Semantics & Pragmatics Volume 14, Article 13, 2021
https://doi.org/10.3765/sp.14.13

This is an EARLY ACCESS version of

Marty, Paul, Jacopo Romoli, Yasutada Sudo & Richard Breheny. 2021. Negative free choice. *Semantics and Pragmatics* 14(13). https://doi.org/10.3765/sp.14.13.

This version will be replaced with the final typeset version in due course. Note that page numbers will change, so cite with caution.

Negative Free Choice*

Paul Marty
University College London
p.marty@ucl.ac.uk

Jacopo Romoli
University of Bergen
jacopo.romoli@uib.no

Yasutada Sudo
University College London
y.sudo@ucl.ac.uk

Richard Breheny
University College London
r.breheny@ucl.ac.uk

Abstract Free Choice (FC) is an inference arising from the interaction between existential modals and disjunction. Schematically, a sentence of the form *permitted(A or B)* gives rise to the inference $\Diamond A \wedge \Diamond B$. Many competing theories of FC have been proposed but they can be classified into two main groups: one group derives FC as an entailment, while the other derives it as an implicature. By contrast, Negative Free Choice (NFC), the corresponding inference from negated universal modals embedding conjunction, e.g., *not(required(A and B))* to $\neg\Box A \wedge \neg\Box B$, has been discussed much less, and its existence has even been questioned in the recent literature. This paper reports on three experiments whose results provide clear evidence that NFC exists as an inference, but also indicate that NFC is far less robust than FC. This leaves us with two theoretical possibilities: the *uniform approach*, which comes in two versions, one deriving both FC and NFC as implicatures, and the other deriving both as entailments, and the *hybrid approach* that derives FC as an entailment and NFC as an implicature. We argue that the observed difference between FC and NFC is straightforwardly explained under the hybrid approach while it poses a challenge for the uniform approach. We end with a brief discussion of the options we see for the uniform approach and their further consequences.

Keywords: free choice, negative free choice, negation, alternatives, implicature

1 Introduction

Free Choice (FC) is an inference arising from the interaction between existential modals and disjunction. For instance, a sentence of the form *permitted(A or B)* gives rise to the

* For very helpful discussion and feedback, we would like to thank Moysh Bar-Lev, Danny Fox, Simon Goldstein, Matt Mandelkern, Guillermo del Pinal, Paolo Santorio, the editor Malte Willer and three anonymous reviewers for this journal, as well as audiences at SALT31, University of Potsdam, University of Tübingen, and Harvard University. This research was supported by the Leverhulme Trust grant RPG-2018-425. Materials, data and scripts of analyses are available open access on the OSF: https://osf.io/3942z/.

inference $\Diamond A \wedge \Diamond B$, as illustrated by (1).

(1) It is permitted that Mia buys apples or bananas.
$\rightsquigarrow$ *It is permitted that Mia buys apples and it is permitted that Mia buys bananas*

FC is a long-standing puzzle for compositional semantics, since it does not follow from the otherwise well motivated semantics of the possibility modal *permitted* as an existential quantifier, and the disjunctive connective *or* as $\vee$, according to which the literal meaning of (1) is simply $\Diamond A \vee \Diamond B$ (von Wright 1968, Kamp 1974). An additional difficulty of the puzzle comes from the fact that the negation of (1) behaves classically, as illustrated by (2). That is, *not(permitted(A or B))* means $\neg(\Diamond A \vee \Diamond B)$ (a.k.a. Double Prohibition), as expected from the standard semantics.

(2) It is not permitted that Mia buys apples or bananas.

The current literature is replete with competing theories of FC. For our purposes, these theories can be classified into two major strands (see Meyer 2016 for a useful overview).[1]

- Approaches that derive FC as an entailment, by abandoning the classical semantics of possibility modals and disjunction (Zimmerman 2000, Aloni 2003, Simons 2005, Aloni 2007, Willer 2017, Aloni 2018, Rothschild & Yablo 2018, Goldstein 2019). This view generally needs an account of Double Prohibition, and different ideas have been put forward.
- Approaches based on implicature which, on the other hand, keep the classical semantics of possibility modals and disjunction, and account for FC as an implicature (Kratzer & Shimoyama 2002, Alonso-Ovalle 2006, Fox 2007, Klinedinst 2007, Chemla 2009a, Franke 2011, Santorio & Romoli 2017, Bar-Lev 2018, Bar-Lev & Fox 2020). On this view, Double Prohibition falls out naturally, given that implicatures are generally not derived in the scope of negation.

In contrast to FC, what we call Negative Free Choice (NFC) is less often discussed. NFC is an inference from a sentence of the form *not(required(A and B))* to $\neg\Box A \wedge \neg\Box B$, as illustrated by (3).

(3) It is not required that Mia buys apples and bananas.
$\rightsquigarrow$ *It is not required that Mia buys apples and it is not required that Mia buys bananas*

1 Kamp (1974) proposes a performative theory, the scope of which remains limited. In particular, it does not generalise to non-performative cases or to non-deontic cases. Properly testing this approach would require comparing cases involving deontic modals with cases involving other modal bases. Given that our current study only looks at deontic modals, we will not discuss Kamp's (1974) approach further.

Like FC, NFC does not follow from the classical semantics of a necessity modal as a universal quantifier and conjunction as $\wedge$. Also, note that, without negation, the sentence has a meaning that is expected from the classical semantics of modals, as illustrated by (4).

(4) It is required that Mia buys apples and bananas.

In the current literature, the empirical status of NFC is debated. Based on introspective judgments, Fox (2007) claims that this inference exists, while Ciardelli et al. (2018: 615) and Goldstein (2019: 8) explicitly deny its existence.[2]

Chemla (2009b) presents quantitative data indicative of the existence of NFC, but it remains inconclusive. Specifically, Chemla's experiment with an inferential task yielded results where FC was endorsed about 90% of the time, while NFC was endorsed about 60% of the time. These results are good evidence that FC is more robust than NFC, but they are arguably not compelling with respect to whether or not NFC is in fact available, as this experiment lacks a baseline for unavailable inferences. As Chemla (2009b: 12) himself notes, absolute values in the results of inferential tasks are not easily interpretable, and it cannot be concluded that 60% is incompatible with the absence of an inference.[3] In addition, as we shall later see in detail, the critical sentences in Chemla's experiment contain a potential confound arising from a plural reading of the relevant conjoined DPs, which is associated to a homogeneity condition, which in turn might have inflated the endorsement rate of NFC (Szabolcsi & Haddican 2004, Breheny 2005, Magri 2014, among others).

Against this background, this paper reports on three experiments that use sentence-picture acceptability tasks. We claim that the results of these experiments provide clear evidence for the existence of NFC as an inference. Crucially, the experiments included various baseline conditions for unavailable inferences, unlike Chemla 2009b. They also replicated Chemla's observation that NFC is less robust than FC, despite the difference in the task.

Specifically, Experiment 1 tested sentences like (1) and (3), and provided initial results that NFC exists along with FC, but is less commonly observed than FC. Experiment 2 tested versions of sentences like (3) that used complex conjunctions of the form *both A and B*, rather than simple conjunctions of the form *A and B*, and correspondingly versions of sentences like (1) with complex disjunctions of the form *either A or B*. We made this change to eliminate the potential confounding factor arising from interpreting simple

2 We should note that the claim by Ciardelli et al. (2018: 615) is based on examples like (i), which involve the configuration $\Diamond\neg(A \wedge B)$, whereas we tested the configuration $\neg\Box(A \wedge B)$ in our study. While the two are classically equivalent, we cannot exclude the possibility that there is a difference between them in terms of the availability of the NFC inference. In addition, while most accounts make the same predictions for the two configurations, not all do. Extending the investigation in this paper to cases like (i) would be thus a natural and important next step.

(i) Mary might not speak both Arabic and Bengali.

3 See also Chemla 2009a, Geurts & Pouscoulous 2009, Gotzner & Romoli 2018 for relevant discussion.

conjoined DPs as plural and the associated homogeneity, which indeed had a sizable effect. Importantly, the results still provide evidence for the existence of NFC, and give further evidence that FC is more robust than NFC. Finally, Experiment 3 compared FC/NFC with Direct Scalar Implicatures (DSI) like (5) and Indirect Scalar Implicatures (ISI) like (6). In the results, DSIs are observed about one third of the time, and ISIs about half of the time, which is consistent with what is known about these inferences in the experimental literature. As we discuss, however, the fact that ISIs were derived more often than DSIs raised a challenge for certain accounts of FC and NFC.

(5) It is permitted that Mia buys bananas.
⇝ *It is not required that Mia buys bananas*

(6) It is not required that Mia buys bananas.
⇝ *it is permitted that Mia buys bananas*

We argue that, when taken together, these results have important theoretical implications, raising issues for pure entailment approaches and pure scalar implicature approaches. Conversely, the results are in line with the hybrid approach we outline below, which combines an entailment account of FC and an implicature approach to NFC.

The rest of the paper is organised as follows. Sections 2–4 present three sentence-picture acceptability experiments, the results of which establish that NFC is an available inference but is far less robust than FC. Next, Section 5 offers a critical discussion of the different theoretical possibilities in light of the experimental results. Finally, Section 6 concludes.

2 Experiment 1: Probing for NFC

All three experiments reported in this paper used a sentence-picture acceptability task. In Experiment 1, participants were presented with sentence-picture items like the two examples in Figure 1, and had to decide whether the sentence was a good description of the situation depicted in the picture. Participants reported their judgement by clicking one of two response buttons, labelled "Good" and "Not good" respectively. Our expectation was that, by using these general labels, we would be able to capture judgments not only about truth and falsity, but more generally about pragmatic adequacy. Specifically, we reasoned that participants would use the "Not good" response button to reject a sentence if it is literally false, or if it is literally true, but odd due to additional inferences.

In the target trials, sentences like (7) were paired with pictures that make them inappropriate if FC/NFC is accessed, but appropriate if not, as in the examples given in Figure 1.

(7) a. It is permitted that Mia buys the popsicle or the fries. FC
b. It is not required that Sam buys the popsicle and the avocado NFC

Figure 1 Examples of sentence-picture items used in the test trials of Experiment 1. These examples correspond to critical trials for FC sentences (on the left) and NFC sentences (on the right). Participants were instructed to interpret the green circle as *allowed but not required*, the red circle as *not allowed*, and the black square as *required* (see subsection 2.2 and Appendix A.1).

Before going on, two remarks are in order about sentences like (7b). Firstly, sentences of the form *not(required(A and B))* may have an independent implicature that $\Diamond A \wedge \Diamond B$. In the critical pictures for NFC, we made sure that this inference would be true while the putative NFC inference would be false, so that a 'Not good' response would be due to NFC (cf. Chemla 2009b).[4] Secondly, note that if the conjunction could have wide scope over the negation, the sentence would then entail NFC. For this reason, we decided to use a finite clause rather than an infinitival complement, so as to create a scope island for the conjunction (we will deal with the potential issue of homogeneity in Experiment 2).

Building upon the method in Marty et al. (2015), participants' responses in the target trials were compared to two sets of baselines: (i) responses to false and true picture controls, obtained by pairing the same sentences with different pictures where the truth value of the sentence doesn't change with or without FC/NFC, and (ii) responses to false and true sentence controls, obtained by pairing the same pictures with different sentences that were uncontroversially either true or false in the situations depicted by those pictures. By including these baselines in the experiment, we were able to verify that the oddness of FC and NFC sentences in their target conditions can be specifically attributed to the corresponding FC/NFC inferences being not true in these conditions.

