. The procedure and tasks were the same as in Experiment I, this time comparing the NEGATED CONJUNCTION to counterfactuals with simple antecedents, either involving negation or not, all summarised in (26) and (31), repeated from above.

(22) If Arthur was on the right, the see-saw would be balanced.
SIMPLE POSITIVE $\bar{A} \mathrel{\Box\!\!\rightarrow} C$

(26) If Arthur and Bill were not both on the left, the see-saw would be balanced.
NEGATED CONJUNCTION $\neg(A \wedge B) \mathrel{\Box\!\!\rightarrow} C$

(29) If Bill was on the right, the see-saw would be balanced.
SIMPLE POSITIVE $\bar{B} \mathrel{\Box\!\!\rightarrow} C$

(30) If Arthur was not on the left, the see-saw would be balanced.
SIMPLE NEGATIVE $\neg A \mathrel{\Box\!\!\rightarrow} C$

(31) If Bill was not on the left, the see-saw would be balanced.
SIMPLE NEGATIVE $\neg B \mathrel{\Box\!\!\rightarrow} C$

Experiment I and its replications (see OSF repository) showed clearly that participants understand the task very well, so we did not include a sanity check in this study.

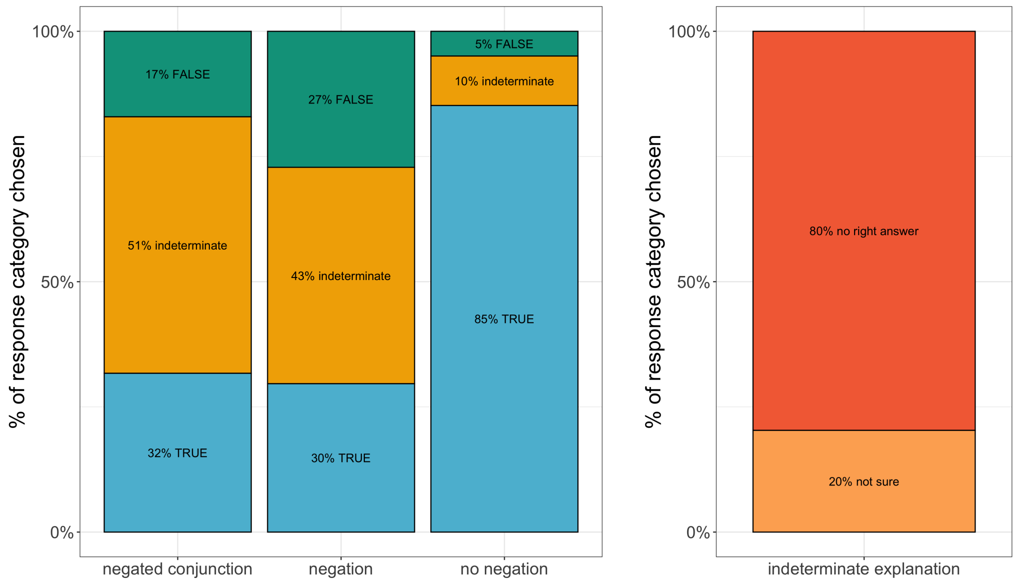


Figure 6 Results of the truth-value judgment task in Experiment II (left), and a detailed look at the Indeterminate choices in Experiment II (right).

3.4.2 Results

For the analyses, the SIMPLE NEGATIVE about Arthur in (30) and the one about Bill in (31) were collapsed into one condition, and so were the SIMPLE POSITIVE ones in (22) and (29).

The results of the truth-value judgment tasks are plotted in Fig. 6. As can be seen, simple antecedents without negation were judged True more often than sentences containing negation, whose response patterns were similar. Table 3.4.2 shows that the statistical results confirm this impression: Responses to simple POSITIVE antecedents differed significantly from those to both simple NEGATIVE antecedents and NEGATED CONJUNCTIONS. The difference between simple NEGATIVE antecedents and NEGATED CONJUNCTIONS was also significant.

As in the previous experiment, the vast majority (80%) of participants who chose the indeterminate answer said they selected it because they felt that there genuinely was no right answer (rather than them being uncertain about it; Fig. 6 on the right). An intercept-only binomial regression confirmed the difference between the two choices was statistically significant (β=1.37, z=4.4, p<.0001).

The results of the picture-selection task are shown in Fig. 3.4.2. After reading the negated conjunction condition, about three-quarters of participants selected

omnibus multinomial analysis

		(intercept)	disjunction, negated	conjunction, negated
Coefficients:	FALSE	-1.099	0.634	0.406
	TRUE	-0.480	0.102	2.634
Std. Errors:	FALSE	0.437	0.514	0.752
	TRUE	0.353	0.441	0.514

pairwise comparisons

contrast	LR statistics	*p-value*
negation | conjunction, negated	64.98	<0.0001***
no negation | conjunction, negated	35.48	<0.0001***
no negation | negation	54.73	<0.0001***

Table 4 Results of the statistical analyses conducted for Experiment II: Parameters of the omnibus model above; likelihood-ratio (LR) statistics and p-values below. Bonferroni-corrected significance is indicated by *.

pictures including both boys on the right; after reading simple negative antecedents, about half the participants made these same choices, and when there was no negation, those pictures were selected by only 11% of the participants in that condition.

The picture choices (Fig. 3.4.2) were analyzed using a binomial model with NEGATED CONJUNCTION as reference level; both β-Estimates comparing NEGATED CONJUNCTION to both SIMPLE NEGATION and NO NEGATION were > .23, and the t-values, > 2.93. Nested-models comparisons confirmed that all differences were significant (all ps < .0001).

3.4.3 Discussion

Recall that the Classical Negation hypothesis predicts that no change should result from swapping out *on the right* with *not on the left*. Conversely, the Alternative Negation hypothesis is compatible with observing differences between the two simple antecedents cases.

In our results, we found again an effect of overt negation across the two tasks: the negated simple antecedents were judged to be true much less than the corresponding positive ones, and the picture with both boys on the right was selected more often in the negated case than in the positive case (though not as much as in the negated conjunction condition). The results are therefore more in line with the

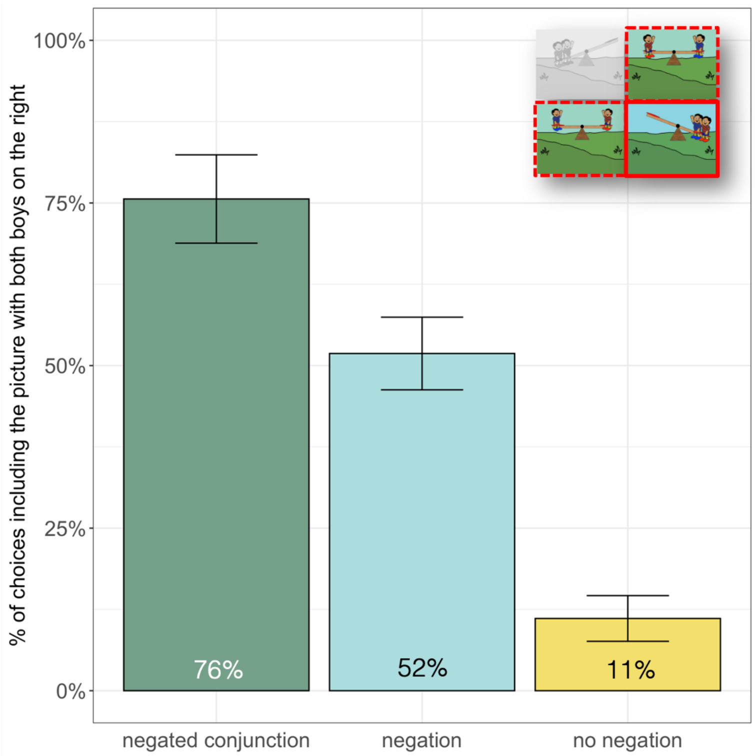


Figure 7 Picture choices in Experiment II. Error bars indicate Standard Errors.

predictions of the Alternative negation approach, and again challenging for the Classical negation one, based on the role of negation in counterfactuals.

Now, one factor to better understand next is the scope of negation with respect to disjunction. In Exp. I, we assumed that participants read (32) as involving a disjunction scoping over negation:

(32) If Arthur or Bill were not on the left, the see-saw would be balanced.

That is, we assumed that (32) is underlying of the form $(\neg A \lor \neg B) \mathrel{\Box\!\!\rightarrow} C$. But we cannot exclude that some of our participants interpreted the negation in (32) as scoping over the disjunction, taking the sentence as having the form $\neg(A \lor B) \mathrel{\Box\!\!\rightarrow} C$. The latter is equivalent to $(\neg A \land \neg B) \mathrel{\Box\!\!\rightarrow} C$, and therefore clearly false in our scenario. Exp. III aims at controlling the scope of negation systematically. In addition to sentences like (32), we also include sentences where negation appears unambiguously below disjunction.

If we observe a similar difference, both in truth-value judgments and in picture selection choices, between (32) and its positive counterpart in (33), then we can be confident that in those cases scope is not playing a role. This in turn means that the points made above in relation to the theoretical approaches could be translated to the clausal cases, to refine the challenge for the Classical negation approach.

3.5 Experiment III

3.5.1 Methods

Participants. 200 self-declared native speakers of English participated in this experiment, like in Exp. I and II. Participants were randomly assigned to one of the four conditions below. Three workers were excluded from the analysis because of a coding error.

Materials. The procedure and tasks were the same as in Exp.s I and II, using the story in Fig. 3. The conditions again involved NEGATED CONJUNCTION and NEGATIVE DISJUNCTION, such as in (26) and (28), repeated below. In addition, we added two novel clausal conditions in one of which negation overtly appears in each disjunct, thereby fixing its scope. (33) is a clausal version of the positive disjunction case; (23) is its negative counterpart and the crucial novel condition.

(26) If Arthur and Bill were not both on the left, the see-saw would be balanced.
NEGATED CONJUNCTION $\neg(A \wedge B) \mathrel{\Box\!\!\rightarrow} C$

(28) If Arthur or Bill were not on the left, the see-saw would be balanced.
NEGATIVE DISJUNCTION $(\neg A \vee \neg B) \mathrel{\Box\!\!\rightarrow} C$

(33) If Arthur was on the right or Bill was on the right, the see-saw would be balanced.
CLAUSAL DISJUNCTION $(\bar{A} \vee \bar{B}) \mathrel{\Box\!\!\rightarrow} C$

(23) If Arthur was not on the left or Bill was not on the left, the see-saw would be balanced.
CLAUSAL NEGATIVE DISJUNCTION $(\neg A \vee \neg B) \mathrel{\Box\!\!\rightarrow} C$

3.5.2 Results

Figure 8 (left) shows the pattern of results. We fitted a multinomial regression with condition as fixed effect across the four conditions with POSITIVE CLAUSAL DISJUNCTION and INDETERMINATE answers as reference level.[12] Table 3.5.2 (upper half) shows the coefficients of this model; planned pairwise comparisons (lower half)

12 Anonymized data and code are available on https://osf.io/2etx7/.

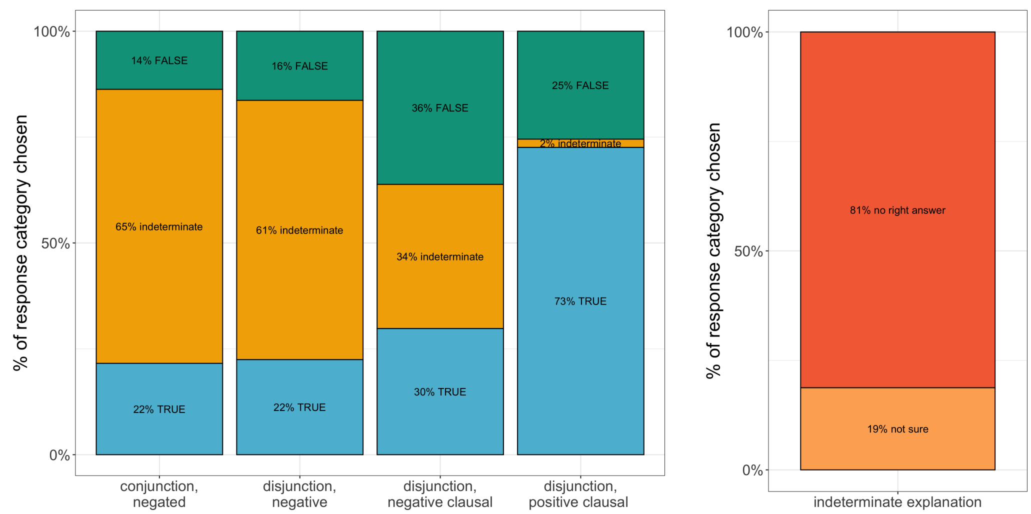


Figure 8 Results of the truth-value judgment task in Experiment III (left), and a detailed look at the Indeterminate choices in Experiment III (right).

revealed that CLAUSAL NEGATIVE DISJUNCTION and CLAUSAL POSITIVE DISJUNCTION resulted in significantly different truth-value judgment patterns. In addition, the comparison between NEGATED CONJUNCTION to NEGATIVE DISJUNCTION, as well as the one between CLAUSAL NEGATIVE DISJUNCTION and NON-CLAUSAL NEGATIVE DISJUNCTION were marginally significant.

As in the previous experiment, the vast majority (81%) of participants who chose the indeterminate answer said they selected it because they felt that there genuinely was no right answer (rather than them being uncertain about it; Fig. 8 on the right). An intercept-only binomial regression confirmed the difference between the two choices was statistically significant (β=1.47, z=5.12, p<.0001).

Again, the picture choices (Fig. 3.4.2) were analyzed using a binomial model with POSITIVE CLAUSAL DISJUNCTION as reference level; all parameters showed high β-estimates (> 1.73) and z-values (> 3.89). Nested-models comparisons confirmed that crucially, NEGATED CLAUSAL DISJUNCTION led to significantly more picture choices including both boys on the right than POSITIVE CLAUSAL DISJUNCTION (p<.00001). None of the other comparisons were significant (ps>.16).

3.5.3 Discussion

In Exp. I, we found a difference between positive and negative disjunctions, which we argued challenged the Classical negation approach. But we could not exclude,

omnibus multinomial analysis

		(intercept)	conjunction, negated	disjunction, negative	disjunction, negative clausal
Coefficients:	FALSE	13.000	0.016	0.021	0.082
	TRUE	37.001	0.009	0.010	0.024
Std. Errors:	FALSE	1.038	1.118	1.111	1.095
	TRUE	1.013	1.072	1.073	1.077

pairwise comparisons

contrast	*LR statistics*	*p-value*
disjunction, clausal | disjunction, negated clausal	27.09	<0.0001***
conjunction, negated | disjunction, negative	12.44	0.058
conjunction, negated | disjunction, negative clausal	10.52	0.021*
disjunction, negative | disjunction, negative clausal	7.96	0.075

Table 5 Results of the statistical analyses conducted for Experiment III: Parameters of the omnibus model above; likelihood-ratio (LR) statistics and *p*-values below. Bonferroni-corrected significance is indicated by *.

that the difference was at least partly due to some participants reading the negation as taking wide scope, instead of the intended narrow scope reading.

Exp. III addressed this potential confound. We controlled for the scope of negation by using clausal versions of the negative and positive disjunction conditions of Exp. I. The results still show a clear difference between clausal positive and negative conditions, which cannot be attributed to scope, thereby strengthening the challenge for the Classical negation approach.

We turn now to Exp. IV, where we investigate all of the critical sentences in the antecedents of the counterfactuals tested in Exp.s I-III, but as unembedded sentences. The goal here is to test the potential overall effect of negation, independently from the presence of a counterfactual. To this end, we also compare directly the clausal positive and negative disjunction condition of Exp. III and the corresponding unembedded positive and negative disjunction condition of Exp. IV.

3.6 Experiment IV

3.6.1 Methods

Participants. 300 self-declared native speakers of English participated in this experiment, using the same criteria as in the other experiments. The sample size was increased in order to have a similar number of participants as in the previous

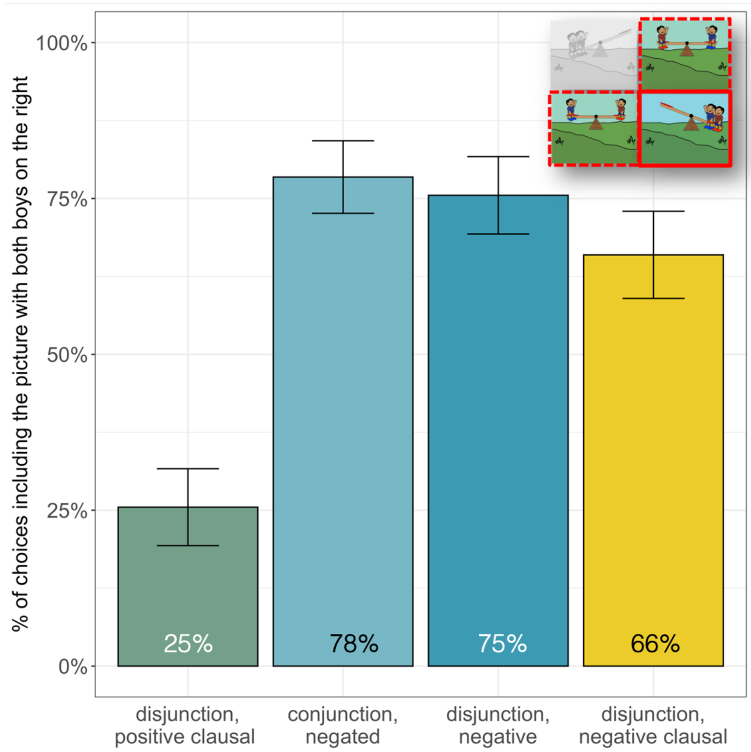


Figure 9 Picture choices in Experiment III.

experiments, distributed over the six conditions below. One worker was excluded from the analysis because of a coding error.

Materials. The procedure and task were just like in the previous experiments, but since the sentences were presented in isolation, the task was a simple picture-sentence matching task, to one of each conditions presented so far (see Fig. 10):

(34) Arthur or Bill are on the right.
POSITIVE DISJUNCTION $(\bar{A} \vee \bar{B})$

(35) Arthur and Bill are not both on the left.
NEGATED CONJUNCTION $\neg(A \wedge B)$

(36) Arthur and Bill are both on the right.
CONJUNCTION $(A \wedge B)$

(28) Arthur or Bill are not on the left.
NEGATIVE DISJUNCTION $(\neg A \vee \neg B)$

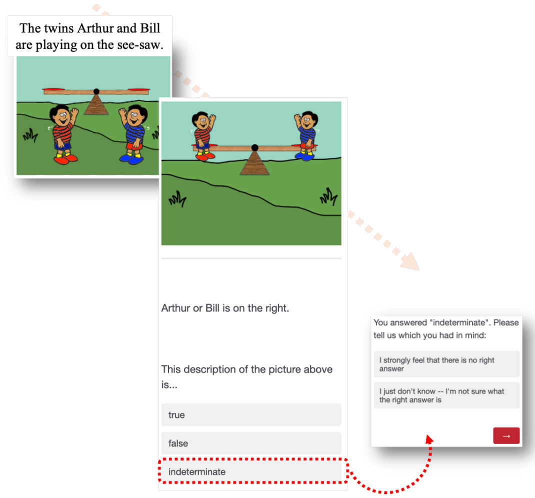


Figure 10 Picture-matching task used in Experiment IV.

(37) Arthur is on the right or Bill is on the right.
CLAUSAL DISJUNCTION $(\bar{A} \vee \bar{B})$

(38) Arthur is not on the left or Bill is not on the left.
CLAUSAL NEGATIVE DISJUNCTION $(\neg A \vee \neg B)$

3.6.2 Results

Fig. 11 shows the picture-matching judgments for each condition. The contrast between NEGATED CONJUNCTION and POSITIVE CONJUNCTION shows that people understood the task well (in fact, answers were 100% "true" and 100% "false", respectively).

For the analysis of Exp. IV, we again conducted a multinomial regression with condition as fixed effect, with indeterminate choices and POSITIVE DISJUNCTION as reference level; the model parameters are reported in the top half of Table 3.6.2.[13] Our planned pairwise comparisons aimed to understand the effect of negation and connective type independently of counterfactuality (top half of Table 3.6.2). These comparisons showed that negation significantly lowered the proportion of "true" judgments for the simple disjunction, but in the case of clausal disjunction there

13 Note that instead of the planned four conditions in the pre-registration, we decided, for completeness' sake, to include all six conditions critical to the paper so far. Since it was not theoretically important to analyze all 15 possible combinations, we focused on the most relevant for our purposes. As for all other experiments, the anonymized data and code are available on https://osf.io/2etx7/.

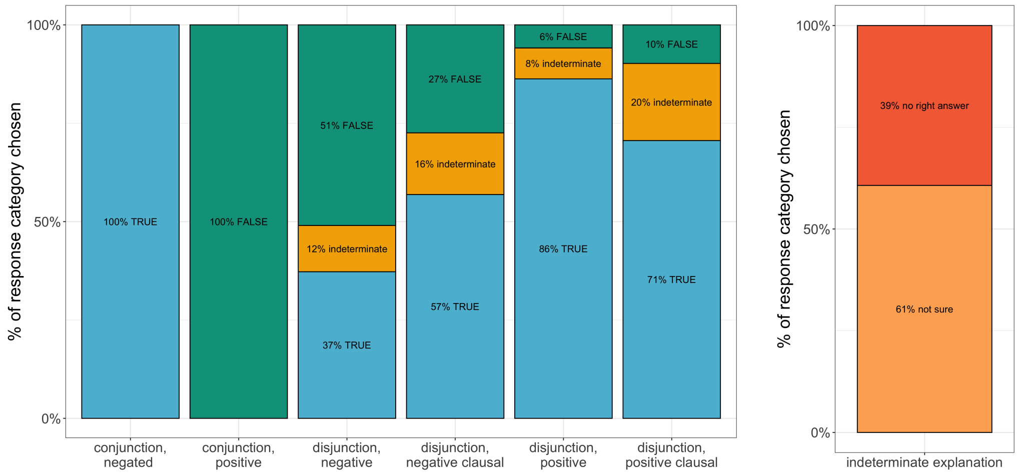


Figure 11 Results of the picture-matching task in Experiment IV (left), and a detailed look at the Indeterminate choices in Experiment IV (right).

was only a numerical trend which wasn't significant.

To understand whether negation and clausality interacted in this task, we performed an additional analysis. The data were subsetted to only disjunctive sentences, and negation and clausality contrast-coded. We then built a multinomial model predicting matching choices by negation, clausality, and their interaction. None of the predictors were significant except, again, negation (z=2.65, $p < .008$, all others: zs<1.16, ps> .16); in addition, the interaction was not (zs<1.41, ps> .15). Finally, the analysis of indeterminate choices showed, in this case, no significant difference between "unsure" answers and truly indeterminate answers (β=0.44, z=1.13, p=0.26).

3.6.3 Discussion

Exp. IV had the goal of testing whether adding negation would have an effect on judgments independently from counterfactuals. When we focus on the disjunctive conditions, we do observe a drop in "true" judgments in the negative conditions with respect to their positive counterparts.

As discussed above, in the non-clausal case, we cannot exclude that the difference between positive and negative is (at least in part) due to wide-scope reading of negation in the negative case. The case of the clausal negative disjunction and the clausal positive case, on the other hand, does not involve scope and therefore more directly gives us a baseline for the effect of negation on judgments. In this

omnibus multinomial analysis

		(intercept)	conjunction, negative	conjunction, positive
Coefficients:	FALSE	-0.287	-0.179	12.019
	TRUE	2.420	11.849	-4.438
Std. Errors:	FALSE	0.764	280	48.5
	TRUE	0.522	174	141.4

		disjunction, negative	disjunction, negative clausal	disjunction, positive clausal
Coefficients:	FALSE	1.754	0.847	-0.406
	TRUE	-1.267	-1.133	-1.139
Std. Errors:	FALSE	0.888	0.883	0.940
	TRUE	0.701	0.657	0.632

pairwise comparisons

contrast	*LR statistics*	*p-value*
disjunction, positive | disjunction, negative	41.38	<0.001**
disjunction, positive clausal | disjunction, negative clausal	5.42	0.13

Table 6 Results of the statistical analyses conducted for Experiment IV: Parameters of the omnibus model above; likelihood-ratio (LR) statistics and *p*-values below. Bonferroni-corrected significance is indicated by *.

case, we observed a drop in judgments in the negative case (which was marginally significant before Bonferroni-correction and not significant after). This difference, if reliable, could be due to additional complexity added by the negation or extra pragmatic conditions imposed by its presence. What is relevant for us here is whether this effect could account for the contrasts between positive and negative counterfactual sentences that we observed in Exp.s I-III.