4 This implicature arises from alternatives like the one in (i), the negation of which entails that it is permitted that Sam buys the popsicle and that he buys the avocado.

(i) It is not permitted that Sam buys the popsicle and the avocado.

2.1 Participants

51 participants, all located in the United Kingdom, were recruited using Prolific and were paid £2.00 for their participation. Of these, 5 were screened out prior to the test trials because their performance in the pre-test phase did not reach the threshold of 30% accuracy we had pre-established, and 1 was further removed prior to analyses because they did not declare English as their native language. The data of the remaining 45 were used in the analyses (33 female, average age 35.4 years). All participants gave written informed consent to the processing of their personal information for the purposes of this study. All data were collected and stored in accordance with the provisions of Data Protection Act 2018, the UK's implementation of the General Data Protection Regulation.

2.2 Materials

Each trial involved a sentence presented below a picture (see Figure 1). For the test trials, pictures were constructed using one of the five frames given and exemplified in Table 1. Every picture displayed two different food items from the following list of 18 food items: *popsicle, pear, broccoli, cookie, pineapple, croissant, burger, hot-dog, cupcake, fries, pizza, donut, banana, carrot, apple, avocado, orange, lettuce.* Each food item was enclosed within one of three symbols: a green circle, a red circle with a red diagonal line through it, and a black square. Participants were instructed to interpret those symbols as representing specific rules concerning what two children, Mia and Sam, are *allowed but not required to buy* (green circle), *not allowed to buy* (red circle) and *required to buy* (black square) at the supermarket, according to their parents (see Appendix A.1).

All sentences were constructed in reference to the pictures they were paired with. In the test trials, sentences were constructed using one of the six frames given in Table 2. The [name] term was the name of one of our two characters, *Mia* or *Sam.* The [L] and [R] terms were food names corresponding to those of the food items displayed on the *left* and on the *right* of the picture, respectively. For sentence types involving only one food term, the food name used in the sentence was pseudo-randomly assigned to match one of the two food items displayed on the picture (i.e., the one on the left or the one on the right).

In addition to the test trials, we included pre-test trials at the start of the experiment. Those trials were designed to assess whether participants understood correctly how to interpret, in isolation, the three symbols enclosing the food items before continuing to the test trials. Pictures in those trials involved only one food item, enclosed within one of the three symbols, and were paired with sentences like *It is permitted/not permitted that Mia buys the pear* or *It is required/not required that Mia buys the banana.* We refer the reader to Appendix B for a complete description of the pre-test trials and a discussion of their purpose in the study.

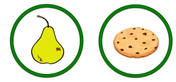
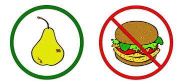
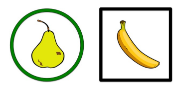
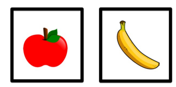
Label	Description of the picture type	Example picture
AA	permitted to buy both items	
AD	permitted to buy one item & not permitted to buy the other	
DD	not permitted to buy either items	
AR	permitted to buy one item & required to buy the other	
RR	required to buy both items	

Table 1 Description and examples of the pictures used in the test trials of Experiment 1. The position of the food items and of the symbols represented on the pictures (on the right vs. on the left) was randomized.

2.3 Design of the study

FC and NFC were the sentence types of primary interest. FC sentences such as (8) were investigated for their well-known ability to readily give rise to free choice inferences:

(8) It is permitted that Mia buys the pear or the burger. FC
$\rightsquigarrow$ *It is permitted that Mia buys the pear and it is permitted that Mia buys the burger* ✓

In the critical trials, these sentences were paired with AD pictures (e.g., pictures illustrating that Mia is Allowed to buy the pear, but Disallowed to buy the burger) that make them false if the free choice inference in (8) is present, but true if it is absent. In the control cases, they were paired with AA and DD pictures that make them respectively true and false, regardless of the presence or absence of free choice inferences. Using FC sentences as a reference point, sentences such as (9) were investigated to test whether they can yield similar conjunctive inferences, which we referred to as NFC:

(9) It is not required that Mia buys the pear and the banana. NFC
$\rightsquigarrow$ *It is not required that Mia buys the pear and it is not required that Mia buys the banana* ?

In a way parallel to FC sentences, these sentences were paired in the critical trials with AR pictures (e.g., pictures illustrating that Mia is Allowed to buy the pear and Required to buy the banana) that make them false if the free choice inference in (9) is present, but true if it is absent. In the control cases, they were paired with AA and RR pictures that make

Label	Description of the sentence type
FC	It is permitted that [name] buys the [L] or the [R].
C1	It is not permitted that [name] buys the [L] or the [R].
C2	It is permitted that [name] buys the [L/R].
NFC	It is not required that [name] buys the [L] and the [R].
C3	It is required that [name] buys the [L] and the [R].
C4	It is not required that [name] buys the [L/R].

Table 2 Schematic description of the sentences used in the test trials of Experiment 1, where [L] and [R] correspond to the names of the food items displayed respectively on the left and on the right of the pictures they were paired with.

them respectively true and false, regardless of the presence or absence of any putative free choice inference. We were interested in comparing participants' judgements to FC and NFC sentences in the critical trials relative to their false and true picture controls as well as relative to one another. C1, C2, C3 and C4 sentences, exemplified below, were added to the experiment to extend and refine our set of comparison points:

(10) a. It is not permitted that Mia buys the pear or the burger. C1
b. It is permitted that Mia buys the pear. C2
c. It is required that Mia buys the pear and the banana. C3
d. It is not required that Mia buys the pear. C4

The C1 and C2 sentences were paired with the same picture types as the FC sentences (i.e., AA, AD and DD) while the C3 and C4 sentences were paired with the same picture types as the NFC sentences (i.e., AA, AR and RR). Those sentence-picture combinations are useful on their own to evaluate whether participants show any difficulties in understanding the building blocks of the pictures used in the test trials. In addition, they provide us with another type of controls for our target sentences. In the critical trials for the FC sentences (i.e., when paired with AD pictures), the C1 sentences are false while the C2 sentences are true; similarly, in the critical trials for the NFC sentences (i.e., when paired with AR pictures), the C3 sentences are false while the C4 sentences are true. Thus, those items give us a second kind of baselines for our set of comparisons in adding false and true sentence controls (i.e., same picture with different sentences) to the previous true and false picture controls (i.e., same sentence with different pictures). Crossing pictures and sentences, we obtained the set of experimental conditions (in bold font) represented in Table 3, where the letters *L* and *R* are used here as cover terms to refer respectively to the food item on the left and the one on the right of the example pictures.[5]

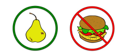
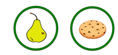
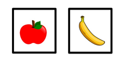
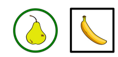
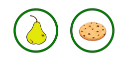
		DD	AD	AA
				
FC	It is permitted that Mia buys the *L* or the *R*.	**False** (4)	**Target** (10)	**True** (4)
C1	It is not permitted that Mia buys the *L* or the *R*.	**True** (4)	**False** (4)	**False** (4)
C2	It is permitted that Mia buys the *L*.	**False** (4)	**True** (4)	**True** (4)
		RR	AR	AA
				
NFC	It is not required that Mia buys the *L* and the *R*.	**False** (4)	**Target** (10)	**True** (4)
C3	It is required that Mia buys the *L* and the *R*.	**True** (4)	**False** (4)	**False** (4)
C4	It is not required that Mia buys the *L*.	**False** (4)	**True** (4)	**True** (4)

Table 3 Summary of the combinations giving rise to the **False**, **True** and **Target** conditions in Experiment 1 (*L* refers to the food item on the left and *R* to the one on the left for those examples). Numbers in parenthesis refer to the number of items included in the study to exemplify the different conditions.

2.4 Sets of trials and randomisation

To diversify the content of the pictures and sentences presented to participants, the study ran 8 sets of trials, all generated in Python using the same template. Every set used all and the same 18 food items from the list in 2.2. In each set, the food items were evenly distributed among both characters and evenly combined with the three rule-related symbols by pseudo-randomly assigning to each character 3 food items that he was permitted to buy, 3 that he was not permitted to buy and 3 that he was required to buy; as a result, each food item was consistently associated with the same rule within a given set of trials. Pictures and sentences were generated for each set on the basis of those assignments. For

5 We note that two of the sentence-picture combinations categorized here as **True** conditions, namely C2-AA and C4-AA, could in principle be rejected by participants due to *ad-hoc* enrichments. For instance, a sentence like *It is permitted that Mia buys the pear* can be enriched on the basis of a contextually determined scale like ⟨*pear, cookie*⟩ to further convey that *it is not permitted that Mia buys the cookie*, in which case it should be rejected rather than accepted when paired with AA pictures. Those considerations are orthogonal however to our research purposes since, in our design, C2-AA and C4-AA items are not among the controls used as baselines for analyzing the FC and NFC sentences. As we report on below, we found in fact no evidence for such enrichments in our study: C2-AA and C4-AA items were accepted by participants more than 95% of the time.

the test trials, pictures were generated using the frames in Table 1 by randomly selecting two food items with the appropriate symbols and randomly determining their ordering on the picture. Sentences for those trials were then generated on the basis of the picture they were paired with using the frames in Table 2 so that the linear order of the food terms in the sentence matches their (left-to-right) ordering on the picture. For sentence types involving only one food term, the food name used in the sentence was pseudo-randomly assigned to match one of the two food items displayed on the picture in accordance with the experimental condition. These steps were repeated to create every instance of the experimental conditions in Table 3. Sentence-picture items for the pre-test trials were generated through a similar process: pictures in those trials were generated by randomly selecting a food item with the appropriate symbol and sentences were then generated by using the food name corresponding to the food item displayed on the picture. These steps were repeated to create every instance of the combinations described in Table 10 in Appendix B. The result was a set of 16 pre-test and 84 test trials designed so that the proportion of expected 'Good' and 'Not Good' responses was well-balanced for both kinds of trials. The process was repeated 8 times to generate the 8 sets of trials that were used in the study.

2.5 Procedure

The experiment was run as an online survey using Gorilla Experiment Builder. Participants were introduced in the instructions to two children, Sam and Mia, and were told that they would see pictures representing by means of different symbols the parental rules concerning what Sam and Mia are allowed to buy, not allowed to buy and required to buy at the supermarket (see Appendix A.1 for the instructions). Participants were shown three examples of enclosed food items, one for each symbol, together with their intended interpretation. They were told that each picture in the study will be accompanied with a sentence that relates to it and they were instructed to click on 'Good' if they consider that sentence a good description of the picture they see and otherwise to click on 'Not good'.

After the instructions, each participant was assigned a set of trials. Each set was pseudo-randomly assigned so as to reach an equal number of participants per set. For each set, the position of the 'Good' and 'Not good' response buttons on the screen was counterbalanced across participants. The experiment started with a block of 16 pre-test trials and then continued with a block of 84 test trials.[6] In each block, trials were presented in random order, with a 1000 ms interstimulus interval, and a self-timed break between both blocks. For each trial, participants were asked to decide whether the sentence was a correct description of the picture. Participants gave their answers by clicking one of both response buttons. Items remained on the screen until participants gave their answer.

6 The survey was set up so that a participant making errors in more than 30% of the pre-test trials could not continue to the test trials. 5 participants were screened-out that way. The mean accuracy rate of the remaining participants was 92.5% (95% CI [94, 91]). See Appendix B for discussion.