In order to address this question, we compare more directly the results for clausal positive and negative unembedded disjunctions of Exp. IV with the corresponding clausal positive and negative disjunctions in the antecedent of counterfactuals of Exp. III. In particular, we target an interaction between negation and matching type, which would suggest that the mere presence of negation cannot alone explain our results.

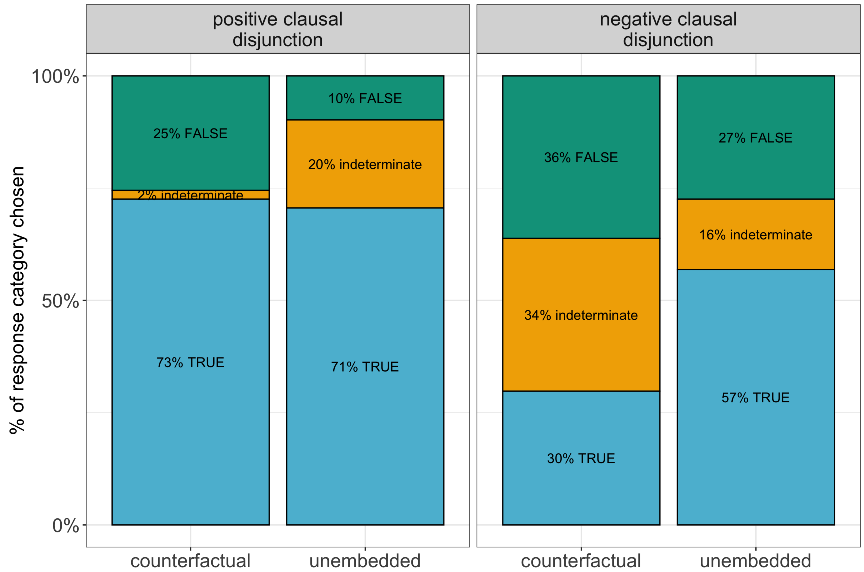


Figure 12 Comparing results of Exp. III and Exp. IV.

3.7 Comparison between Exp. III and Exp. IV

To compare between experiments, we constructed another multinomial model, making indeterminate the reference level. Since the design allowed it, we contrast-coded both the level of embedding (counterfactual vs. unembedded) and the condition (NEGATIVE VS. POSITIVE CLAUSAL DISJUNCTION), and predicted truth-value judgment from presence or absence of negation, level of embedding, and their interaction.

Table 3.7 shows the results, always relative to the "indeterminate" answers. Focusing on the True responses, we find a main effect of negation, but no effect of embedding. Crucially, we observe a significant interaction between negation and embedding: the difference between positive and negative was larger in the counterfactual case vs the unembedded one.

The interaction is critical for us because it confirms that a general effect of the presence of negation cannot account for the difference between positive and negative in the counterfactual case. Rather, it suggests that it is negation *in combination with* counterfactuals that is responsible for (at least part) of the differences that we observed in Exp. 1-3 between the positive and negative conditions.

omnibus multinomial analysis

		(intercept)	negation	embedding	interaction negation X embedding
Coefficients:	FALSE	0.623	0.313	0.690	0.939
	TRUE	1.512	0.934	0.227	0.938
Std. Errors:	FALSE	0.325	0.325	0.325	0.325
	TRUE	0.301	0.301	0.301	0.301
p-values:	FALSE	0.06	0.336	0.034*	<0.01*
	TRUE	<0.0001***	<0.002**	0.450	<0.02*

Table 7 Results of the statistical comparison between Exp. III and Exp. IV.

3.8 Summary of the results

We tested counterfactuals with complex antecedents in a simple scenario across a series of four experiments. Participants were asked to perform a truth-value judgment task, and a picture selection task. Overall, we consistently found an effect of overt negation.

The first finding (Exp. I) concerns counterfactuals which include overt negations in disjunctive antecedents, i.e. counterfactuals of the form in (39)-a. In the truth-value judgment task, these counterfactuals are assessed as low as counterfactuals with negated conjunctions in antecedents, like (39)-b, and differently from their positive counterpart in (39)-c. In line with this, participants selected picture choices including both boys or the right more in the negative cases in (39-a) and (39-b) than in the positive one in (39-c). Crucially, we replicated these results in Exp. III, where the scope of negation with respect to disjunction was more explicitly controlled for.

(39) a. $(\neg A \vee \neg B) \,\square\!\!\rightarrow C$
b. $\neg(A \wedge B) \,\square\!\!\rightarrow C$
c. $(\bar{A} \vee \bar{B}) \,\square\!\!\rightarrow C$

The second finding concerns counterfactuals with truth-conditionally equivalent antecedents that, respectively, do not and do include overt negation (Exp. II). We represent them schematically in (40).

(40) a. $\bar{A} \,\square\!\!\rightarrow C$
b. $\neg A \,\square\!\!\rightarrow C$

We find that these counterfactuals are *not* judged to be equivalent, and trigger different responses in both tasks. In particular, (40)-b was associated with a lower truth-value judgment score than (40-a) and with more picture choices involving both boys on the right. Finally, in Exp. IV – and by comparing Exp. IV and Exp. III – we find that the relevant effects are associated with the presence of overt negation specifically in combination with counterfactuals.

Overall, these results are in line with the **Alternative Negation** approach, while they raise a challenge for the the **Classical Negation** approach. We now turn to discuss the theoretical landscape in detail.

4 Positive account

In this section, we propose a semantics that explains both the original data in CZC and our new data in §3. To do so, we use an inquisitive semantics that builds on standard inquisitive frameworks (see Ciardelli, Groenendijk & Roelofsen 2018a, as well as Ciardelli 2016), but develops a novel treatment for negation. The account has some obvious similarities to the account in Schulz 2019, though there is a crucial difference in the treatment of negation and its alternatives.

We choose an inquisitive system because it makes the comparison with other accounts easier, since both CZC and Schulz operate within an inquisitive framework. But we want to emphasize that this was not the only choice. For example, frameworks based on alternative semantics (like Alonso-Ovalle 2009), or frameworks based on syntactic alternatives (like Santorio 2018) might have also been deployed. The only key assumption concerning counterfactuals is that the operation that maps a counterfactual antecedent to alternatives should be semantically encoded, rather than scalar. (We discuss why we do not adopt a scalar system in §5.)

4.1 Formal semantics

Inquisitive semantics (see Ciardelli, Groenendijk & Roelofsen 2018a, among others) is an attempt to provide a unitary semantics for declarative and interrogative sentences. In this framework, the meaning of a sentence is not equated directly to its truth-conditions, understood as the set of worlds in which the sentence is true. Rather, it is modeled in terms of the information states supporting that sentence, where information states are sets of sets of worlds.

What matters for us is that the basic notion of the semantics is a notion of support of a sentence at an information state. This opens up the possibility of giving a nonclassical treatment of connectives. We are going to define a fairly standard inquisitive semantics, following also CZC up to a point. Crucially, though, we define a different semantics for negation.

Language. We use a propositional language where sentences are formed recursively from atomic sentences, connectives, and a counterfactual operator.

$$\mathsf{A} \quad ::= \quad p \mid \bot \mid \mathsf{A} \wedge \mathsf{A} \mid \mathsf{A} \vee \mathsf{A} \mid \neg \mathsf{A} \mid \mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{A}$$

Model-theoretic background. We assume a background model consisting of a triple $\langle \mathcal{W}, V, \preceq_w \rangle$. $\mathcal{W}$ is a set of possible worlds; V is a valuation function mapping a pair of an atomic sentence and a world to a member of $\{0,1\}$; $\preceq_w$ is a three-place comparative similarity relation among worlds, in the style of Lewis 1973b. $\preceq_w (w_1, w_2)$ holds iff w_1 is more similar to w than w_2 is.

In addition, following Ciardelli 2016, we define a **counterfactual operation** on propositions $\Rightarrow$. $p \Rightarrow q$ maps propositions p and q to a third, counterfactual proposition. Our account preserves standard minimal change semantics for the conditional. So $\Rightarrow$, following on standard counterfactual semantics in Lewis and Kratzer, is defined in terms of a relation of comparative closeness, $\preceq_w$, as follows:

(41) $\quad p \Rightarrow q \quad = \quad \{w : \forall w' : w' \in \textsc{max}_{\preceq, w}(p), w' \in q\}$

I.e., $p \Rightarrow q$ is the proposition that includes all the worlds w such that all the maximally close p-worlds to w are q-worlds.

Semantics. The key notion of the semantics is the notion of a sentence being **supported at a state**. A **state** is modeled simply as a set of worlds, i.e. a subset of $\mathcal{W}$. Thus our semantics departs from classical accounts, on which meaning is characterized via a recursive assignment of truth conditions. But this departure is not specific to inquisitive systems. Rather, it is shared with several frameworks that are broadly in the dynamic tradition (see e.g. Veltman 1985, Veltman 1996). The distinctive feature of inquisitive semantics is integrating the notion of support with a notion of an **alternative**. An alternative of sentence ϕ is a state s such that (i) s supports ϕ, and (ii) there is no state s' that is a superset of s and that also supports ϕ. Informally, alternatives are the largest, and hence least informative, states that support a sentence.

In traditional systems, declarative sentences are invariably associated with just one alternative i.e., there is invariably a single largest state that makes a sentence true. But this is not true in inquisitive semantics. In particular, disjunctions are associated to two alternatives, i.e. the alternatives denoted by each disjunct. Using lowercase italics to denote propositions:

$$Alt(\mathsf{A} \vee \mathsf{B}) = \{a, b\}$$

This feature is at the basis of the nonclassical behavior of disjunction in inquisitive systems. In particular, sentences that are taken to be classically equivalent to disjunctions might have different alternatives and this is what explains the data about counterfactual antecedents.

Below are the clauses for atomic sentences and for conjunction and disjunction; negation will be discussed in the next section.

$$\begin{array}{lll} s \vDash p & \text{iff} & \text{for all } w \in s, V(p,w) = 1 \\ s \vDash \mathsf{A} \wedge \mathsf{B} & \text{iff} & s \vDash \mathsf{A} \text{ and } s \vDash \mathsf{B} \\ s \vDash \mathsf{A} \vee \mathsf{B} & \text{iff} & s \vDash \mathsf{A} \text{ or } s \vDash \mathsf{B} \end{array}$$

The entry for counterfactuals (which is based on the entry in Ciardelli 2016) exploits the definition of alternatives, as well the '$\Rightarrow$' operation defined above.

(42) $s \vDash \mathsf{A} \;\Box\!\!\rightarrow \mathsf{B}$ iff $\forall p \in Alt(\mathsf{A})\; \exists q \in Alt(\mathsf{B})$ such that $s \subseteq p \Rightarrow q$

I.e., a counterfactual $\mathsf{A} \;\Box\!\!\rightarrow \mathsf{B}$ is supported at a state just in case s is included in the proposition $p \Rightarrow q$, for all p that are alternatives of A, and for some q that is an alternative of B. (We are going to sketch a refinement of the semantic clause for counterfactuals to accommodate undefinedness in §4.6.)

Notice that, in combination with the inquisitive meaning for disjunction, the meaning in (42) immediately vindicates Simplification, repeated below.

(6) **Simplification**: $(\mathsf{A} \vee \mathsf{B}) \;\Box\!\!\rightarrow \mathsf{C} \vDash \mathsf{A} \;\Box\!\!\rightarrow \mathsf{C}, \mathsf{B} \;\Box\!\!\rightarrow \mathsf{C}$

Of course, given the discussion in Section 1, this also means that our semantics has to have some nonclassical features. Since Simplification is valid, and since we retain the classical comparative closeness analysis of counterfactual meaning (and hence Failure of Antecedent Strengthening), Substitution has to fail. We address this point below.

Up to here, everything is in agreement with standard inquisitive semantics for connectives and conditionals. Let us now introduce our point of departure.

4.2 Negation

We treat negation as alternative-sensitive: $\neg$A has multiple alternatives, similarly to disjunction and differently from conjunction. Crucially, and differently from other accounts, these alternatives are partly determined via a contextual parameter.

Building on much work in formal pragmatics, we assume a notion of Question Under Discussion, or QUD (see Stalnaker 1978, 2002, Roberts 2012 a.o.). The driving idea behind the notion of a QUD is the following. At any stage in discourse, speakers are trying to address one or more questions that have been explicitly or implicitly raised. We can think of these questions as being built into the conversational score, i.e. the overall body of information that speakers are tracking in conversation.

Following standard accounts, we can model a QUD as a partition on the worlds that are live options in conversation (the *context set*, in Stalnaker's terminology).