2.6 Results

We analysed the data by modelling response-type likelihood using logit mixed-effects regression models (Jaeger 2008). Analyses were conducted using the `lme4` (Bates et al. 2011) and languageR libraries for the R statistics program (R Development Core Team 2013). β value, standard errors, z-values and p-values are shown in the tables accompanying the relevant analyses with the R pseudo-code describing the models they were obtained from.

2.6.1 Control sentences

Responses to the control sentences were as expected: participants strongly rejected these sentences in the **False** conditions (all mean rejection rates $\geq$ 90%) and uniformly accepted them in the **True** conditions (all rejection rates $\leq$ 11%) with an overall mean accuracy reaching 93% (95% CI [94, 92]). These results show that participants performed the task appropriately and, crucially, that they had no difficulty in mapping each symbol to its intended meaning and no difficulty in interpreting the various combinations of those symbols across picture types.

2.6.2 Target sentences

Figure 2 shows the mean rates of rejection for the target sentences by condition (i.e., by picture type). In the following, we report two sets of analyses, the statistical details of which are given in Appendix C .

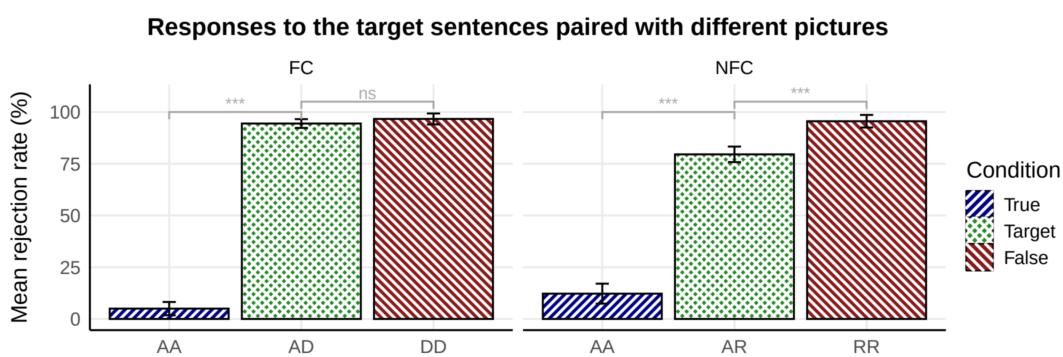


Figure 2 Mean rejection rate (in %) for the target FC and NFC sentences in Experiment 1 as a function of the condition (i.e., by picture type). FC and NFC sentences were more rejected in their **Target** conditions than in their **True** picture controls. Error bars represent 95% CIs.

The first set of analyses assessed whether responses to FC and NFC sentences in their **Target** conditions differ from those in their corresponding **True** and **False** conditions, i.e., in conditions where these sentences were paired with their true and false picture controls. For each sentence type, the model included Condition as a fixed effect (2 levels: **Target** vs. **False**, or **Target** vs. **True**), a random effect for subject and a random slope for Condition per subject wherever appropriate.[7] For FC sentences, the mean rejection rate in the **Target** condition was very high ($M = 94\%$, 95% CI [96, 92]), much higher than in the **True** condition ($M = 5\%$, 95% CI [8, 1], $\beta = -14.74$, $p < .001$) and roughly the same as in the **False** condition ($M = 96\%$, 95% CI [99, 94], $\beta = 0.65$, *ns*). For NFC sentences, the mean rejection rate in the **Target** condition was slightly lower ($M = 79\%$, 95% CI [83, 75]), falling between the one obtained in the **True** condition ($M = 12\%$, 95% CI [17, 7], $\beta = -6.41$, $p < .001$) and the **False** condition ($M = 95\%$, 95% CI [98, 94], $\beta = 2.54$, $p < .001$).

To evaluate further the differences between FC and NFC sentences, we examined the effects of sentence type and condition on participants' responses to these sentences in the **Target** vs. **False** conditions. The model included Sentence (2 levels: FC, NFC), Condition (2 levels: **Target**, **False**) and their interaction as fixed effects, a random effect for subject and a random slope for Condition per subject.[8] The model showed a main effect of sentence (FC>NFC, $\beta = -1.86$, $p < .001$) as well as an interaction between Sentence and Condition ($\beta = 1.56$, $p < .05$) such that the mean rejection rate for FC sentences was comparatively greater than the one for NFC sentences in the **Target** conditions.

In sum, NFC sentences received a relatively high rate of rejection in the **Target** conditions in comparison to their true and false picture controls, albeit slightly lower than FC sentences in the same conditions. Those results provide preliminary support for the view that NFC sentences yield free choice inferences, which made them inappropriate descriptions in the **Target** condition, just like FC sentences.

2.6.3 Comparison between Control and Target sentences

Figure 3 shows the mean rates of rejection for FC and NFC sentences in the **Target** conditions along with our second set of baselines: the mean rates of rejection for their corresponding true and false sentence controls (i.e., control sentences paired with the same pictures as FC and NFC in the **Target** conditions).

We carried out the exact same analyses as above by replacing in our models the previous picture controls with the new sentence controls (see Table 12 in Appendix C for details). The patterns of results for those novel comparisons were in line with those reported above. When paired with the target AD pictures, the mean rejection rate for FC sentences was much higher than that for C2 sentences ($M = 3\%$, 95% CI [6, 1], $\beta = -18.25$, $p < .001$) but similar to that for C1 sentences ($M = 90\%$, 95% CI [94, 85], $\beta = -1.28$, *ns*). When paired with the target AR pictures, the mean rejection rate for NFC sentences fell between

7 For the **Target** vs. **False** comparisons, the models including a maximal random effect structure resulted in singular fits. To allow non-singular fits, the random slope for Condition per subject were removed from those models.

8 Models including Sentence in the random effect structure failed to reach convergence.

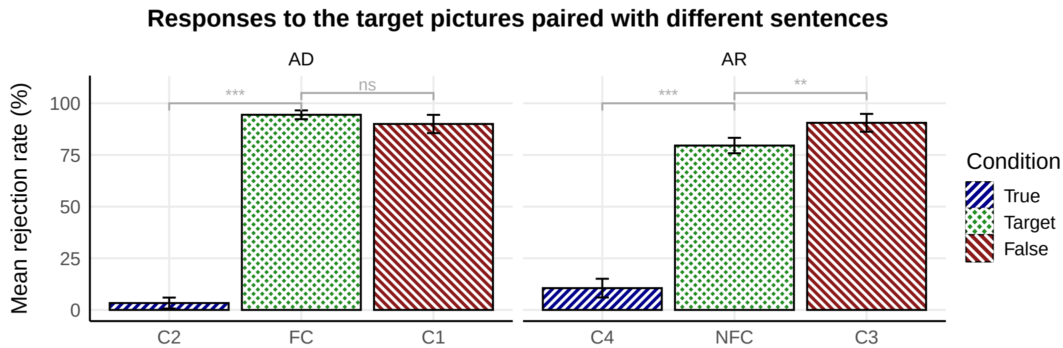


Figure 3 Mean rejection rate (in %) for the target and control sentences of Experiment 1 when paired with the target pictures (i.e., AR and AD) as a function of the condition (i.e., by sentence type). FC and NFC sentences were more rejected than their **True** sentence controls when paired with the target pictures. Error bars represent 95% CIs.

that for C4 sentences ($M = 10\%$, 95% CI [15, 6], $\beta = -7.42$, $p < .001$) and that for C3 sentences ($M = 90\%$, 95% CI [94, 86], $\beta = 5.53$, $p < .001$). In a way similar as above, we further examined the effects of sentence type and condition on participants' responses to AD and AR pictures. The model included Picture (2 levels: AR, AD), Condition (2 levels: **Target**, **False**) and their interaction as fixed effects, a random effect for subject and a random slope for Picture per subject.[9] The model showed a main effect of Picture (AR $<$ AD, $\beta = -1.83$), a main effect of Condition (Target $<$ False, $\beta = -0.79$, $p < .05$) , and an interaction between Picture and Condition ($\beta = -2.13$, $p < .001$) showing that the mean rejection rate for FC sentences was comparatively greater than the one for NFC sentences in the **Target** conditions, just like with the previous baselines.

This second set of comparisons confirm to a large extent the results from the first set: NFC sentences also received a relatively high rate of rejection in the **Target** conditions in comparison to their true and false sentence controls. Crucially, those findings rule out the possibility that the patterns of responses in our **Target** conditions be accounted for only in terms of the pictures used in those conditions, that is independently of the FC and NFC sentence they were paired with. Finally, the present results show that the differences previously noted between FC and NFC sentences reproduce using other baselines.

2.7 Discussion

The results of Experiment 1 offer empirical evidence that NFC sentences give rise to free choice inferences while suggesting that these inferences are less readily available for these

9 All models including Condition in the random effect structure failed to reach convergence.

negative sentences than for their positive counterparts. As we already mentioned, however, there is in fact another factor that could explain our findings for NFC sentences and that pertains to the homogeneity effect of conjoined DPs when interpreted as a plural (Szabolcsi & Haddican 2004, Breheny 2005, Magri 2014 among others).[10]

In a nutshell, it has been observed that simple conjunctions under negation like *Mia didn't buy the banana and the carrot* can have a reading entailing that Mia didn't buy the banana and she didn't buy the carrot. This reading however goes away if *both* is added to the sentence (or when *and* is stressed), as in *Mia didn't buy both the banana and the carrot.* A standard way to capture this reading goes as follows (see Szabolcsi & Haddican 2004) among others): the conjunction *the banana and the carrot* can refer to the plural individual constituted by the banana and the carrot. In turn, this plural is associated with a homogeneity condition when it combines with non-collective predicates, according to which either all of the members of the plural satisfies the predicate or none of them does.[11] In the case above, if Mia didn't buy the plurality constituted by the banana and the carrot, then, given homogeneity, she didn't buy either of its members i.e. she didn't buy the banana and she didn't buy the carrot. Finally, the addition of *both* would prevent interpreting the conjoined DPs as a plural individual, thereby blocking the reading of the sentence just described.

What is most relevant for us is that this homogeneity effect – which we call PLURAL HOMOGENEITY to distinguish it from the homogeneity arising from modals assumed by some accounts of FC we discuss below – could indirectly give rise to what we have called negative free choice and explain the results for NFC sentences without the need for a theory of free choice. To illustrate, consider one of our NFC sentences again in (11). If the conjoined DPs are read as a plural, the embedded sentence will give rise to a homogeneity inference of the form $A \leftrightarrow B$, i.e. Mia buys the banana if and only if she buys the carrot. This inference, will then project into the universal modal. While different accounts of homogeneity vary in their predictions about projection (see Križ 2015, Križ & Chemla 2015 for discussion), most accounts would predict universal projection in case of a universal modal. Hence the resulting inference would be that it's required that Mia buys the banana if and only if she buys the carrot, $\Box(A \leftrightarrow B)$.

(11) It is not required that Mia buys the banana and the carrot.

As a result, NFC would be now entailed: if it is not true that Mia is required to buy bananas and carrots and if she is required to buy one if and only if she buys the other, then it follows that she is not required to buy one and that she is not required to buy the other.[12]

10 We thank an anonymous reviewer for helping us clarify this point.

11 The source of the homogeneity of plurals remains controversial. Some proposal captures it as a presupposition (Gajewski 2005), others as derived from contextual restriction (Breheny 2005), truth-value gaps (Križ 2015), or implicatures (Magri 2014, Bar-Lev 2018).