Against this background, our account of negation is the following: (i) negation is sensitive to the QUD; (ii) ¬A rules out all the A-compatible answers to the QUD; i.e., ¬A is supported at a state s just in case s entails an answer to the QUD that is incompatible with A. As we point out below, the idea that negation is alternative-sensitive is often invoked in the literature.

To formalize this, we need to introduce a further notion, i.e. the notion of a state supporting a question. Following standard inquisitive systems (see e.g. Ciardelli et al. 2018a), we say that a state s supports a question Q just in case s entails a full answer to Q.

$$s \vDash Q \quad \text{iff} \quad \text{for some } p \in Q, s \vDash p$$

Our semantics for negation is the following ('|A|' stands for the possible worlds content of A):[14]

(43) $\quad s \vDash \neg_Q \mathsf{A} \quad \text{iff} \quad \text{for all } t \subseteq s, t \vDash Q \text{ and } t \cap |\mathsf{A}| = \varnothing$

n words, a state s supports $\neg_Q$A just in case every substate of s entails an answer to Q, and every substate of s is incompatible with A. Notably, we assume that negation has an extra argument for a QUD, and that that argument is filled in by context.

Let us go through an example. Consider a conversation in which there are four live possibilities, in which each of the two twins Arthur and Bill is on one of the sides of the see-saw (visualized in Figure 13): Suppose first that the QUD concerns the location of Arthur (and not that of Bill). In this situation, the QUD partitions the context into two subsets of worlds, each involving two possibilities.

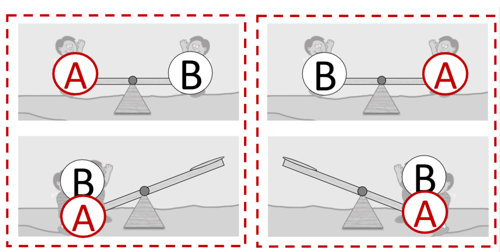


Figure 13 QUD of a conversation in which only Arthur's location matters.

14 The possible worlds content |A| of a sentence A is defined as the union of the alternatives of A: $|\mathsf{A}| = \bigcup Alt(\mathsf{A})$.

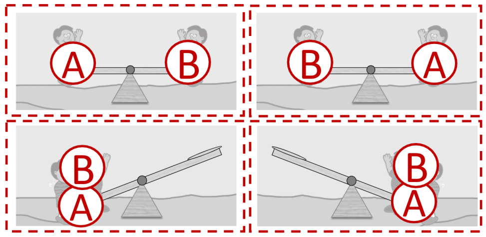
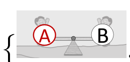


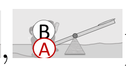
Figure 14 QUD of a conversation in which Arthur's and Bill's locations matter.

Against this background, consider an assertion of:

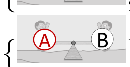
(44) Arthur is not$_Q$ on the right.

According to the entry in (43), (44) is supported by any state s such that (i) all subsets of s provide a complete answer to the QUD, and (ii) no subsets of s support *Arthur is on the right.* Assuming the obvious semantics for *Arthur is on the right*, this means that (44) is supported by the following three states:

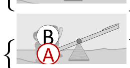
(45) a. { 

, 

}

b. { 

}

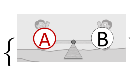
c. { 

}

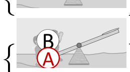
Notice that, since (45-b) and (45-c) are subsets of (45-a), (44) has just one alternative, namely (45-a) itself.

Consider now a case where the QUD is more fine-grained: suppose that it concerns the location of both Arthur and Bill. In this case, the QUD partitions our sample context into four singletons (see Figure 14).

In this case, we find that (44) is supported by the following two states:

(46) a. { 

}

b. { 

}

Both of them are alternatives for (44).

Of course, by yoking the denotation of negation to the QUD, our account introduces an element of context dependence. We won't give a full theory of how QUDs are determined in a given context. But we need one specific assumption, which we take to be independently plausible. We assume that contexts like the one in our experiments (or CZC and Schulz ones), raise a QUD that is fine-grained enough to settle both the truth of the proposition regarding the single individuals (e.g. be it Arthur and Bill or switch A and switch B).[15]

Before moving on, let us emphasize two ways in which our account connects to existing work.

First, several theories of focus and alternatives point to a connection between negation, alternatives, and the QUD. For example, Tian, Ferguson and Breheny (2016; see also Tian & Breheny 2019) argue that the processing of negation requires consideration of an explicit or implicit QUD. Negation is also known to interact in substantial ways with focus alternatives, which in turn are linked to the QUD (see Beaver & Clark 2008 among others). Giving a general theory of how negation and QUD interact goes beyond the purposes of this paper. But the idea that there should be a link between negation and QUD connects to existing work on negation and alternatives.

Second, our account is similar to Schulz's in important respects. In particular, Schulz also treats negation as alternative-generating. The key difference is that, on her account, alternatives are determined exclusively by the lexical material that is in the scope of negation. Conversely, on our account alternatives are generated partly via the QUD, hence contextually. As we will see in §5, this makes for an important difference in predictions.

4.3 Predictions: CZC and our data

Let us now explain how our approach fares with the original data presented by CZC, as well as for our novel data in Exp.s I-IV. First of all, we notice that we can account for the basic contrast between disjunction and negated conjunction in counterfactual antecedents. Consider first counterfactuals with disjunctive antecedents. Via the semantics in (42), the support conditions we get are below (we mark the proposition denoted by a sentence with the corresponding boldface capital letters).

(47) $s \vDash (\mathsf{A} \vee \mathsf{B}) \mathrel{\Box\!\!\rightarrow} \mathsf{C}$ iff $\forall p \in \{\mathbf{A}, \mathbf{B}\} : \exists q \in Alt(\mathsf{C}) : s \subseteq p \Rightarrow q$

15 This is loosely inspired by the structural conception of alternatives, due to Katzir 2007, Fox & Katzir 2011 that has become widely accepted in the literature on implicature. Katzir defines a syntactic notion of alternative. For Katzir, alternatives to a clause A are syntactic strings that are no more complex than A. In turn, and roughly, alternatives that are no more complex than A are strings that may be obtained from A via deletion and replacement of subconstituents, using a relevant fragment of the lexicon as substitution source. One key constraint that Katzir adopts is that material that has been recently pronounced is in the substitution source, and counts as no more complex than the sentence itself.

To take a concrete case, consider (48):

(48) If Arthur or Bill was on the right, the see-saw would be balanced.

On the assumption that the alternatives of *Arthur or Bill was on the right* are the propositions expressed by the two disjuncts, and that the consequent denotes only one alternative, the predicted support conditions are below (we use again boldface to denote propositions).

(49) $s \vDash$ (48) iff $\forall p \in \{\textbf{A right}, \textbf{B right}\}$,
$s \subseteq \{w : \forall w' \in \textsc{max}_{w,\leq}(p), w' \in \textbf{Balanced}\}$

I.e., (48) is supported at a state s just in case the closest worlds to s-worlds where Arthur is on the right are worlds where the see-saw is balanced, and the closest worlds where Bill is on the right are worlds where the see-saw is balanced. Given a plausible choice of ordering or premise set, (48) will be supported by the information state of an agent with the relevant information. The key element here is the lexical entry for disjunction, on which $\mathsf{A} \vee \mathsf{B}$ is supported at a state just in case one of the disjuncts is supported. For the case of (48), this entry determines that the alternatives for a disjunction are the propositions expressed by the two disjuncts.

Consider now the counterpart of (48) involving negated conjunction:

(50) If Arthur and Bill were not$_Q$ both on the left, the see-saw would be balanced.

Since the entry for negation is QUD-dependent, we need to fix a QUD. We assume that, in our context, the relevant QUD is the following:

Q: $\{\textbf{A left \& B left}, \textbf{A left} \,\&\, \overline{\textbf{B left}}, \overline{\textbf{A left}} \,\&\, \textbf{B left}, \overline{\textbf{A left}} \,\&\, \overline{\textbf{B left}}\}$

In this case, the alternatives for *Arthur and Bill are not both on the left* are three: $\textbf{A left} \,\&\, \overline{\textbf{B left}}$, $\overline{\textbf{A left}} \,\&\, \textbf{B left}$, $\overline{\textbf{A left}} \,\&\, \overline{\textbf{B left}}$. So we get the following support conditions:

(51) $s \vDash$ (50) iff $\forall p \in \{\textbf{A left} \,\&\, \overline{\textbf{B left}}, \overline{\textbf{A left}} \,\&\, \textbf{B left}, \overline{\textbf{A left}} \,\&\, \overline{\textbf{B left}}\}$,
$s \subseteq \{w : \forall w' \in \textsc{max}_{w,\leq}(p), w' \in \textbf{Balanced}\}$

Of course, the presence of the third alternative makes a difference. Assuming a closeness ordering that reflects the causal structure of the scenario, the closest worlds where neither boy is on the left are worlds where the see-saw is not balanced. So the counterfactual turns out to be not supported in the relevant scenario.

So far, these predictions mimic those of CZC. But we get a crucial divergence in the following two cases. Take (52-a) first.

(52) If Arthur or Bill were not$_Q$ on the left, the see-saw would be balanced.

We assume the same background QUD as for (50):

$$Q\text{: }\{\textbf{A left \& B left}, \textbf{A left} \,\&\, \overline{\textbf{B left}}, \overline{\textbf{A left}} \,\&\, \textbf{B left}, \overline{\textbf{A left}} \,\&\, \overline{\textbf{B left}}\}$$

Then we get, via the meaning of negation:

$$Alt(\text{A is not left}) = \{\overline{\textbf{A left}} \,\&\, \textbf{B left}, \overline{\textbf{A left}} \,\&\, \overline{\textbf{B left}}\}$$
$$Alt(\text{B is not left}) = \{\textbf{A left} \,\&\, \overline{\textbf{B left}}, \overline{\textbf{A left}} \,\&\, \overline{\textbf{B left}}\}$$

Once we take the disjunction of the two negated clauses, we get back the same set of alternatives as the one we obtained for (50).

$$Alt(\text{Ais not left or B is not left}) = \{\overline{\textbf{A left}} \,\&\, \textbf{B left}, \textbf{A left} \,\&\, \overline{\textbf{B left}}, \overline{\textbf{A left}} \,\&\, \overline{\textbf{B left}}\}$$

Hence we predict that (50) and (52) have analogous support conditions. This explains the data we observe in §3.[16]

Take now the case where the antecedent of a counterfactual involves overt negation, but no binary connective.

(53) If Arthur was not$_Q$ on the left, the see-saw would be balanced.

If we keep assuming the same QUD, we get that the antecedent of (53) has two alternatives, i.e. the two in which Arthur is not on the left. As a result, the support conditions of (53) are predicted to be:

(54) $$s \vDash (53) \quad \text{iff} \quad \forall p \in \{\overline{\textbf{A left}} \,\&\, \textbf{B left}, \overline{\textbf{A left}} \,\&\, \overline{\textbf{B left}}\}, \; s \subseteq \{w : \forall w' \in \textsc{max}_{w,\leq}(p), w' \in \textbf{Balanced}\}$$

Also in this case, the semantics predicts that (53) is not supported in the relevant scenario. Again, this is in line with the data in §3.

4.4 Predictions: Schulz's data

Recall that participants in Schulz' experiment were asked to evaluate the target sentence in (18), repeated below, against the background of a version of CZC's scenario where the electricity is off in the whole building.

16 This extends to the clausal counterpart in (i), assuming the two instances of negation are associated with the same QUD.

(i) If Arthur was not$_Q$ on the left or Bill was not$_Q$ on the left, the see-saw would be balanced.

(18) If the electricity was working and switch A and switch B were not both up, then light would (still) be off.

Assume a background QUD where the salient propositions are whether the electricity is on, whether switch A is up, and whether switch B is up. Using '**E**', '**A**', and '**B**' to stand for the relevant propositions, the QUD is:

$$Q\text{: } \{\mathbf{EAB}, \mathbf{\bar{E}AB}, \mathbf{E\bar{A}B}, \mathbf{EA\bar{B}}, \mathbf{\bar{E}\bar{A}B}, \mathbf{E\bar{A}\bar{B}}, \mathbf{\bar{E}A\bar{B}}, \mathbf{\bar{E}\bar{A}\bar{B}}\}$$

Given our entry for negation, the antecedent *the electricity was working and switch A and switch B were not both up* denotes three alternatives, namely: $\mathbf{E\bar{A}B}$, $\mathbf{EA\bar{B}}$, and $\mathbf{E\bar{A}\bar{B}}$. As a result, the support conditions of the conditional are:

(55) $$s \vDash (18) \quad \text{iff} \quad \forall p \in \{\mathbf{E\bar{A}B}, \mathbf{EA\bar{B}}, \mathbf{E\bar{A}\bar{B}}\}, \\ s \subseteq \{w : \forall w' \in \textsc{max}_{w,\leq}(p), w' \in \mathbf{Off}\}$$

This predicts that the conditional is not accepted at the state of the context, since (given plausible assumptions about the ordering, at least) in the closest worlds where $\mathbf{E\bar{A}\bar{B}}$ is true the light is not off. This is in line with Schulz's experimental finding.