12 More schematically, $\neg\Box(A \wedge B) \wedge \Box(A \leftrightarrow B)$ entails that $\neg\Box A$ and $\neg\Box B$. To illustrate, assume that $\neg\Box A$ is false instead: given the second conjunct it would mean that $\neg\Box B$ would have to be false as

In order to control for this potential confound, we ran a follow-up experiment to Experiment 1 in which we added *both* to our original NFC sentences, which, as discussed, is not compatible with the homogeneity reading of conjunction thereby blocking the other potential route to negative free choice described above. We turn to this second experiment in the next section.

3 Experiment 2: Controlling for the homogeneity of conjunction

The materials and method used in Experiment 2 were the same as in Experiment 1 with only one exception: the sentence types used in Experiment 1 were minimally altered so as to control for the possible effect of conjunction homogeneity on participants' rejection of NFC sentences. Specifically, Experiment 2 tested novel NFC sentences like the one in (12) which involve complex conjunctions of the form *both A and B*, where the presence of *both* permits us to remove the homogeneity of conjunction.

(12) It is not required that Mia buys both the pear and the banana. NFC-both
$\rightsquigarrow$ *It is not required that Mia buys the pear and it is not required that Mia buys the banana* ?

Just like the NFC sentences in Experiment 1, the novel NFC-both sentences were paired in the target trials with pictures that make them false if the free choice inference in (12) is present, but true if it is absent. Thus, if negative free choice inferences can be generated independently of the homogeneity of conjunction, participants should still reject these sentences to a large extent in the target trials.

3.1 Participants

50 participants, all located in the United Kingdom, were recruited using Prolific and were paid £2.00 for their participation. Of these, 1 was screened out prior to the test trials because their performance in the pre-test phase did not reach the pre-established threshold of 30% accuracy (the same as in Experiment 1, see Appendix B). The data of the remaining 49 were used in the analyses (36 female, average age 34.8 years). The consent and data collection procedures were the same as in Experiment 1.

3.2 Materials and Design

The materials used in Experiment 2 differed from the materials used in Experiment 1 along one unique dimension: the test trials of Experiment 2 involved novel sentences obtained by modifying the ones used in the test trials of Experiment 1. Specifically, the novel sentences were constructed using one of the six frames given in Table 4 (see Table 2 for comparisons).

well. But this would be in conflict with the first conjunct. Same for $\neg\Box$B.

Label	Description of the sentence type
FC-either	It is permitted that [name] buys either the [L] or the [R].
C1-either	It is not permitted that [name] buys either the [L] or the [R].
C2	It is permitted that [name] buys the [L/R].
NFC-both	It is not required that [name] buys both the [L] and the [R].
C3-both	It is required that [name] buys both the [L] and the [R].
C4	It is not required that [name] buys the [L/R].

Table 4 Schematic description of the sentences used in the test trials of Experiment 2, where [L] and [R] correspond to the names of the food items displayed respectively on the left and on the right of the pictures they were paired with.

Crucially, the NFC sentences from Experiment 1 were modified so as to involve complex conjunctions of the form *both A and B*, i.e., by adding the homogeneity remover *both* to the original sentences (e.g., *It is not required that Mia buys both the pear and the banana*). For the sake of parallelism, the FC sentences from Experiment 1 were modified along the same lines so as to involve similarly complex disjunctions of the form *either A or B* (e.g., *It is permitted that Mia buys either the pear or the burger*). We made this second modification to make the structures of our target sentences more similar to each other which, in turn, should prevent differences in complexity to affect our evaluation of the factors of interest. For completeness, the C1 and C3 control sentences from Experiment 1, which also involved embedded instances of conjunction or disjunction, were adjusted accordingly; the C2 and C4 control sentences, on the other hand, were left unchanged.

The rest of the design of Experiment 2 was identical to that of Experiment 1 in all respects. Thus, Table 3 also stands as a summary of the sentence-picture combinations giving rise to the **False**, **True** and **Target** conditions in Experiment 2 (including the number of items exemplifying each condition), provided the modifications of the sentence types we just described. The study ran 8 different sets of trials generated via the same program as the one used to generate the sets of trials in Experiment 1 (see Section 2.4 for details).

3.3 Procedure

The procedure was identical to the one used in Experiment 1 (see Section 2.5 for details).

3.4 Results

The data were analysed using the data analysis pipelines created to analyse the data from Experiment 1. The results from Experiments 1 & 2 are thus directly comparable.

3.4.1 Control sentences

Responses to the control sentences were very similar to those found in Experiment 1: participants strongly rejected these sentences in the **False** conditions (all mean rejection rates $\geq$ 83%) and uniformly accepted them in the **True** conditions (all rejection rates $\leq$ 13%) with an overall mean accuracy reaching 91% (95% CI [92, 90]).

3.4.2 Target sentences with their baselines

Figure 4 shows the mean rates of rejection for the FC-either and NFC-both sentences by condition (i.e., with their true and false picture controls), and Figure 5 shows the mean rates of rejection for these sentences in the **Target** conditions along with their corresponding true and false sentence controls. The details of the statistical analyses are given in Appendix C.

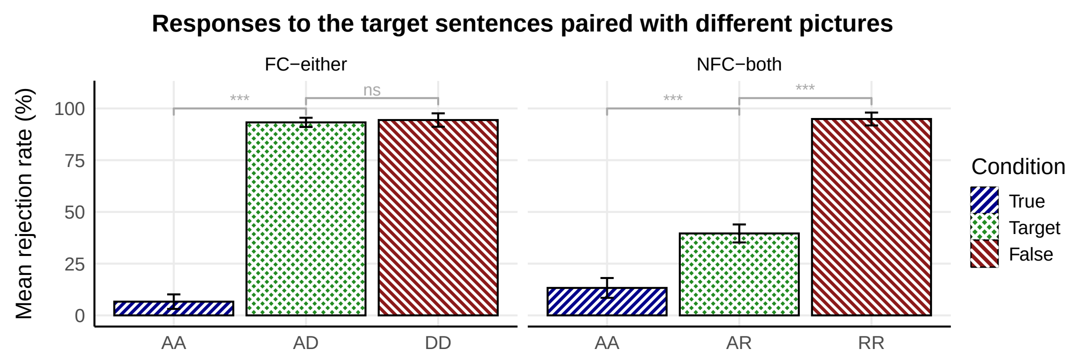


Figure 4 Mean rejection rate (in %) for the target sentences of Experiment 2 as a function of the condition (i.e., by picture type). FC and NFC sentences were more rejected in their **Target** conditions than in their **True** picture controls. Error bars represent 95% CIs.

The patterns of results for the FC-either sentences were similar to those found for the FC sentences. That is, the mean rejection rate for these sentences in the **Target** condition was very high ($M = 93\%$, 95% CI [95, 91]), much higher than that for their **True** picture and sentence controls (all Ms <7%, all $|\beta s| > 16$, all $ps < .001$), but no different from that for their **False** picture controls ($M = 94\%$, 95% CI [97, 91], $\beta = 0.25$, *ns*). Interestingly, in contrast to FC sentences, FC-either sentences were also found to be rejected more often than their **False** sentence controls (C1-either), which similarly involved complex disjunctions ($M = 83\%$, 95% CI [88, 77], $\beta = -4.36$, $p < .01$). These findings confirm those from Experiment 1 and further illustrate the robustness of FC inferences.

By contrast, the patterns of results for the NFC-both sentences showed some similarities, but also some differences with those found for the NFC sentences. Starting with the

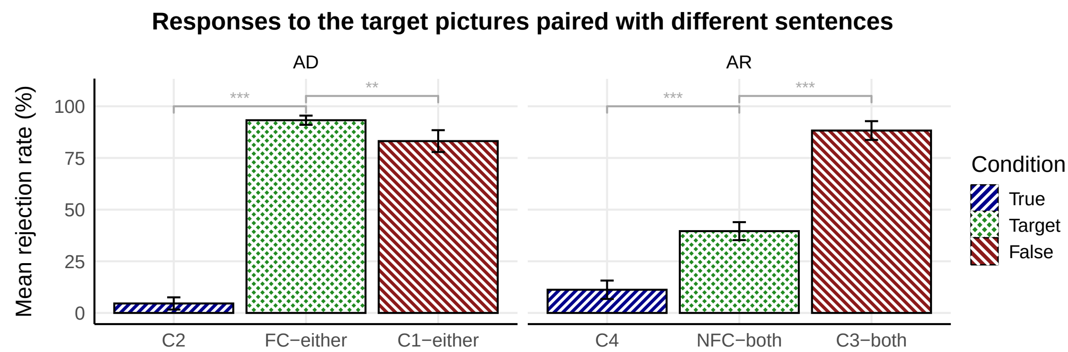


Figure 5 Mean rejection rate (in %) for the target and control sentences of Experiment 2 when paired with the target pictures (i.e., AR and AD) as a function of the condition (i.e., by sentence type). FC and NFC sentences were more rejected than their **True** sentence controls when paired with the target pictures. Error bars represent 95% CIs.

similarities, the mean rejection rate for NFC-both sentences also fell between that for their **True** picture and sentence controls (all Ms <14%, all $|\beta s| > 4.76$, all ps $< .01$) and that for their **False** picture and sentence controls (all Ms >88%, all $|\beta s| > 5.11$, all ps $< .001$). These findings confirm that NFC inferences exist and, crucially, they establish that these inferences can be generated independently of the homogeneity of conjunction.

However, the mean rejection rate for NFC-both sentences in the **Target** conditions was also found to be substantially lower than that for NFC sentences in these same conditions, with 39% of rejection (95% CI [43, 35]) contra 79% (95% CI [83, 75]) for NFC-sentences. Taken at face value, these results suggest that the elimination of conjunction homogeneity in Experiment 2 diminished by about 50% the rate of responses compatible with NFC-like inferences in Experiment 1 (or, conversely, that conjunction homogeneity boosted this type of responses in Experiment 1). Relatedly, these results reveal that, once the homogeneity of conjunction is factored out, free choice inferences are in fact considerably less available in their negative than in their positive form, as evidenced by the increased differences between the NFC sentences and their FC counterparts in Experiment 1 vs. Experiment 2.

3.5 Discussion

We designed Experiment 2 as a minimal variant of Experiment 1 with the aim of controlling for and quantifying over the effect of conjunction homogeneity on participants' responses to NFC sentences in Experiment 1. Overall, the results of this follow-up study confirm the main findings from Experiment 1: in line with what we found for NFC-sentences, the rate of rejection for NFC-both sentences in the target trials fell between that of their true and false baselines, indicating that NFC inferences are still detectable in the absence of conjunction homogeneity. We interpret these results as establishing that NFC inferences do exist independently of the homogeneity of conjunction.

Yet the present results also invite us to refine some of our previous observations in significant ways. Specifically, we found that, in comparison to the original NFC sentences, participants rejected far less often the novel NFC-both sentences in the target trials. These results suggest that the homogeneity of conjunction actually affected participants' responses to the original NFC sentences, essentially by boosting the proportion of NFC-like responses, and should be thus regarded as a major explanatory factor in accounting for the high rate of rejection observed for these sentences in Experiment 1. This first refinement subsequently leads to another one: once the additional effect of conjunction homogeneity is factored out, the discrepancy in robustness between FC and NFC inferences becomes clear-cut. In particular, unlike for NFC sentences, we found no evidence in either experiments that participants judged FC or FC-either sentences differently in their target and in their false conditions, with rates of rejection above 90% in the target conditions. At first sight, these rates are at odds with the rates of scalar implicatures commonly observed for other cases in the experimental literature.