4.5 Answering the logical challenges

Finally, let us consider how our account addresses the challenges originally raised by CZC. Recall that CZC challenge two main tenets of standard counterfactual logics:

(56) **Substitution**: $\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \vDash \mathsf{A}' \mathrel{\Box\!\!\rightarrow} \mathsf{C}$ (with A and A′ truth-conditionally equivalent)

(57) **Negated Conjunction**: $\neg\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}, \neg\mathsf{B} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \vDash \neg(\mathsf{A} \wedge \mathsf{B}) \mathrel{\Box\!\!\rightarrow} \mathsf{C}$

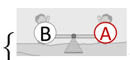
As we pointed out above, all accounts of the phenomenon have to drop Substitution, at least at some level of theorizing. This is not surprising. The data under discussion include instances of the Simplification inference, and as we pointed out in §1, vindicating Simplification requires rejecting Substitution at some level (provided we don't want to endorse a semantics on which Antecedent Strengthening holds).

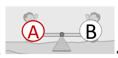
Most importantly, our account validates Negated Conjunction.[17] The apparent

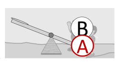
17 We are here assuming that the formulation of Negated Conjunction involves standard Boolean negation. (Notice that CZC claim that **Negated Conjunction** fails on the same understanding of negation.) It is possible that, in some cases, Negated Conjunction might fail once we bring an alternative-sensitive negation into the picture. We have not found any instances of this failure, but we are also not in a position to exclude it.

failure of Negated Conjunction observed by CZC is due to the fact that they use an instance of the inference that mixes clauses that do and do not involve overt negation. Once we add overt negation both in the premises and in the conclusion, as we have done above in Exp. 2, we observe the pattern of results expected. Since Negated Conjunction is valid, the semantics for counterfactuals employs a notion of minimal change, as on classical accounts.

One further observation: despite the fact that Substitution in general fails, on our account DeMorgan's laws still hold—provided that we formulate the relevant sentences with overt negation. In particular, the inference $(\neg \mathsf{A} \vee \neg \mathsf{B}) \mathrel{\Box\!\!\rightarrow} \mathsf{C} \vDash \neg(\mathsf{A} \wedge \mathsf{B}) \mathrel{\Box\!\!\rightarrow} \mathsf{C}$ turns out to be valid. The reason is that negation introduces alternatives regardless of its position and therefore e.g. $\neg(\mathsf{A} \wedge \mathsf{B})$ and $\neg \mathsf{A} \vee \neg \mathsf{B}$ will end up associated with the same alternatives, namely those in (58).

(58) { 

, 

, 

}

4.6 Capturing the indeterminate choices

The entry in (42) predicts that counterfactuals are supported at a state s just in case, for all alternatives of the antecedent p, and for some alternatives of the consequent q, the conditional $p \Rightarrow q$ is supported by s. Many participants in our experiments (as well as in CZC's experiments) judge counterfactuals indeterminate in cases where different alternatives of the same antencedent yield different results.

There are different options of how to capture these indeterminate judgments in our proposal; a simple one is to introduce undefinedness in the semantics.[18] Doing this in detail goes beyond the scope of this paper, but in the follow we sketch the gist of this move. The basic idea is that counterfactuals display homogeneity (see, a.o., Križ 2015) with respect to the alternatives generated by the antecedent.

To introduce undefinedness in an inquisitive semantics, we need a more complex picture of inquisitive meaning.[19] Each sentence is assigned both support conditions and definedness conditions. In particular, counterfactuals are defined at s just in case all alternatives of the antecedent yield a conditional that is supported at s, or none does.[20]

18 A different option would be pragmatic: we could hold that some participants opt for the indeterminate choice rather than false because they perceive that the sentence has more than one possible reading, depending on the understood QUD. As a result, rather than committing to false they choose the indeterminate response (see Bar-Lev 2018 for a similar proposal with different data involving plural definites). See also Ciardelli et al. 2018b: sec. 5.6 for discussion of indeterminacy.

19 This treatment is loosely based on Ciardelli, Groenedijk and Roelofsen's treatment of presupposition in inquisitive systems in Ciardelli et al. 2012.

20 Of course, once we introduce this extra component of meaning, all clauses in the semantics should

$$(59)\quad \begin{array}{lll} \mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{B} \text{ defined at } s & \text{iff} & \forall p \in Alt(\mathsf{A})\ \exists q \in Alt(\mathsf{B}) \text{ s. t. } s \subseteq p \Rightarrow q, \text{ or} \\ & & \forall p \in Alt(\mathsf{A})\ \exists q \in Alt(\mathsf{B}) \text{ s. t. } s \nsubseteq p \Rightarrow q \\ s \vDash \mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{B} & \text{iff} & \forall p \in Alt(\mathsf{A})\ \exists q \in Alt(\mathsf{B}) \text{ such that } s \subseteq p \Rightarrow q \end{array}$$

Under this modification of the proposal, the counterfactual cases above involving negation (i.e. negated conjunction, negative disjunction etc) would all come out undefined, as one of their alternatives yield a different result than the others in the given context. How undefinedness is mapped to truth-value judgments is a complex issue (see von Fintel 2004 among others for discussion), but we could assume that some participants mapped it to indeterminate and others to false.

4.7 How far does overt negation go?

Now we address the apparent discrepancy between our finding and one of CZC's findings. Recall from §1: CZC run a follow-up to their main experiment, where they replace *down* with *not up*. The relevant sentence is the following:

(60) If switch A or switch B was not up, the light would be off.

60% of their non-excluded participants judge the sentence true. This is in apparent conflict with the result of our experiment, where we found that sentences involving disjunctions of negations get endorsed at comparatively low rates, and induce selection of the picture where both twins are on the right-hand side of the see-saw.

We are not in a position to give a full account of this discrepancy. To do this, further work seems to be needed. But we want to notice a possible point of weakness in CZC's data, as well as suggest a possible explanation for the discrepancy.

First, we notice that the exclusion rate of participants in the relevant experiment is extremely high (71.66%). CZC themselves are unclear about the reason behind this. But an exclusion rate of this sort does raise some worries about the reliability of this particular data point.[21]

Second, even if we take CZC's data at face value, we might explain the difference between CZC's data and ours by postulating that negation interacts differently with the QUD, depending on how high in the sentence it appears. In particular, it may be that negation is lower in (60) than in our (61), repeated below from above.

be rewritten to incorporate a recursive computation of definedness conditions (and this will require making decisions about how the relevant kind of undefinedness projects across various connectives). This project is not central to our goals in this paper, so we leave it as a sketch here.

21 Relatedly, we know that about 18% of all of CZC's participants (60% of the 30% of participants whose data were analyzed) judged (60) true. However, we have no information about how their *excluded* participants judged (60), and it could very well be that the minority of the 70% of participants that were excluded judged the sentence true, which would make CZC's results comparable to ours.

(61) If Arthur or Bill were not on the left, the see-saw would be balanced.

It might be that (60) negation doesn't take scope over the whole verb phrase, but rather only over the adverb *up*. Conversely, in (61) negation takes scope over the whole VP. We conjecture that this difference in scope is crucial in terms of generating alternatives. In particular, it might be that negation only generates alternatives if it takes scope over the VP. If this was correct, it would explain the difference between (60) and (61). In (61), negation does not contribute to the generation of alternatives, but in (60) it does.

To fully develop this hypothesis, one should provide a full theory of the interaction between negation and QUD, and test it with different positions of negation. We think this is a promising direction to explore, but have to leave this for future work.

5 Comparison with other accounts

In the previous section, we have developed a positive account of counterfactuals that predicts both the data already known in the literature, and the new data that we have presented in §3. In this section, we review in detail three alternative accounts of the data. The first is due to CZC themselves (2018b), the second to Kathrin Schulz (2019), and the third to Bar-Lev and Fox (2018, 2020).

Before starting, it's helpful to highlight three choice points on which the three accounts differ from each other.

(1) The nature of Simplification. The first choicepoint is whether we regard the Simplification effect as hardwired in the semantics, or due to broadly pragmatic/scalar effects.

(2) Minimal Change. A second choice point concerns whether accounts retain or jettison the idea that the basic truth conditions of counterfactual appeal to a notion of minimal change.

(3) Alternatives: lexical vs contextual. Alternatives may be generated exclusively on the basis of the linguistic material in the scope of alternative-sensitive operators (what we call the 'lexical' conception of alternatives), or may be generated also on the basis of other factors (what we call the **contextual** conception).

These three choice points allow us to classify some of the accounts of CZC's data: CZC's own account, Kathrin Schulz's (2019), and Moysh Bar-Lev and Danny Fox's (2018, 2020), as well as our own; see Table 8.

Let us start with how our account fares with respect to the three main choice points outlined above. First, we use a semantic, rather than a scalar algorithm for computing alternatives. Second, we preserve the classical idea that semantics for

	Simplification	Minimal Change	Generation of Alternatives
Our account	semantic	yes	contextual
CZC	semantic	no	lexical
Schulz	semantic	yes	lexical
Bar-Lev & Fox	scalar	yes	contextual

Table 8 Comparison between other accounts and our account.

counterfactuals relies on a notion of minimal change. Third, and crucially, we use a contextual accounts of alternatives. This means that the alternatives to a sentence are not generated merely by the prejacent of alternative-sensitive operators. Rather, they are also partly determined by the broader linguistic context. In particular, our account crucially exploits the idea that negation is QUD-sensitive.

Both CZC and Schulz operate within an inquisitive semantics, which as we have seen produces a semantic account of Simplification. In addition, both accounts use a lexical conception of alternatives. But these accounts differ on choice point (ii). CZC advocate abandoning the minimal change requirement, in favor of an alternative account, 'background semantics'. Conversely, Schulz preserves a classical minimal change semantics.

Bar-Lev & Fox's account exploits a traditional semantic framework, on which meanings are truth conditions. Bar-Lev and Fox supplement this framework with a scalar account of Simplification, based on an exhaustivity operator with a particularly strong meaning. In combination with some assumptions about alternatives and negation, they predict CZC's data.

As the table indicates, our account departs from all of these accounts. We think the main upshot of this debate is not that counterfactuals provide evidence for or against a certain semantic framework. Rather, the takeaway lessons are (i) the meanings of disjunction and negation are associated with alternatives, which interact with the meaning of counterfactuals, and (ii) that context plays a key role in fixing the alternatives. In the following, we compare our account to the others in more detail.

5.1 CZC's account

Account. CZC try to account for their data by making two main maneuvers. On the one hand, they adopt an inquisitive semantics for connectives; on the other, they abandon comparative closeness semantics for counterfactuals, replacing it with a different kind of premise semantics (what they call 'background semantics').

With respect to the first point, the key difference with respect to the semantics in §4 is their entry for negation, which is in (62).

(62) $s \vDash \neg \mathsf{A}$ iff $\forall t \subseteq s$: if $t \neq \varnothing$ then $t \nvDash \mathsf{A}$

On this semantics, negated sentences have a non-inquisitive meaning: they denote a set that includes just one alternative. (In other words, negation acts as a 'flattener' of alternatives.) As a result, negated conjunctions and disjunctions of negations have different meanings. In particular, $\neg(\mathsf{A} \wedge \mathsf{B})$ has a non-inquisitive meaning and denotes only one alternative, namely the complement of W with respect to $\mathsf{A} \wedge \mathsf{B}$. Conversely, $\neg \mathsf{A} \vee \neg \mathsf{B}$ has an inquisitive meaning and denotes the set including the alternatives corresponding to the negated disjuncts.

CZC's second maneuver consists in defining a meaning for the metalanguage connective $\Rightarrow$ that allows for failures of **Negated conjunction**. A detailed description of their system would take us far from our main goals, but we explain informally the main ideas.

CZC adopt a premise semantics for counterfactuals (see §1), broadly in the style of Kratzer. The key difference is that some information in the premise set can be erased by counterfactual antecedents. CZC introduce a notion of a counterfactuals antecedent 'calling into a question' a proposition in the premise set (in turn, based on a notion of causal dependence). Propositions in the premise set that are called into question are simply discarded, and counterfactuals are evaluated using the propositions that are leftover in the premise set.

The resulting system is still broadly in keeping with the main ideas behind Kratzer semantics (as emphasized by Schulz 2019). Crucially, though, the semantics gives up on the idea of comparative closeness. The reason is that erasing premise sets is equivalent to changing the position of worlds in a closeness ordering, on the basis of counterfactual antecedents. (For a somewhat similar technology, see Santorio 2019.) This is incompatible with the basic idea of comparative closeness semantics, which assumes that worlds are evaluated against a fixed, stable closeness ordering.

Predictions. Let us examine the predictions of CZC's theory *vis-à-vis* our data. CZC correctly predict that (63)-a and (63)-b should get different levels of endorsement, and give rise to the selection of different pictures. However, CZC also predict that (63)-c and (63)-d should pattern with (63)-a, rather than (63)-b. As we have seen, this prediction is incorrect.

(63) a. If Arthur or Bill were on the right, the see-saw would be balanced.
b. If Arthur and Bill were not both on the left, the see-saw would be balanced.
c. If Arthur or Bill were not on the left, the see-saw would be balanced.
d. If Arthur was not on the left or Bill was not on the left, the see-saw would be balanced.

Also, CZC predict that the sentences in (64), which involve no binary connectives and only differ in virtue of the presence of overt negation in the antecedent, also behave analogously. Again, this prediction appears to be incorrect.

(64) a. If Arthur was on the right, the see-saw would be balanced.
b. If Arthur was not on the left, the see-saw would be balanced.

In summary, CZC's theory falls short of recognizing the alternative-generating role of negation.[22]

5.2 Schulz's account

Account. Schulz (2019) also works with inquisitive semantics, but she crucially makes the assumption, with which we agree, that negation contributes to generating alternatives. To introduce Schulz's meaning for negation, we need to first introduce her notion of a partition induced by a sentence, $Q(\mathsf{A})$.

$$Q(\mathsf{A}) = \{s\text{: for all } w, w' \in s, \text{ and for all atomic formulas } p \text{ in } \mathsf{A}, V(w,p) = V(w',p)\}$$

I.e., the partition induced by A is the set of states that agree on the answers to all the atomic formulas figuring in A. For example, the partition induced by the conjunction $\mathsf{A} \wedge \mathsf{B}$ is the set of the following four states:

$$\{s_{\mathsf{AB}}, s_{\bar{\mathsf{A}}\mathsf{B}}, s_{\mathsf{A}\bar{\mathsf{B}}}, s_{\bar{\mathsf{A}}\bar{\mathsf{B}}}\}$$

22 An anonymous referee suggests that CZC could appeal to context dependence to get out of this predicament. They could suggest that the default (or even the only) interpretation of (64)-involves a 'background' (CZC's replacement parameter for the ordering source) that includes the information that Billy is on the left, while the default (or only) interpretation of (64)-a involves a background includes no information, at all, so that worlds where Billy is on the right are in play. The referee recognizes that this option would be ad hoc, but suggests that our proposal to build reference to the QUD is no less stipulative. We disagree with the referee's assessment here, especially in light of the fact that, we pointed out above, negation routinely interacts with alternatives.

Given this, her entry for negation is the following:

(65) $s \vDash \neg \mathsf{A}$ iff $s \vDash Q(\mathsf{A})$ and $s \perp \mathsf{A}$

(The notion of a state contradicting a sentence, $s \perp \mathsf{A}$, is understood as the state and the possible worlds content of the sentence having an empty intersection.)

Given the entry in (65), Schulz also assigns a different meaning to $\bar{\mathsf{A}} \vee \bar{\mathsf{B}}$ and $\neg(\mathsf{A} \wedge \mathsf{B})$ differ, but in a different way than in Ciardelli et al. 2018b. As for CZC, $\bar{\mathsf{A}} \vee \bar{\mathsf{B}}$ has two alternatives, namely $\{|\bar{\mathsf{A}}|, |\bar{\mathsf{B}}|\}$. But, differently from CZC's semantics, $\neg(\mathsf{A} \wedge \mathsf{B})$ denotes three alternatives, namely $\{|\bar{\mathsf{A}}\mathsf{B}|, |\mathsf{A}\bar{\mathsf{B}}|, |\bar{\mathsf{A}}\bar{\mathsf{B}}|\}$.

Predictions. This difference allows Schulz to make different predictions in some cases where negation is involved, including the variant of the two switches scenario in Schulz 2019 (see also McHugh & Cremers 2019 for another scenario where Schulz's account makes the right predictions). At the same time, the account fails to predict some of the data in §3. In particular, the same cases that created trouble for CZC also create trouble for Schulz.

(66) a. If Arthur or Bill were on the right, the see-saw would be balanced.
b. If Arthur or Bill were not on the left, the see-saw would be balanced.
c. If Arthur was not on the left or Bill was not on the left, the see-saw would be balanced.

(67) a. If Arthur was on the right, the see-saw would be balanced.
b. If Arthur was not on the left, the see-saw would be balanced.

Schulz predicts that the three sentences in (66) should be equivalent and so should be the pair of sentences in (67). The reason is that, while she correctly argues that negation introduces alternatives, she takes alternatives to be determined exclusively by the lexical material in the scope of negation. In both (66)-b and (67)-b, negation takes scope over just an atomic sentence. Hence the presence of negation is predicted to have no effect on the generation of alternatives. Our account is very similar in formalism and in spirit, but we crucially let alternatives be determined via a contextually determined QUD.

5.3 Bar-Lev and Fox's account

A third account of the phenomenon, due to Bar-Lev and Fox (2020), is based on a scalar account of Simplification. As we discuss below, this approach can account for all of our results. However, the assumptions it makes about negation are problematic when we move beyond counterfactuals.

Account. Bar-Lev & Fox's (2020) develop an account of the phenomenon that is based on a scalar account of Simplification. In the background, they adopt a semantic account to scalar implicature (see Chierchia 2004, Chierchia et al. 2012, Magri 2009, Meyer 2013, Chierchia 2013 among others). 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. More precisely, EXH takes as arguments a sentence and a set of alternatives and returns the conjunction of the sentence with the negation of a subset of the alternatives — the alternatives that are 'innocently excludable.' Informally, EXH looks at all the maximal consistent subsets of alternatives to a sentence, and negates all alternatives that are in all of those subsets. The effect is to strengthen the sentence as much as possible, while avoiding contradictions and arbitrary choices between alternatives. The meaning of EXH is given in (68), while the definition of innocent exclusion is in (69), where 'C' stands for a set of salient alternatives to a sentence.

(68) $\textsc{exh}(A)(p)(w) = p_w \wedge \forall q \in \mathsf{IE}(p,A)[\neg q_w]$

(69) $\mathsf{IE}(p,C) := \bigcap \left\{ C' \;\middle|\; \begin{array}{l} C' \subseteq C \text{ and } C' \text{ is a maximal subset of } C \\ \text{such that } \{\neg q : q \in C)\} \cup \{p\} \text{ is consistent} \end{array} \right\}$

Let us illustrate how this works for a simple disjunctive sentence that gives rise to the implicature that Arthur and Bill are not both on the right.

(70) Arthur or Bill is on the right.
$\rightsquigarrow$ *Arthur and Bill are not both on the right*

(70) is parsed as involving a covert exhaustivity operator, as in (71). We assume that the alternatives of (70) are in (72).[23]

(71) EXH[Arthur or Bill is on the right]

(72) $\left\{ \begin{array}{ll} \text{A or B is on the right} & (\mathsf{A} \vee \mathsf{B}) \\ \text{A is on the right} & \mathsf{A} \\ \text{B is on the right} & \mathsf{B} \\ \text{A and B are on the right} & (\mathsf{A} \wedge \mathsf{B}) \end{array} \right\}$

Given the alternatives in (72), only the conjunctive alternative $(\mathsf{A} \wedge \mathsf{B})$ is excludable. This is because there are only two maximal consistent subsets of excludable alternatives, $\{\mathsf{A}, \mathsf{A} \wedge \mathsf{B}\}$ and $\{\mathsf{B}, \mathsf{A} \wedge \mathsf{B}\}$, and only the conjunctive alternative appears in both. This yields the intuitively correct prediction, i.e. the implicature in (70).

This algorithm, by itself, does not derive Simplification, but it can be refined

23 For relevant discussion on alternatives see Breheny et al. (2017) and references therein.

in various ways to do so. In particular, Bar-Lev and Fox argue that the algorithm should not only conjoin the prejacent with the negation of innocently excludable alternatives, but it should also conjoin—'includes'—the prejacent with a subset of other non-negated alternatives. This refined version of the algorithm allows it to make predictions about a whole range of other phenomena, including free choice inferences (see e.g. Fox 2007, among others) and Simplification.

The new algorithm requires a definition of innocently *includable* alternatives, spelled out as follows: innocently includable alternatives are those that are in all maximal subsets of alternatives that can be conjoined consistently with the assertion and with the negation of innocently excludable alternatives. Based on (72), the definition of EXH is then straightforward: EXH conjoins the prejacent with all the innocently includable alternatives and the negation of all innocently excludable ones.

(73) $\textsc{exh}(C)(p)(w) = p_w \wedge \forall q \in \mathsf{IE}(p,C)[\neg q_w] \wedge \forall r \in \mathsf{II}(p,C)[r_w]$

(74) a. $\mathsf{II}(p,C) := \bigcap \left\{ C'' \;\middle|\; \begin{array}{l} C'' \subseteq C \text{ and } C'' \text{ is a maximal subset of } C \text{ s.t.} \\ \{r : r \in C''\} \cup \{\neg q : q \in \mathsf{IE}(p,C)\} \cup \{p\} \text{ is consistent} \end{array} \right\}$

In addition to this, as Bar-Lev and Fox point out, they need an extra assumption to derive the data about antecedents with negated conjunctions. They need to assume that EXH is an alternative of negation. As we will see, this assumption is key for deriving both CZC's data and ours, but it leads to problems elsewhere in the theory of implicature.

Predictions, 1/2: counterfactual data. We show first that, once we strengthen the meaning of EXH with innocently includable alternatives, Simplification is derived as an implicature. Consider (75) and the alternatives in (76).

(75) EXH[If A or B was on the right, the see-saw would be balanced]

(76) $\left\{ \begin{array}{lr} \text{If A or B right, see-saw balanced} & (\bar{\mathsf{A}} \vee \bar{\mathsf{B}}) \mathrel{\Box\!\!\rightarrow} \mathsf{C} \\ \text{If A right, see-saw balanced} & \bar{\mathsf{A}} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \\ \text{If B right, see-saw balanced} & \bar{\mathsf{B}} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \\ \text{If A and B right, see-saw balanced} & (\bar{\mathsf{A}} \wedge \bar{\mathsf{B}}) \mathrel{\Box\!\!\rightarrow} \mathsf{C} \end{array} \right\}$

As in the case of simple disjunction, the only innocently excludable alternative is $(\bar{\mathsf{A}} \wedge \bar{\mathsf{B}}) \mathrel{\Box\!\!\rightarrow} \mathsf{C}$. But now we have a further way of strengthening the basic meaning of the assertion, i.e. running the inclusion algorithm. There is only one subset of includable alternatives, $\{\bar{\mathsf{A}} \vee \bar{\mathsf{B}}) \mathrel{\Box\!\!\rightarrow} \mathsf{C}, \bar{\mathsf{A}} \mathrel{\Box\!\!\rightarrow} \mathsf{C}, \bar{\mathsf{B}} \mathrel{\Box\!\!\rightarrow} \mathsf{C}\}$. Thus we can include them all and obtain the Simplification inferences in (77).

(77) If A or B was on the right, the see-saw would be balanced.

$\rightsquigarrow$ If A was on the right, the see-saw would be balanced.

$\rightsquigarrow$ If B was on the right, the see-saw would be balanced.

With no further changes to the system, we would get analogous results for the case of negated conjunctions. So the innocent inclusion algorithm, by itself, does not predict CZC's and our data. It is here that Bar-Lev and Fox's extra assumption comes in. This assumption consists in stipulating that EXH is an alternative of negation. This predicts that antecedents with negated conjunctions give rise to a different (and stronger) Simplification implicature. For illustration, the sentence in (78) gives rise to the alternatives in (79).