Before getting to the theoretical consequences of these findings, there is a possible explanation for the discrepancy between FC and NFC that we would like to explore and which relates to current research on *scalar diversity* (Van Tiel et al. 2016 among others). Indeed, recent experimental works have found that scalar items belonging to different scales are associated with scalar inferences that can be more or less robust and, furthermore, that weak and strong scalar terms belonging to the same scale may be associated with direct and indirect scalar inferences to different degrees (see Marty et al. 2020 for an example involving the scale ⟨possible, certain⟩). Thus, it is possible that the differences we observed between FC and NFC sentences is just a reflection of scalar diversity. That is, it is possible that, in general, the use of the weak scalar term *permitted* more readily gives rise to scalar implicatures than its negated stronger scale-mate, *not required*. To investigate this possibility, we carried out a third experiment in which we compared FC and NFC inferences directly with the direct implicatures arising from *permitted* and with the indirect ones arising from *not required*.

4 Experiment 3: A comparison with direct and indirect scalar implicatures

Experiment 3 compared the robustness of the free choice inferences arising from FC and NFC sentences with that of the scalar implicatures arising from related sentences like those in (13) and (14). Specifically, positive sentences like (13) were investigated for their ability to give rise to direct scalar implicatures (i.e., the inference from *permitted* to *permitted but not required*), and negative sentences like (14) for their ability to give rise to so-called 'indirect' scalar implicatures (i.e., the inference from *not required* to *not required, but permitted*).

(13) It is permitted that you eat the banana. SI
$\rightsquigarrow$ *It is not required that you eat the banana*

(14) It is not required that you eat the burger. ISI
⇝ *It is permitted that you eat the burger*

Just like the FC and NFC sentences in the previous experiments, these sentences were paired in the critical trials with pictures that make them false if the relevant inferences are present, but true otherwise. We hypothesized that if the differences we observed between FC and NFC sentences are to be attributed to the use of *permitted* vs. *not required*, then similar differences should be observed between SI and ISI sentences, i.e., SI sentences should give rise to more scalar implicatures than ISI sentences.

4.1 Participants

55 participants, all located in the United Kingdom, were recruited using Prolific and were paid £1 for their participation. Of these, 5 were removed at the initial stage of the data analysis because their performance to the control trials did not reach the pre-established threshold of 30% accuracy. The data of the remaining 50 were used in the final analyses (25 female, average age 32.8 years). The consent and data collection procedures were the same as in Experiments 1 and 2.

4.2 Materials and Design

The materials for Experiment 3 were based on the materials for Experiments 1 and 2. Sentences were constructed using one of the 8 frames in Table 5.

Concretely, the sentence types included in Experiment 3 were the same as those included in the test or pre-test trials of Experiment 2, with some minor alterations (i.e., use of a 2nd person pronoun instead of a character's name and use of the verb *eat* instead of *buy*). Thus, in particular, the SI and ISI sentences were variants of the T1 and T4 sentences from the pre-test phase of Experiments 1 & 2 (see Table 10 in Appendix B). In this study, however, these sentences were tested for their ability to give rise to scalar implicatures in order to collect further reference points for our comparisons between FC and NFC. Thus, just like the other target sentences, these sentences were paired in the critical trials with pictures that make them false if the relevant scalar implicature is present, but true otherwise. In the control cases, they were paired with pictures that make them respectively true and false, regardless of the presence or absence of an implicature (as in the pre-test trials of Experiments 1 & 2). Everything else being equal, the other sentence types were paired with the same picture types as their counterparts in Experiments 1 and 2. Crossing pictures and sentences, we obtained the set of experimental conditions (in bold font) represented in Table 6. Each experimental condition were instantiated 3 times in the study by varying the contents of the pictures and sentences, giving rise to a total of 48 items.

Label	Description of the sentence type
FC	It is permitted that you eat either the [L] or the [R].
C1	It is not permitted that you eat either the [L] or the [R].
C2	It is permitted that you eat the [L/R].
NFC	It is not required that you eat both the [L] and the [R].
C3	It is required that you eat both the [L] and the [R].
C4	It is not required that you eat the [L/R].
SI	It is permitted that you eat the [food].
ISI	It is not required that you eat the [food].

Table 5 Schematic description of the target and control sentences tested in Experiment 3. For SI and ISI sentences, [food] corresponds to the name of the sole food item displayed on the pictures they were paired with. For the other sentence types, [L] and [R] correspond to the names of the food items displayed respectively on the left and on the right of the pictures they were paired with.

4.3 Procedure

The experiment was run on Ibex Farm, a hosted version of the Ibex software, created by Alex Drummond. Participants were first introduced to the general display and response procedure used in the study, and then invited to complete three example trials involving pictures and sentences unrelated to our experimental purposes. Next, participants were presented with the instructions specific to the study (see Appendix A.2). They were told that the study describes a new diet program and that they would see pictures representing by means of different symbols what they are allowed to eat, what they have to eat, and what they are not allowed to eat according to this program. To illustrate these rules, participants were shown examples of enclosed food items, one for each symbol, together with their intended interpretation. Crucially, they were told that the pictures would sometimes depict only one rule, sometimes two rules at once. The instructions were thus essentially the same as in Experiments 1 and 2 (see Appendix A.2 for details).

The experiment started with a block of 6 unannounced practice trials and then continued with a block of 48 test trials. The practice trials involved simple sentences of the form *It is required/not permitted that you eat the [food]* and were used to further illustrate the experimental display and procedure prior to the test trials. In each block, trials were presented in random order, with a 1000 ms interstimulus interval. All participants saw the same list of practice and test trials. As in Experiments 1 and 2, participants were asked for each trial to decide whether the sentence was a correct description of the picture. They gave their answers by pressing one of two response keys, labelled ‘Good’ and ’Not good’.

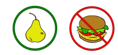
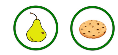
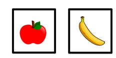
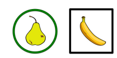
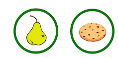
		DD	AD	AA
				
FC	It is permitted that you eat either the *L* or the *R*.	**False** (3)	**Target** (3)	**True** (3)
C1	It is not permitted that you eat either the *L* or the *R*.	**True** (3)	**False** (3)	**False** (3)
C2	It is permitted that you eat the *L*.	**False** (3)	**True** (3)	**True** (3)
		RR	AR	AA
				
NFC	It is not required that you eat both the *L* and the *R*.	**False** (3)	**Target** (3)	**True** (3)
C3	It is required that you buy eat the *L* and the *R*.	**True** (3)	**False** (3)	**False** (3)
C4	It is not required that you eat the *L*.	**False** (3)	**True** (3)	**True** (3)
		D	A	R
				
SI	It is permitted that you eat the *food*.	**False** (3)	**True** (3)	**Target** (3)
ISI	It is not required that you eat the *food*.	**Target** (3)	**True** (3)	**False** (3)

Table 6 Summary of the combinations giving rise to the **False**, **True** and **Target** conditions in Experiment 3 (for SI and ISI, *food* refers to the unique food item used in the picture; for the other sentence types, *L* refers to the food item on the left of the picture and *R* to the one on the left). Numbers in parenthesis refer to the number of items included in the study to exemplify the different conditions.

Items remained on the screen until participants gave their answer.

4.4 Results

The data were analysed using the data analysis pipelines from Experiments 1 and 2.

4.4.1 Control sentences

Responses to the control C1-C4 sentences were very similar to those found in Experiments 1 and 2: participants strongly rejected these sentences in the **False** conditions (all mean rejection rates $\geq$ 86%) while they uniformly accepted them in the **True** conditions (all rejection rates $\leq$ 10%) with an overall mean accuracy reaching 90% (95% CI [92, 87]).

4.4.2 Target sentences

Figure 6 shows the mean rates of rejection for the target SI, ISI, FC and NFC sentences by condition. The details of the statistical analyses reported in the following are given in Table 15 in Appendix C.

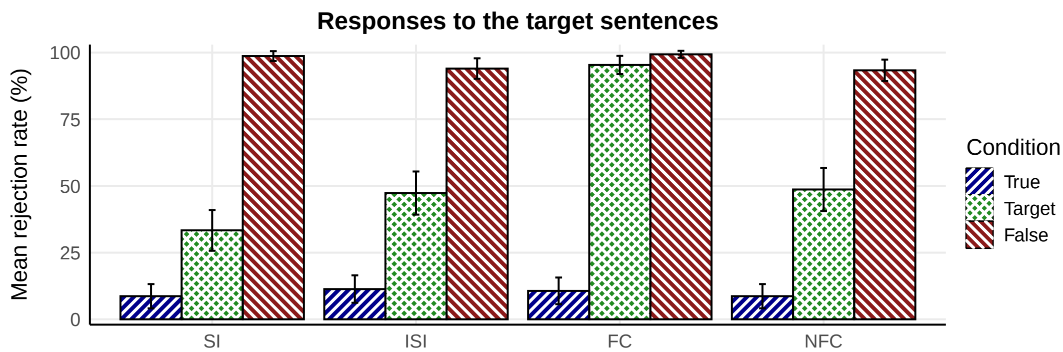


Figure 6 Mean rejection rate (in %) for the target sentences of Experiment 3 as a function of the condition. SI-related sentences gave rise to more SIs in their negative than in their positive forms (ISI>SI) while FC-related sentences gave rise to more FC inferences in their positive than in their negative forms (FC>NFC). Error bars represent 95% CIs.

The results for the FC and NFC sentences were in line with those observed for the corresponding sentences in Experiment 2. That is, for FC sentences, the rejection rate in the **Target** conditions ($M = 95\%$, 95% CI [98, 92]) was much higher than that in the **True** conditions but not significantly different from that in the **False** conditions; by contrast, for NFC sentences, the rejection rate in the **Target** conditions ($M = 48\%$, 95% CI [56, 40]) was in between those obtained in the **True** and **False** conditions. In that respect, the patterns of results for the SI and ISI sentences were found to be similar to those observed for the NFC sentences: these sentences also gave rise to intermediate rejection rates in the **Target** conditions, with a mean rejection of 33% (95% CI [40, 25]) for the SI sentences and a slightly higher rate of 47% (95% CI [55, 39]) for the ISI sentences. In sum, all target sentences except the FC sentences gave rise to intermediate rejection rates in their **Target** conditions.

To evaluate the differences between these four sentence types along the factors of interest, we examined the effects of inference type (scalar implicatures vs. free choice) and sentence polarity (positive vs. negative) on participants' responses in the **Target** conditions. The model included Inference type (2 levels: SI, FC), Polarity (2 levels: Positive, Negative) and their interaction as fixed effects, a random effect for subject, a random effect for item and a random slope for Inference per subject.[13] The model showed a mean effect of Inference

13 Models including a richer random effect structure either didn't converge or resulted in singular fits (random effect variance estimated near zero).

(FC>SI, $\beta = 5.06$, $p < .001$), a main effect of Polarity (Positive>Negative, $\beta = 0.93$, $p < .01$) and a significant interaction between both factors ($\beta = -4.90$, $p < .001$) such that the difference between the mean rejection rates for positive vs. negative FC sentences was greater than that between the mean rejection rates for positive vs. negative SI sentences.[14]. Taken together, these results show that FC-related sentences gave rise to much more FC-type responses in their positive than in their negative forms whereas SI-related sentences gave rise to slightly more SI-type responses in their negative than in their positive forms.

4.5 Discussion

The main finding here is that SI sentences involving *permitted* do not give rise to more scalar implicatures than their negative variants involving *not required*. If anything, our results suggest that the direct SIs associated with the use of *permitted* is *less* robust than the indirect ones associated with the used of *not required*. This result therefore rules out the possibility that the discrepancies observed between FC and NFC across our experiments is to be attributed to a parallel contrast in robustness between the implicatures triggered by the use of *permitted* vs. *not required*.