(78) EXH[If A and B were not both on the left, the see-saw would be balanced]

(79)
$$\left\{\begin{array}{ll} \text{If A and B not both on the left, see-saw balanced} & \neg(A \wedge B) \;\Box\!\!\rightarrow C \\ \text{If A not left, see-saw balanced} & \neg A \;\Box\!\!\rightarrow C \\ \text{If B not left, see-saw balanced} & \neg B \;\Box\!\!\rightarrow C \\ \text{If A or B not left, see-saw balanced} & \neg(A \vee B) \;\Box\!\!\rightarrow C \\ \text{If EXH[A and B left], see-saw balanced} & \text{EXH}(A \wedge B) \;\Box\!\!\rightarrow C \\ \text{If EXH[A was left], see-saw balanced} & \text{EXH}(A) \;\Box\!\!\rightarrow C \\ \text{If EXH[B was left,] see-saw balanced} & \text{EXH}(B) \;\Box\!\!\rightarrow C \\ \text{If EXH[A or B were left], see-saw balanced} & \text{EXH}(A \vee B) \;\Box\!\!\rightarrow C \end{array}\right\}$$

As Bar-Lev & Fox (2020) show, the addition of the new alternatives (in particular the alternatives $\text{EXH}(A) \;\Box\!\!\rightarrow C$ and $\text{EXH}(B) \;\Box\!\!\rightarrow C$) has the effect that the alternative $\neg(A \vee B) \;\Box\!\!\rightarrow C$ is now innocently includable. As a result, (78) now has a stronger meaning, schematically represented in (80). The crucial extra element is underlined.

(80) EXH[If A and B were not both on the left, the see-saw would be balanced] =
$\neg(A \wedge B) \;\Box\!\!\rightarrow C \wedge \neg A \;\Box\!\!\rightarrow C \wedge \neg B \;\Box\!\!\rightarrow C \wedge \underline{\neg(A \vee B) \;\Box\!\!\rightarrow C}$

In other words, given the extra assumption about alternatives, (78) simplifies to the three conditionals in (81).

(81) If A and B were not both on the left, the see-saw would be balanced.

$\rightsquigarrow$ If A was not on the left, the see-saw would be balanced.

$\rightsquigarrow$ If B was not on the left, the see-saw would be balanced.

$\rightsquigarrow$ If neither A nor B was on the left, the see-saw would be balanced.

Given the last inference in (81), which is false in the given context, Bar-Lev and

Fox predict the main result of our Experiment I in §3: lower endorsement than (77) for (81) in the truth-value judgment task.

In addition, their account also predicts the results about disjunctions of negation, both clausal and non-clausal, in Experiments I and III. Consider again the sentences in (82):

(82) a. If Arthur or Bill were on the right, the see-saw would be balanced. $(\bar{A} \vee \bar{B}) \mathrel{\Box\!\!\rightarrow} C$
b. If Arthur or Bill were not on the left, the see-saw would be balanced. $(\neg A \vee \neg B) \mathrel{\Box\!\!\rightarrow} C$
c. If Arthur was not on the left or Bill was not on the left, the see-saw would be balanced. $(\neg A \vee \neg B) \mathrel{\Box\!\!\rightarrow} C$

Bar-Lev and Fox predict that (82-b) and (82-c) pattern together. This is because (82-b) and (82-c) (but not (82-a)) give rise to alternatives including EXH. This makes the alternative non-excludable and then includable and, in turn, lead to the stronger simplification meaning in (84), which is false in our scenario.

(83)
$$\left\{\begin{array}{ll} \text{If A not on the left or B not on the left, balanced} & (\neg A \vee \neg B) \mathrel{\Box\!\!\rightarrow} C \\ \text{If A was not on the left, balanced} & \neg A \mathrel{\Box\!\!\rightarrow} C \\ \text{If B was not on the left, balanced} & \neg B \mathrel{\Box\!\!\rightarrow} C \\ \text{If A not on the left and B not on the left, balanced} & (\neg A \wedge \neg B) \mathrel{\Box\!\!\rightarrow} C \\ \text{If } \textsc{exh}\text{[A on the left] or } \textsc{exh}\text{[B on the left], balanced} & (\textsc{exh}(A) \vee \textsc{exh}(B)) \mathrel{\Box\!\!\rightarrow} C \\ \text{If } \textsc{exh}\text{[A was on the left], balanced} & \textsc{exh}(A) \mathrel{\Box\!\!\rightarrow} C \\ \text{If } \textsc{exh}\text{[B was on the left], balanced} & \textsc{exh}(B) \mathrel{\Box\!\!\rightarrow} C \\ \text{If } \textsc{exh}\text{[A on the left] and } \textsc{exh}\text{[B on the left], balanced} & (\textsc{exh}(A) \wedge \textsc{exh}(B)) \mathrel{\Box\!\!\rightarrow} C \end{array}\right\}$$

(84) $(\neg A \mathrel{\Box\!\!\rightarrow} C) \wedge (\neg B \mathrel{\Box\!\!\rightarrow} C) \wedge ((\neg A \wedge \neg B) \mathrel{\Box\!\!\rightarrow} C)$

What about Experiment II? Bar-Lev and Fox do explain the results of the truth-value judgment task — they explain the differences in the endorsement rates of (85-a) and (85-b).

(85) a. If A was not on the left, the see-saw would be balanced. $\neg A \mathrel{\Box\!\!\rightarrow} C$
b. If A was on the right, the see-saw would be balanced. $\bar{A} \mathrel{\Box\!\!\rightarrow} C$

The reason is that (85-a) is associated with an extra alternative, which leads to the computation of an extra implicature, which happens to be false in the context. The alternatives associated with (85-a) are:

(86)
$$\left\{\begin{array}{ll} \text{If A was not on the left, balanced} & \neg A \mathrel{\Box\!\!\rightarrow} C \\ \text{If } \textsc{exh}\text{[A was on the left], balanced} & \textsc{exh}(A) \mathrel{\Box\!\!\rightarrow} C \end{array}\right\}$$

As it turns out, $\textsc{exh}(A) \mathrel{\Box\!\!\rightarrow} C$ is excludable giving rise to the implicature in (87) (here shown for the case of A, the corresponding inference is predicted for B). (87) is saying that if Arthur was not on the left the see-saw would be balanced, but that it's not true that, if Arthur was on the the left and Bill was not, the see-saw would be balanced. This is false, hence (85-a) is correctly predicted to be endorsed at a lower level than (85-b).

(87) $\neg A \mathrel{\Box\!\!\rightarrow} C \wedge \neg((A \wedge \neg B) \mathrel{\Box\!\!\rightarrow} C)$

Turning to the results of the picture selection task, the implicature approach could account for our results under the assumption that participants selected the picture(s) they thought were compatible with the *exhaustified* antecedent they were asked about e.g. *What would it look like if Arthur Bill were on the right?*[24] This account for the difference in picture selections of the main cases above, e.g. POSITIVE DISJUNCTION VS. NEGATED CONJUNCTIONS and NEGATIVE DISJUNCTIONS, as they lead to different meanings when exhaustified. To illustrate, consider positive disjunction first, (88), which, exhaustified against the alternatives in (89), give rise to the familiar exclusive disjunction meaning, incompatible with both boys being on the right. This explains why participants tended not to choose the picture representing this situation in the positive disjunction condition.

(88) $\textsc{exh}$[Arthur or Bill were on the right] = $(A \vee B) \wedge \neg(A \wedge B)$

(89) $$\left\{\begin{array}{lr} \text{A or B were on the right} & (A \vee B) \\ \text{A was on the right} & A \\ \text{B was on the right} & B \\ \text{A and B were on the right} & (A \wedge B) \end{array}\right\}$$

The case of negated conjunction in (90), on the other hand, is associated to more alternatives, given the assumption that EXH is an alternative of negation.

(90) Arthur and Bill were not both on the left.

(91) $$\left\{\begin{array}{ll} \text{A and B were not on the left} & \neg(A \wedge B) \\ \text{A was not on the left} & \neg A \\ \text{B was not on the left} & \neg B \\ \text{A or B were not on the left} & \neg(A \vee B) \\ \textsc{exh}\text{[A and B were on the left]} & \textsc{exh}(A \wedge B) \\ \textsc{exh}\text{[A was on the left]} & \textsc{exh}(A) \\ \textsc{exh}\text{[B was on the left]} & \textsc{exh}(B) \\ \textsc{exh}\text{[A or B were on the left]} & \textsc{exh}(A \vee B) \end{array}\right\}$$

24 Thanks to an anonymous reviewer for very helpful discussion on this point.

The presence of the alternative $\textsc{exh}(\mathsf{A} \vee \mathsf{B}) = (\mathsf{A} \vee \mathsf{B}) \wedge \neg(\mathsf{A} \wedge \mathsf{B})$ makes the alternative $\neg(\mathsf{A} \vee \mathsf{B})$ not excludable anymore. Neither of them is includable either, given the presence of the other. The result is that no implicature is derived in this case. The meaning therefore remains compatible with the situation in which neither boy is on the left, which, in turn, can account for the higher selection of the picture were both boys are on the right.

The same can be shown for the negative disjunction case, whether clausal or not. For the same reasons as in the negated conjunction case, (92), exhaustified against the alternatives in (92), unlike its positive counterpart, remains compatible with both boys being on the right (neither being on the left).

(92) Arthur was not on the left or Bill was not on the left = $\neg\mathsf{A} \vee \neg\mathsf{B}$

(93)
$$\left\{\begin{array}{ll} \text{Arthur was not on the left or Bill was not on the left.} & \neg\mathsf{A} \vee \neg\mathsf{B} \\ \text{A was not on the left} & \neg\mathsf{A} \\ \text{B was not on the left} & \neg\mathsf{B} \\ \text{Arthur was not on the left and Bill was not on the left.} & \neg\mathsf{A} \wedge \neg\mathsf{B} \\ \textsc{exh}\text{[A on the left] or } \textsc{exh}\text{[B on the left]} & \textsc{exh}[\mathsf{A}] \vee \textsc{exh}[\mathsf{B}] \\ \textsc{exh}\text{[A was on the left]} & \textsc{exh}(\mathsf{A}) \\ \textsc{exh}\text{[B was on the left]} & \textsc{exh}(\mathsf{B}) \\ \textsc{exh}\text{[A on the left] and } \textsc{exh}\text{[B on the left]} & \textsc{exh}[\mathsf{A}] \wedge \textsc{exh}[\mathsf{B}] \end{array}\right\}$$

Finally, the implicature approach can also account for the fact that when participants were presented with negated simple antecedents like (85-a), they tended to select the picture where both boys are on the right, to a higher rate than those who saw the positive counterpart in (85-b). This is because the exhaustified meaning of the positive antecedent is again different from that of the negated one. (94) leads to the inference that only Arthur was on the right, while (96) is compatible with Bill not being on the left.

(94) Arthur was on the right = $\bar{\mathsf{A}} \wedge \neg\bar{\mathsf{B}}$

(95)
$$\left\{\begin{array}{ll} \text{A was on the right} & \bar{\mathsf{A}} \\ \text{B was on the right} & \bar{\mathsf{B}} \end{array}\right\}$$

(96) Arthur was not on the left = $\neg\mathsf{A}$

(97)
$$\left\{\begin{array}{lr} \text{A was not on the left} & \neg\mathsf{A} \\ \text{B was not on the left} & \neg\mathsf{B} \\ \textsc{exh}\text{[A was on the left]} & \textsc{exh}[\mathsf{A}] \\ \textsc{exh}\text{[B was on the left]} & \textsc{exh}[\mathsf{B}] \end{array}\right\}$$

Predictions, 2/2: problems with indirect implicatures. We saw that the implicature account manages to predict all of the data about counterfactuals. As we briefly show now, however, to achieve these results, this account needs to make assumptions that are problematic elsewhere. In particular, the stipulation that EXH is an alternative of negation creates trouble in the computation of indirect implicatures.

Consider a simple negated conjunction as in (98), which gives rise the 'indirect' implicature that at least one of Arthur and Bill is on the left.

(98) Arthur and Bill are not both on the left.
$\rightsquigarrow$ *Arthur or Bill is on the left*

This inference is derived by exhaustifying (99) against the alternatives in (100), in parallel to the simple disjunctive case seen above: the alternative $\neg(A \vee B)$ is excludable and its exclusion gives rise to the observed implicature.

(99) EXH[not[Arthur and Bill are both on the left]] =
$\neg(A \wedge B) \wedge (A \vee B)$

(100) $\left\{ \begin{array}{lr} \text{A and B are not both on the left} & \neg(A \wedge B) \\ \text{A is not on the left} & \neg A \\ \text{B is not on the left} & \neg B \\ \text{A or B is not on the left} & \neg(A \vee B) \end{array} \right\}$

This is uncontroversial for any account of implicatures. The problem is that, as we saw above in (90), as soon as we make the crucial assumption that EXH is an alternative of negation, we lose the account of (99).

The presence of the extra alternatives make everything not excludable or includable. The result is that the indirect implicature above is not derived anymore. This was crucial as an account of our data above, but creates problem for the simple case of indirect implicatures.

Bar-Lev & Fox (2020: fn. 59) themselves discuss this problem. They however suggest that the problem can be alleviated once we take into account the sensitivity of implicatures to the focus structure of the sentence. Following Fox & Katzir (2011) among others, they argue that only alternatives of items within the focus constituent of a sentence are considered for implicature computation. Therefore, in a sentence like (98), negation will be replaced by EXH if and only if it is part of the focus constituent. This leaves room for deriving the indirect implicature above when negation is not part of the focused constituent of the sentence, as it happens e.g. in (101).

(101) $[\text{A and B}]_{\text{F}}$ are not both up.

We know from the literature on alternatives, however, that this prediction is problematic (see Romoli 2013, Trinh & Haida 2015, Breheny, Klinedinst, Romoli & Sudo 2017 among others). In particular, we can control for the focus constituent of the sentence by adding an explicit question in the context. Consider (102):

(102) **Cynthia**: You look surprised. What's going on?
Donna: Arthur and Bill are not both on the left. (Maybe they learned to use a see-saw!)

Cynthia's question makes it clear that the whole sentence by Donna is in focus (i.e. the whole sentence is a possibile answer to the question; Rooth 1992 among many others). Nonetheless, (102) intuitively still gives rise to the inference that one of Arthur and Bill is on the left, against Bar-Lev & Fox's (2020) prediction.

In sum, the assumption that negation and EXH are alternatives can account for the effects we found with counterfactuals, but under-generates as we move beyond those to implicatures of simple negative sentences.

6 Conclusion

Classical semantics for counterfactuals are based on a notion of comparative similarity and minimal change. These semantics suffer from some well-known difficulties related to disjunctive antecedents, which have been used in various ways to support the claim that truth-conditionally equivalent sentences are not substitutable in counterfactual antecedents. In their study, CZC present a particularly dramatic instance of this difficulty, showing that negative disjunctions and negated conjunctions cannot be replaced while keeping fixed truth value. They propose a new semantics for counterfactuals, which builds on inquisitive semantics (see Ciardelli et al. 2018a) and gives up on comparative similarity and minimal change.

In this paper, we have presented a study consisting of a series of experiments that start from CZC's general format, but using a different scenario, and involving an extra task based on the selection of pictures. Our results replicate the basic effect found by CZC, but also suggest that that effect is linked to the presence of overt negation. We have suggested that the correct account of the phenomenon (i) treats counterfactual antecedents as generating alternatives via mechanisms hardwired in the semantics; (ii) holds on to minimal change semantics; (iii) appeals in part to context to generate alternatives. We have developed an inquisitive semantics for counterfactuals that accounts for all the experimental data. We have also emphasized, though, that the main theoretical upshot concerns the nature of alternatives in conditionals and the role of negation, rather than inquisitive semantics *per se*.

Finally, our studies were conducted against the backdrop of a rich literature in psychology (e.g., De Vega 2008, Nieuwland 2013, Nieuwland & Martin 2012, García-Madruga et al. 2001, Kaup et al. 2006, Khemlani et al. 2014, Orenes et al. 2014, Baggio et al. 2016: for a review, see Kulakova & Nieuwland, 2016), adding a novel paradigm that can be easily adapted to various logical scenarios, populations, and languages. We thus hope that this work is not only informative for formal semantics of conditionals, but also for work on the cognition of counterfactuality.

References

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

Baggio, Giosuè, Paolo Cherubini, Doris Pischedda, Anna Blumenthal, John-Dylan Haynes & Carlo Reverberi. 2016. Multiple neural representations of elementary logical connectives. *NeuroImage* 135. 300–310.

Bar-Lev, Moshe. 2018. *Free choice, homogeneity and innocent inclusion*: The Hebrew University of Jerusalem dissertation.

Bar-Lev, Moshe & Danny Fox. 2020. Free choice, simplification, and innocent inclusion. *Natural Language Semantics* .

Beaver, David I. & Brady Z. Clark. 2008. *Sense and sensitivity: How focus determines meaning*. Wiley-Blackwell.

Breheny, Richard, Nathan Klinedinst, Jacopo Romoli & Yasutada Sudo. 2017. The symmetry problem: current theories and prospects. *Natural Language Semantics* 1–26.

Burgess, John P. 1981. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic* 22(1). 76–84.

Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Adriana Belletti (ed.), *Structures and Beyond: The Cartography of Syntactic Structures*, vol. 3, 39–103. Oxford: Oxford University Press.

Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice, and intervention.* Oxford University Press.

Chierchia, Gennaro, Danny Fox & Benjamin Spector. 2012. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning volume 3*, Berlin: Mouton de Gruyter.

Ciardelli, Ivano. 2016. Lifting conditionals to inquisitive semantics. In *Proceedings of SALT 26*, 732–752.

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2012. Inquisitive semantics—nasslli 2012 lecture notes.

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2018a. *Inquisitive semantics*. Oxford University Press.

Ciardelli, Ivano, Linmin Zhang & Lucas Champollion. 2018b. Two switches in the theory of counterfactuals: A study of truth conditionality and minimal change. *Linguistic and Philosophy* 41. 577–621.

De Vega, Manuel. 2008. Levels of embodied meaning: From pointing to counterfactuals. *Symbols and embodiment* 285–308.

Fiddick, Laurence, Leda Cosmides & John Tooby. 2000. No interpretation without representation: The role of domain-specific representations and inferences in the wason selection task. *Cognition* 77(1). 1–79.

Fine, Kit. 1975. Truth, vagueness and logic. *Synthese* .

von Fintel, Kai. 2001. Counterfactuals in a dynamic context. *Current Studies in Linguistics Series* 36. 123–152.

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. Palgrave.

Fox, Danny & Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19(1). 87–107.

Franke, Michael. 2011. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics* 4(1). 1–82. http://dx.doi.org/10.3765/sp.4.1.

García-Madruga, Juan A, Sergio Moreno, Nuria Carriedo, Francisco Gutiérrez & Philip N Johnson-Laird. 2001. Are conjunctive inferences easier than disjunctive inferences? a comparison of rules and models. *The Quarterly Journal of Experimental Psychology Section A* 54(2). 613–632.

Gillies, Anthony S. 2007. Counterfactual scorekeeping. *Linguistics and Philosophy* 30(3). 329–360.

Icard, Thomas F. 2017. From programs to causal models. In *Proceedings of the 21st amsterdam colloquium*, 35–44.

Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistic and Philosophy* 30(6). 669–690.

Kaufmann, Stefan. 2013. Causal premise semantics. *Cognitive science* 37(6). 1136–1170.

Kaufmann, Stefan. 2017. The limit assumption. *Semantics & Pragmatics* 18(10). http://dx.doi.org/https://doi.org/10.3765/sp.10.18.

Kaup, Barbara, Jana Lüdtke & Rolf A Zwaan. 2006. Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics* 38(7). 1033–1050.

Khemlani, Sangeet, Isabel Orenes & Philip N Johnson-Laird. 2014. The negations of conjunctions, conditionals, and disjunctions. *Acta Psychologica* 151. 1–7.

Klinedinst, Nathan. 2007. *Plurality and possibility*: UCLA dissertation.

Klinedinst, Nathan. 2009. (simplification of)^2 disjunctive antecedents. *MIT Working Papers in Linguistics* 60.

Kratzer, Angelika. 1981. The notional category of modality. In H. J. Eikmeyer & H. Rieser (eds.), *Words, worlds, and contexts: New approaches to word semantics*, Berlin: de Gruyter.

Kratzer, Angelika. 1986. Conditionals. In *Chicago linguistics society 22*, 1–15.

Kratzer, Angelika. 2012. *Modals and conditionals: New and revised perspectives*, vol. 36. Oxford University Press.

Kraus, Sarit, Daniel Lehmann & Menachem Magidor. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial intelligence* 44(1). 167–207.

Križ, Manuel. 2015. *Aspects of homogeneity in the semantics of natural language*: University of Vienna dissertation.

Kulakova, Eugenia & Mante S Nieuwland. 2016. Understanding counterfactuality: A review of experimental evidence for the dual meaning of counterfactuals. *Language and Linguistics Compass* 10(2). 49–65.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* .

Lewis, David K. 1973a. *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lewis, David K. 1973b. *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lewis, David K. 1973c. Counterfactuals and comparative possibility. *Journal of Philosophical Logic* 2(4). 418–446.

Litman, Leib, Jonathan Robinson & Tzvi Abberbock. 2017. Turkprime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods* 49(2). 433–442.

Magri, Giorgio. 2009. *A theory of individual-level predicates based on blind mandatory scalar implicatures*: Massachusetts Institute of Technology dissertation.

McHugh, Dean & Alexandre Cremers. 2019. Negation and alternatives in conditional antecedents. In *Proceedings of the 22nd amsterdam colloquium*, 289–298.

Meyer, Marie-Christine. 2013. *Ignorance and grammar*: MIT dissertation.

Nieuwland, Mante S. 2013. "If a lion could speak . . . ": Online sensitivity to propositional truth-value of unrealistic counterfactual sentences. *Journal of Memory and Language* 68(1). 54 – 67. http://dx.doi.org/https://doi.org/10.1016/j.jml.2012.08.003. http://www.sciencedirect.com/science/article/pii/S0749596X12000915.

Nieuwland, Mante S & Andrea E Martin. 2012. If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition* 122(1). 102–109.

Orenes, Isabel, David Beltrán & Carlos Santamaría. 2014. How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and*

Language 74. 36–45.

Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5. 1–69.

Romoli, Jacopo. 2013. A problem for the structural characterization of alternatives. *Snippets* 27. 14–15.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 117–121.

Santorio, Paolo. 2018. Alternatives and truthmakers in conditional semantics. *Journal of Philosophy* 115(10). 513–549. http://dx.doi.org/10.5840/jphil20181151030.

Santorio, Paolo. 2019. Interventions in premise semantics. *Philosophers' Imprint* 19.

Schulz, Katrin. 2019. The similarity approach strikes back: Negation in counterfactuals. *Proceedings of Sinn und Bedeutung* 22(2). 343–360. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/110.

Sperber, Dan, Francesco Cara & Vittorio Girotto. 1995. Relevance theory explains the selection task. *Cognition* 57(1). 31–95.

Stalnaker, Robert. 1968. A theory of conditional. In N. Recher" (ed.), *Studies in logical theory*, Oxford.

Stalnaker, Robert. 1978. Assertion. *Syntax and Semantics* 9.

Stalnaker, Robert. 2002. Common ground. *Linguistic and Philosophy* .

Stalnaker, Robert C. 1981. A defense of conditional excluded middle. In William Harper, Robert C. Stalnaker & Glenn Pearce (eds.), *Ifs*, 87–104. Reidel.

Tian, Ye & Richard Breheny. 2019. Pragmatics and negative sentence processing. In Chris Cummins & Napoleon Katsos (eds.), *The oxford handbook of experimental semantics and pragmatics*, 195–207. Oxford University Press.

Tian, Ye, Heather Ferguson & Richard Breheny. 2016. Processing negation without context – why and when we represent the positive argument. *Language, Cognition and Neuroscience* 31(5). 683–698. http://dx.doi.org/10.1080/23273798.2016.1140214. https://doi.org/10.1080/23273798.2016.1140214.

Trinh, Tue & Andreas Haida. 2015. Constraining the derivation of alternatives. *Natural Language Semantics* 23. 249–270.

Veltman, Frank. 1976. Prejudices, presuppositions, and the theory of counterfactuals. In *Amsterdam papers in formal grammar. proceedings of the 1st amsterdam colloquium*, 248–281. University of Amsterdam.

Veltman, Frank. 1985. *Logics for conditionals*: University of Amsterdam dissertation.

Veltman, Frank. 1996. Defaults in update semantics. *Journal of Philosophical Logic* 25(3). 221–261.

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

von Fintel, Kai. 2004. 'would you believe it? the king of france is back! presuppositions and truth-value intuitions". *In Descriptions and Beyond, edited by Marga*

Reimer and Anne Bezuidenhout. Oxford University Press. .
Willer, Malte. 2018. Simplifying with free choice. *Topoi* 37(3). 379–392. http://dx.doi.org/10.1007/s11245-016-9437-5.
Wittenberg, Eva & Roger Levy. 2017. If you want a quick kiss, make it count: How choice of syntactic construction affects event construal. *Journal of memory and language* 94. 254–271.