5 General Discussion

Let us summarise the main findings. We consistently observed across the three experiments that NFC is an available inference, contrary to claims by Ciardelli et al. 2018 and Goldstein 2019, and that FC is a very robust inference while NFC is much less so, ultimately extending the main findings reported in Chemla 2009b. Finally, the results from Experiment 3 establish that the difference in robustness between FC and NFC cannot be attributed to *permitted* being a better prompt for scalar reasoning than *not required* in our experiments.

As mentioned in Section 1, Chemla's (2009b) experimental evidence was suggestive but not conclusive with respect to the existence of NFC for two reasons. First, the lack of suitable baselines for the inference task did not allow a straightforward interpretation of the endorsement rate for NFC observed in the experiment. Second, the use of simple conjunctions could have given rise to PLURAL HOMOGENEITY effects which, in turn, would give rise to a meaning entailing NFC. In our experiments, we compared the availability of FC and NFC against several control conditions that provided baselines for acceptance and rejection for the same sentences and for the same pictures. In addition, in Experiments 2 and 3, we used complex conjunctions (i.e., *both A and B*), rather than simple conjunctions (i.e., *A and B*), in the target sentences for NFC, so as to avoid potential homogeneity effects coming from simple conjunctions. Comparing the results of Experiments 1 and 2, we observed a sizable effect of this manipulation on the rejection rates for NFC sentences,

14 Clearly, given the direction of the main effects, both of them were driven by the very distinct behavior of FC sentences and, specifically, by the very rate of rejection responses received by these sentences in their **Target** conditions.

suggesting that part of the responses to the target NFC conditions in Experiment 1 were due to such homogeneity effects, rather than to NFC itself (similarly for the results of Chemla 2009b). Importantly, however, NFC still exhibits higher rejection rates than the true control conditions in Experiments 2 and 3, suggesting that it is an available inference.

Let us now evaluate the theoretical possibilities in light of these findings. As mentioned in the introduction, there are two major approaches to FC: the entailment approach (Aloni 2003, 2007, 2018, Goldstein 2019, Rothschild & Yablo 2018, Simons 2005, Willer 2017, Zimmerman 2000) and the implicature approach (Alonso-Ovalle 2006, Bar-Lev 2018, Bar-Lev & Fox 2020, Fox 2007, Franke 2011, Klinedinst 2007, Kratzer & Shimoyama 2002, Santorio & Romoli 2017). Similarly, NFC can be analyzed as an entailment or as an implicature: the entailment approach (Aloni 2018, Willer 2017) and the implicature approach (Fox 2007, Bar-Lev 2018, Bar-Lev & Fox 2020). The existence of NFC is not in itself a problem for the entailment or the implicature approach to FC, as both may derive NFC on certain assumptions. In particular, certain entailment theories of FC like Aloni (2003, 2007) and Goldstein (2019) are not incompatible with the implicature approach to NFC, as we will discuss in greater detail later, although these authors do not discuss this possibility explicitly. Thus, we end up with four possible types of theories, as shown in Table 7 (though we are not aware of any existing theory deriving FC as an implicature and NFC as an entailment, i.e., Theory II in Table 7).

	FC	NFC	
I	entailment	entailment	Aloni 2018, Willer 2017
II	implicature	entailment	—
III	entailment	implicature	Aloni 2003, 2007, Goldstein 2019
IV	implicature	implicature	Fox 2007, Bar-Lev 2018, Bar-Lev & Fox 2020

Table 7 Four possible approaches to FC and NFC

This leaves us with two types of approach, which we will call *the hybrid approach* (III), deriving FC as an entailment and NFC as an implicature, and the *uniform approach* either treating both as entailments (I) or both as implicatures (IV).

In the remainder of this paper, we consider both types of approach in turn. As we explain, for the hybrid approach, the observed difference in robustness between FC and NFC falls out more or less straightforwardly given the widely replicated tendency for implicatures to be less robust than entailments in various experimental tasks, including tasks similar to ours (Bott & Noveck 2004, Marty et al. 2015, Van Tiel et al. 2016 among others). On the other hand, the difference between FC and NFC constitutes a challenge for either version of the uniform approach. In particular, the challenge for these approaches is to identify what makes NFC weak, or conversely what makes FC particularly robust, without having this mechanism extends from one type of free choice inferences to the other. We will discuss the different options we see for the uniform approaches and how to reconcile them with the findings from Experiment 3.

5.1 Uniform approaches and their challenges

5.1.1 FC and NFC as entailments

Approach I in Table 7 derives both FC and NFC as entailments. The robustness of FC is therefore captured by this approach, for it is exactly what one would expect if FC is an entailment. As mentioned, however, the difference between FC and NFC is more challenging for this approach: if NFC were an entailment as well, it should, on the face of it, be at least as robust as FC.

One way to make this approach compatible with the diminished rate of NFC would be to assume that sentences of the form *not(required(A and B))* are ambiguous between two readings, one with NFC and one without it, possibly attributing the reading without NFC to a mechanism allowing for the 'cancellation' of free choice. This is arguably needed anyway for both FC and NFC sentences, given that these sentences can be read without their free choice inference in certain contexts. A typical example of FC 'cancellation' is given in (15), where the continuation makes it clear that the sentence is to be interpreted without FC.[15]

(15) It is permitted that Mia buys the pear or the banana. But I don't remember which one (she is permitted to buy).

While slightly less natural, the same can be reproduced for NFC sentences, as in (16), suggesting that we also need a way of reading these sentences without the inference.

(16) It is not required that Mia buys the pear and the banana. But I don't remember which one (she is not required to buy).

One could think, therefore, that the mechanism preventing the free choice readings of (15) and (16) is at play in our results for those participants who accepted the target sentences. The challenge for this strategy is of course to explain why this mechanism would be more available, or more solicited, when interpreting NFC than FC sentences, which were almost always interpreted with the free choice inference. One possibility here would be to treat negation as ambiguous between a meaning that gives rise to NFC and one that does not, e.g., by giving negation an appropriate alternative-sensitive denotation (see Schulz 2019, Romoli et al. 2020 for discussion). This line of explanation would have the immediate advantage of applying to NFC sentences without extending to FC sentences, as the latter do not involve negation. In developing this account, however, one would have to make sure it does not spoil the predicted robustness of DOUBLE PROHIBITION sentences like (17),

15 We should note here that it is controversial whether examples like (15) actually involve cancellation of the FC inference or rather invite a wide scope reading of disjunction with respect to the modal (see Fusco 2019 for discussion).

the inference of which appears as robust as FC in our results (see Experiment 1, Sentence C1 with Picture AD, which were rejected 90% of the time).[16]

(17) It is not permitted that Mia buys the pear or the banana.

In sum, our results are challenging for the uniform approach deriving both FC and NFC as entailments. As discussed, one could supplement this approach with an ambiguity entry of negation in order to account for the diminished rate of NFC. While we acknowledge that this is possible, we think that developing such an account to make the right predictions for NFC and related sentences like (17) is a non-trivial task.

5.1.2 FC and NFC as implicatures

Approach IV in Table 7 derives both FC and NFC as implicatures. The difference in robustness between FC and NFC is challenging for this approach as well. In this case, the results for NFC are in line with the idea that NFC is derived as an implicature, but explaining the robustness of FC is challenging. The current literature contains some ideas that have been put forward to explain why FC is a very robust inference, in contrast to other implicatures. Yet the observed discrepancy between FC and NFC, together with the results for DSI and ISI, remains unexpected. In what follows, we review how this approach works for FC and NFC, and discuss the remaining challenges in more detail.

The type of implicatures that the implicature approach to FC and NFC makes use of is generated in reference to alternative expressions, or alternatives, for short.[17] We can identify two major types of implicature theories with respect to which alternatives are used to derive FC and NFC.

According to Fox (2007), the crucial alternatives for deriving FC from *permitted(A or B)* mean $\Diamond A \wedge \neg\Diamond B$ and $\neg\Diamond A \wedge \Diamond B$, and these alternatives are derived by means of the alternatives *permitted(A)* and *permitted(B)* with their own implicatures.[18] These alternatives give rise to the implicature $\Diamond A \leftrightarrow \Diamond B$, which, together with the literal meaning of *permitted(A or B)*, entails FC. The key assumption on this approach is, therefore, the nested computation of implicatures, where implicatures can be derived from alternatives that have their own

16 Another option would be to argue that the difference lies in another reading of NFC where conjunction takes wide scope over the modal but below negation (i.e., $\neg(\Box A \wedge \Box B)$), a configuration which, in most accounts, does not give rise to a free choice inference. While we controlled for the scope of conjunction using finite clauses, we cannot exclude that this factor played a role nonetheless. We thank an anonymous reviewer for suggesting this possibility to us.

17 We assume here that alternatives are linguistic expressions (Sauerland 2004, Katzir 2007, Fox & Katzir 2011 among many others). Not every theory of implicature uses linguistic alternatives, but this assumption is often taken to be crucial in constraining the theory. See discussion in the work cited here as well as in Breheny et al. 2018.

18 Fox (2007) makes specific assumptions about the set of alternatives for this nested level of implicatures, but the details do not concern us here.

implicatures. Under Fox's (2007) theory, deriving NFC from *not(required(A and B))* can be done in a parallel fashion. This time, the crucial alternatives are derived from the nested computation of implicatures on *not(required(A))* and *not(required(B))*. With their own implicatures, these alternatives mean $\neg\Box A \wedge \Box B$ and $\Box A \wedge \neg\Box B$, respectively. In negating the meanings of these alternatives, we obtain the implicature $\neg\Box A \leftrightarrow \neg\Box B$ which, together with the literal meaning of *not(required(A or B))*, entails NFC.

The second implicature theory of FC and NFC is due to Bar-Lev (2018) and Bar-Lev & Fox (2020). According to this theory, the crucial alternatives for the FC inference of *permitted(A or B)* are simply *permitted(A)* and *permitted(B)*, without nested implicatures. Instead of being excluded, the meanings of these alternative are instead *included*, which directly derives the FC inference, $\Diamond A \wedge \Diamond B$. Similarly, the NFC inference of *not(required(A and B))* is derived by including the meanings of the alternatives *not(required(A))* and *not(required(B))*.

How can we explain the difference in robustness between FC and NFC under these theories? Proponents of the second version of the implicature approach to FC have proposed an idea that is relevant here (Bar-Lev & Fox 2020): they conjecture that implicatures derived by *inclusion* of alternatives are generally very robust, in contrast to implicatures derived by *exclusion* of alternatives. Note for instance that the DSI and ISI of *permitted(A)* and *not(required(A))* would be derived by *excluding* alternatives, namely, *required(A)* and *not(permitted(A))* respectively. This idea, however, cannot explain the difference between FC and NFC, as both of them would be derived by inclusion under their theory.[19]

Another way to explain the robustness of FC relative to other implicatures would be to resort to the form of the crucial alternatives. That is, under both theories above, the crucial alternatives are *subconstituents* of the original sentence, while the alternatives for the DSI of *permitted(A)* and the ISI of *not(required(A))* are derived by *lexical substitution*. Specifically, the former is derived by negating the alternative *required(A)*, and the latter is derived by negating the alternative *not(permitted(A))*. Given this, one could assume that *subconstituent alternatives* are salient alternatives simply by virtue of being made up of uttered expressions, and more often used to derive inferences than *lexical alternatives*, which are not as salient and can often be ignored (Chemla & Bott 2014, Singh et al. 2016, Tieu et al. 2016 among others). This would explain the difference between FC and DSI/ISI.[20] This idea, however, would not explain by itself the difference between FC and NFC, as both

19 A version of this theory deriving FC by inclusion but NFC by exclusion using Fox's (2007) route remains possible; however, it would need to make peculiar assumptions about the alternatives. Specifically, it would have to assume that *permitted(A or B)* has *permitted(A)* and *permitted(B)* as alternatives, which are included to give rise to FC, while *not(required(A and B))* has *not(required(A))* and *not(required(B))* with their own implicatures as alternatives, which are excluded to give rise to NFC. For the time being, we cannot think of a principled reason for such an asymmetry in the set of alternatives between both cases.

20 Note that, for Fox 2007, the nested level of implicature would require lexical alternatives. That is, in the case of *not(required(A and B)*, the implicature for the subconstituent alternative *not(required(A)* would be derived in reference to its alternative *not(required(B))*. However, one could assume that the latter alternative is also contained in the uttered sentence and, therefore, salient.

of them are derived from subconstituent alternatives, as explained above. In sum, neither of the two main ideas in the literature that are invoked to account for the robustness of FC predicts the difference between the latter and NFC.

In parallel to the discussion about the uniform approach based on entailment, one promising direction is to supplement the implicature approach with an account of the difference between FC and NFC linked to negation. We briefly mention two possible ways of implementing this account.

The first is actually discussed by Bar-Lev & Fox (2020) in relation to other data. In essence, they argue that negation introduces alternatives and, as a result, sentences involving negation can have less inferences than usually predicted, depending on how these extra alternatives compose and interact with the rest of the alternatives. In the case of NFC, one can show that, when these alternatives are factored in, no inference arises. Given that this only happens for NFC, it could account for the difference between NFC and FC. The challenge for this line of explanation, however, is that, as it stands, it incorrectly extends to indirect SIs in predicting that indirect SIs should be less robust than direct ones (see Romoli et al. 2020 for discussion).

The second implementation capitalises on the pragmatics of negation and how it interacts with the relevance of alternatives. The idea is quite simple: if the alternatives for NFC are perceived as less relevant than those for FC, possibly as a result of the pragmatics of negative sentences, this could make the NFC inference less robust. This is not implausible given that it is widely acknowledged that the space of alternatives needs to be restricted to 'relevant' ones, or else the theory would both undergenerate and overgenerate (see Katzir 2007, Fox 2007, Magri 2011, Breheny et al. 2018 among many others). However, the main difficulty for this idea comes from the observation that ISI-related sentences, which also involve negation, behaved quite similarly as DSI-related sentences according to the results of Experiment 3. Note that previous experimental studies on ISIs (e.g., Chemla 2009b, Cremers & Chemla 2014) did not observe a difference comparable to FC vs. NFC either.

Finally, one possible way of making sense of these results would be to combine the effect of negation and the effect of lexical vs. subconstituent alternatives (or, alternatively, the effect of inclusion vs. exclusion) in such a way that the latter effect is diminished for ISIs, in comparison to DSIs. However, it is not obvious to us at this point how such an interaction between these two factors could be given theoretical motivation.[21]

21 Another option to account for differences in strength between pragmatic inferences, which is discussed in Champollion et al. (2019) and implemented in an implicature account, comes from the observation that some inferences like free choice are impervious to differences in the prior beliefs of the speaker/hearer. At this stage, however, it is unclear to us how this line of explanation would extend to account for the difference between FC and NFC, which are derived in the very same way in Champollion et al. 2019 and, more generally, on the implicature approach.

5.2 The hybrid approach

Our results are explained by the hybrid approach in a more straightforward fashion. Indeed, a number of previous studies have found, using various experimental methods, that entailments give rise to very robust effects, while implicatures exhibit more intermediate behavior (van Tiel et al. 2016, Bott & Noveck 2004 among many others). In this respect, note that, in our results, rates for FC inferences were at ceiling, in line with previous studies on FC (Chemla 2009b, Chemla & Bott 2014). This is immediately accounted for if FC is an entailment. Similarly, the lower rates for NFC, especially in Experiments 2 and 3, can be taken as evidence that NFC is instead a type of implicature, along with DSI and ISI.

One concrete implementation of the hybrid approach can be constructed simply by combining an alternative semantic theory of FC (Aloni 2003, 2007, Simons 2005, Goldstein 2019), and the implicature approach to NFC. The alternative semantic theory of FC makes two core assumptions: (i) a disjunction introduces its disjuncts as alternative propositions, and (ii) a possibility modal combines point-wise with the alternatives of its complement.

To illustrate, let us sketch here a version of the alternative semantic theory of FC. To do so, we will adopt the version by Goldstein (2019), as it is most straightforwardly compatible with our results. On this theory, a sentence meaning is a set of propositions (also called 'alternatives' in this system). An atomic sentence simply denotes a singleton set with the proposition it expresses in the standard system, as shown in (18). Negation combines with the meaning of a sentence A, and it returns the singleton set containing the set of worlds in which no alternative of A is true, as show in (19).

(18) $[\![\text{p}]\!] = \{\lambda w\,.p(w) = 1\}$

(19) $[\![\text{not A}]\!] = \{W - \bigcup[\![\text{A}]\!]\} = \{\overline{\bigcup[\![\text{A}]\!]}\}$

Crucially, on this theory, disjunction denotes a set of alternatives corresponding to its disjuncts, as in (20), and a possibility modal universally quantifies over the alternatives of its complement, as in (21).

(20) $[\![\text{A or B}]\!] = [\![\text{A}]\!] \cup [\![\text{B}]\!]$

(21) $[\![\text{permitted(A)}]\!] = \{\lambda w\,.\forall p \in [\![\text{A}]\!] \Diamond p(w) = 1\}$

Combining these two elements gives us FC as an entailment. We illustrate this outcome in (22) for a simple case where A and B are atomic sentences, denoting $\{\mathbf{A}\}$ and $\{\mathbf{B}\}$ respectively. The combination of disjunction with each disjunct as alternatives and the possibility modal requiring each alternatives of its prejacent to be possible, gives us the conjunction of possibilities about each disjunct.

(22) $[\![\text{permitted(A or B)}]\!] = \{\lambda w\,.\forall p \in [\![\text{A or B}]\!] \Diamond p(w) = 1\} =$
$\{\lambda w\,.\forall p \in \{\mathbf{A},\mathbf{B}\} \Diamond p(w) = 1\} = \{\lambda w\,.\Diamond\mathbf{A}(w) = \Diamond\mathbf{B}(w) = 1\}$

As mentioned, any entailment approach to FC needs to say something about how Double Prohibition is to be accounted for. Goldstein (2019) proposes that possibility modals are associated with a homogeneity presupposition, according to which either all alternatives of its prejacent are possible or none of them is. That is, he suggests to modify the definition in (21) as shown in (23).[22]

(23) $[\![\text{permitted(A)}]\!] = \{\lambda w : \exists v \in \{1,0\} \forall p \in [\![\text{A}]\!] \Diamond p(w) = v \;.\; \forall p \in [\![\text{A}]\!] \Diamond p(w) = 1\}$

As Goldstein (2019) shows, a negated possibility modal embedding disjunction will now only be defined and true when neither of the disjunct is possible. To illustrate, consider again a case where A and B are atomic sentences:

(24) $[\![\text{not permitted(A or B)}]\!]$ is defined in some world w only if
$\exists v \in \{1,0\} \forall p \in [\![\text{A or B}]\!] \Diamond p(w) = v$, hence only if
$\Diamond \mathbf{A}(w) = \Diamond \mathbf{B}(w) = 1$ or $\Diamond \mathbf{A}(w) = \Diamond \mathbf{B}(w) = 0$;
when defined it is true in w iff $\forall p \in [\![\text{A or B}]\!] \Diamond p(w) = 1$ is false
$\Diamond \mathbf{A}(w) \neq 1$ or $\Diamond \mathbf{B}(w) \neq 1$

(24) is defined and true only when neither A nor B are possible: the truth-conditions entail that at least one of the two possibility claims is not true and, given the definedness conditions, it follows that they both have to be false. As a result, (24) entails that it is not permitted A and it is not permitted B. This is in line with the intuitive robustness of double prohibition, also apparent in our results of Experiment 1, as mentioned above.[23]

Crucially, however, this account of FC and double-prohibition does not predict NFC. To illustrate, consider the meaning of conjunction, which is defined on this approach as the pointwise intersection of the alternatives of each conjunct, as shown in (25).

(25) $[\![\text{A and B}]\!] = \{p \cap q \mid p \in [\![\text{A}]\!], q \in [\![\text{B}]\!]\}$

One standard way to deal with the necessity modal is to define the necessity modal as the dual of the possibility modal (Aloni 2003, 2007):

22 We follow the notation in Goldstein (2019), where in a configuration of the form $\lambda : \phi.\psi$, ϕ is the presupposition and ψ the assertion. Alternative ideas in the literature include: (i) Aloni (2003, 2007), which is based on disjunction being ambiguous, and (ii) Aloni (2018) and Willer (2017), who make use of a 'bilateral system', a system that assigns positive and negative meanings to each expression.

23 An anonymous reviewer raises the worry that the homogeneity component which allows the derivation of double-prohibition, being itself not an entailment, could predict a less robust inference. We think that this is an empirical question and that one should eventually compare double-prohibition to other homogeneity-based phenomena under negation. This said, some results in the literature do suggest that homogeneity-based inferences with plural definites are quite robust (Križ & Chemla 2015, Tieu et al. 2019b among others). Moreover, Tieu et al. (2019a) tested double prohibition cases in mixed contexts with a ternary task and found overwhelming intermediate judgments, in line with the predictions of a homogeneity-based account.

EARLY ACCESS

(26) $[\![\text{required(A)}]\!] = [\![\text{not(permitted(not(A)))}]\!]$

Assuming that A and B are atomic sentences, *required(A and B)* will be interpreted as the singleton set in (27). Crucially, given the definition of negation above, *not required(A and B)* will simply be the complement of this singleton set, namely (28). The meaning in (28) does not entail NFC: it comprises all worlds in which it is not the case that both A and B are required (i.e., it is compatible with one of the two being required).

(27) $[\![\text{required(A and B)}]\!] = [\![\text{not(permitted(not(A and B)))}]\!]$
$= \{\overline{\Diamond(\overline{\mathbf{A}} \cup \overline{\mathbf{B}})}\} = \{\Box(\mathbf{A} \cap \mathbf{B})\}$

(28) $[\![\text{not(required(A and B))}]\!] = \{\overline{\Box(\mathbf{A} \cap \mathbf{B})}\}$

In sum, this approach derives FC as an entailment, double-prohibition as a combination of entailment and presupposition, but, crucially for us, it does not derive NFC. It is therefore compatible with an implicature approach to NFC along the lines we discussed in Section 5.1.[24] The combination of the entailment approach to FC and the implicature approach to NFC can explain our experimental results.

5.3 Distributivity inference

Before closing, we would like to briefly mention a potential challenge for the two types of approach we discussed above, which comes from a type of inference related to FC and NFC. As is well known, a sentence of the form *required(A or B)* can give rise to an inference to $\Diamond A$ and $\Diamond B$, sometimes referred to as the 'distributivity inference.' (Kratzer & Shimoyama 2002, Simons 2005, Alonso-Ovalle 2006, Santorio & Romoli 2017 among others). Neither of the approaches above derives this inference as an entailment. For instance, adopting essentially the system sketched for the hybrid approach above, Aloni (2007) proposes to capture this inference as a scalar implicature, along the lines of Kratzer & Shimoyama 2002 and Alonso-Ovalle 2006. This is certainly a theoretical possibility, but it needs to be seen whether this approach is empirically supported. In particular, if the distributivity inference turns out to be as robust as FC, then one might want to account for it as an entailment, rather than as a scalar implicature. Although we leave this as an open empirical issue in this paper, it should be mentioned that Simons (2005) and Fusco (2015) derive both FC and the distributivity inference as entailments (and are compatible with the implicature approach to NFC). Thus, there is a theoretical debate here that is similar in nature to the debate on

24 This would mean that, in addition to the 'semantic' alternatives associated with the meaning of sentences in the alternatives semantics outlined above, there are also 'formal' alternatives, from which implicatures are derived (see Cremers et al. 2017 for more discussion and a similar hybrid approach to the derivation of the inferences associated with modified numerals). We also notice that *permitted(A or B)* will not have any implicatures, because its literal meaning is already strong, entailing FC, so there is no redundancy with respect to FC.

FC: some theories derive the distributivity inference as an implicature and others derive it as an entailment. In order to adjudicate between these approaches, more experimental work is necessary.

6 Conclusion

Our three experiments provide strong prima-facie evidence for the existence of NFC. Our results also establish that NFC is less robust than FC, consistent with Chemla (2009b), and that it is more similar to DSIs and ISIs. It bears pointing out that all the sentences we investigated involved deontic modality and that NFC sentences were all of the form $\neg\Box(A \wedge B)$ (that is, we did not investigate classically equivalent NFC configurations such as $\Diamond\neg(A \wedge B)$, which have also been discussed in the literature). Both these factors could play a role and should be investigated further. Having said that, we think that the results of our studies are currently best explained by a hybrid approach, according to which FC is an entailment while NFC is an implicature, and pose challenges for a uniform approach, whether both FC and NFC are conceived entailments or else as implicatures. While we do not think these challenges are necessarily insurmountable, they are important issues that remain to be solved for any uniform approach to FC and NFC.

A Instructions to participants

A.1 Instructions for Experiments 1 & 2

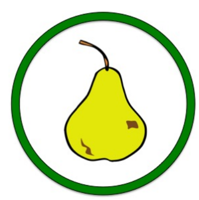
In this study, we will ask for your judgments about certain kinds of sentences in English. These sentences will involve two children, Sam and Mia. Here they are:

Sam

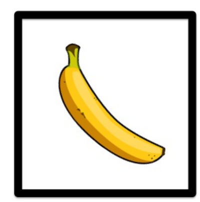
Mia

Sam and Mia are going to the supermarket. Their parents have some rules about what Sam and Mia are allowed to buy, what they have to buy, and what they are not allowed to buy. Here is how we'll represent these rules. Take a look at the examples below:

The **red circle** around the burger means that *it is not permitted* to buy the burger.



The **green circle** around the pear means that *it is permitted but not required* to buy the pear.



The **black square** around the banana means that *it is required* to buy the banana.

You will see many pictures depicting such rules. Each picture will be followed by a sentence that relates to it. Your task is to decide if that sentence is or not a good description of the picture you see. You will click on "Good" if you consider the sentence a good description of the picture; otherwise click on "Not good".

A.2 Instructions for Experiment 3

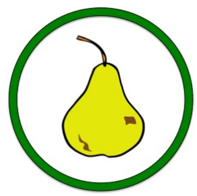
In this study, you will see pictures, each of which followed by a sentence describing it. Your task is to indicate if that sentence is or is not a good description of the picture you see. You will press '1' if you consider the sentence a good description of the picture; otherwise you will press '0'.

Before getting to the study, let's start with a couple of practice trials. Those trials will help you get familiar with the display that will be used throughout the study.

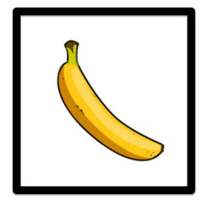
****

You are ready for the study! The study describes a new diet program which includes various rules about what you're allowed to eat, what you have to eat, and what you're not allowed to eat. Here is how we'll represent those rules:

The **red circle** around the burger means that *it is forbidden* to buy the burger.



The **green circle** around the pear means that *it is optional* to buy the pear.



The **black square** around the banana means that *it is obligatory* to buy the banana.

You will see many pictures depicting such rules, and sometimes those pictures will depict two rules at once.

B Pre-test trials in Experiments 1 & 2

Pre-test trials were designed to assess whether participants understood correctly how to interpret in isolation the three symbols enclosing the food items. Pictures in those trials involved only one food item, enclosed within one of the three symbols, as shown in Table 8.

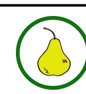
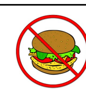
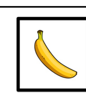
Label	Description of the picture type	Example picture
A	permitted to buy the item	
D	not permitted to buy the item	
R	required to buy the item	

Table 8 Description and examples of the pictures used in the pre-test trials in Experiments 1 & 2.

Sentences in those trials were constructed using one of the four frames given in Table 9. The [name] term was the name of one of our two characters, *Mia* or *Sam*, and the [food] term was the name of the food item displayed on the picture.

Label	Description of the sentence type
T1	It is permitted that [name] buys the [food].
T2	It is not permitted that [name] buys the [food].
T3	It is required that [name] buys the [food].
T4	It is not required that [name] buys the [food].

Table 9 Description of the sentences used in the pre-test trials in Experiments 1 & 2, where [food] corresponds to the name of the food item displayed on the pictures they were paired with.

A summary of the items used in the pre-test trials is given in Table 10. In a nutshell, T1 and T2 sentences were paired with A and D pictures, and T3 and T4 sentences were paired with A and R pictures. The former items were used to verify participants' understanding of the red circle, the latter to verify participants' understanding of the black square, and their combinations to verify participants' understanding of the green circle. We conjectured that participants performing too low on those items did not master the intended interpretation of the symbols (when used in isolation) and therefore did not have the background necessary to understand the building blocks of the more complex pictures used in the test items. For these reasons, we set up our online surveys so that a participant making errors in more than 30% of the pre-test trials could not continue to the test trials. In total, 5 participants in Experiment 1 and 1 participant in Experiment 2 were screened-out that way. This outcome did not affect their payment. The mean accuracy rate of the remaining participants was 92.5% (95% CI [94, 91]) in Experiment 1 and 93.5% (95% CI [95, 92]) in Experiment 2.

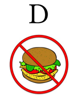
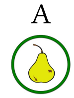
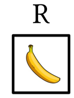
		D 	A 	R 
T1	It is permitted that Mia buys the *food*.	False (2)	True (2)	N/A
T2	It is not permitted that Mia buys the *food*.	True (2)	False (2)	N/A
T3	It is required that Mia buys the *food*.	N/A	False (2)	True (2)
T4	It is not required that Mia buys the *food*.	N/A	True (2)	False (2)

Table 10 Summary of the sentence-picture combinations used in the pre-test trials in Experiments 1 & 2 (*food* is used as a cover term to refer to the food item on the picture). Numbers in parenthesis refer to the number of items included in the pre-test phase to exemplify the different combinations.

C Tables for the statistical analyses in Experiments 1-3

		β	Std. Error	z-value	Pr(|z|)
WITHIN SENTENCE TYPE					
FC					
Target vs. **False**	`Response~Condition+(1|Subject)`				
	(Intercept)	4.477	0.802	5.577	< .001
	ConditionFalse	0.652	0.509	1.281	.2
Target vs. **True**	`Response~Condition+(1+Condition|Subject)`				
	(Intercept)	7.464	1.708	4.370	< .001
	ConditionTrue	−14.745	2.616	−5.637	< .001
NFC					
Target vs. **False**	`Response~Condition+(1|Subject)`				
	(Intercept)	2.491	0.466	5.343	< .001
	ConditionFalse	2.541	0.4688	5.422	< .001
Target vs. **True**	`Response~Condition+(1+Condition|Subject)`				
	(Intercept)	3.138	0.729	4.301	< .001
	ConditionTrue	−6.416	1.158	−5.540	< .001
BETWEEN SENTENCE TYPES	Responses to FC vs. NFC sentences in the **Target** vs. **False** condition				
	`Response~Sentence*Condition+(1+Condition|Subject)`				
	(Intercept)	4.094	0.448	9.124	< .001
	Sentence	−1.862	0.269	6.906	< .001
	Condition	−0.700	0.598	−1.170	.242
	Sentence:Condition	1.562	0.613	2.547	< .05

Table 11 Outputs of the Generalized linear mixed models used to analyse participants' responses to the target sentences in Experiment 1. Note: R pseudo-code shown in the first line of every section.

		β	Std. Error	z-value	Pr(|z|)
WITHIN PICTURE TYPE	`Response~Condition+(1+Condition|Subject)`				
AD					
Target vs. **False**					
	(Intercept)	7.593	1.761	4.311	$< .001$
	ConditionFalse	−1.288	2.728	−0.472	.637
Target vs. **True**					
	(Intercept)	7.554	1.729	4.368	$< .001$
	ConditionTrue	−18.257	3.467	−5.266	$< .001$
AR					
Target vs. **False**					
	(Intercept)	3.144	0.730	4.307	$< .001$
	ConditionFalse	5.535	1.987	2.785	$< .01$
Target vs. **True**					
	(Intercept)	3.198	0.754	4.239	$< .001$
	ConditionTrue	−7.424	1.741	−4.265	$< .001$
BETWEEN PICTURE TYPES	Responses to AR vs. AD pictures in the **Target** vs. **False** condition				
	`Response~Picture*Condition+(1+Picture|Subject)`				
	(Intercept)	4.315	0.636	6.784	$< .001$
	Picture	−1.834	0.756	−2.423	$< .05$
	Condition	−0.798	0.366	−2.179	$< .05$
	Sentence:Condition	2.133	0.499	4.269	$< .001$

Table 12 Outputs of the Generalized linear mixed models used to compare participants' responses to the control and target sentences in Experiment 1. Note: R pseudo-code shown at the top of the main sections.

		β	Std. Error	*z*-value	Pr(|z|)
WITHIN SENTENCE TYPE					
FC-either					
Target vs. **False**	`Response~Condition+(1|Subject)`				
	(Intercept)	4.402	0.717	6.140	< .001
	ConditionFalse	0.254	0.409	0.621	.5
Target vs. **True**	`Response~Condition+(1+Condition|Subject)`				
	(Intercept)	7.252	1.635	4.436	< .001
	ConditionTrue	−16.598	2.721	−6.100	< .001
NFC-both					
Target vs. **False**	`Response~Condition+(1|Subject)`				
	(Intercept)	−0.590	0.382	−1.546	.1
	ConditionFalse	5.118	0.454	11.257	< .001
Target vs. **True**	`Response~Condition+(1+Condition|Subject)`				
	(Intercept)	−1.061	0.632	−1.678	.09
	ConditionTrue	−7.593	1.925	−3.944	< .001
BETWEEN SENTENCE TYPES	Responses to FC-either vs. NFC-both sentences in the **Target** vs. **False** condition				
	`Response~Sentence*Condition+(1+Condition|Subject)`				
	(Intercept)	4.110	0.436	9.410	< .001
	Sentence	−4.631	0.306	−15.107	< .001
	Condition	0.230	0.946	0.244	.8
	Sentence:Condition	4.759	0.590	8.065	< .001

Table 13 Outputs of the Generalized linear mixed models used to analyse participants' responses to the target sentences in Experiment 2. Note: R pseudo-code shown in the first line of every section.

		β	Std. Error	z-value	Pr(|z|)
WITHIN PICTURE TYPE	`Response~Condition+(1+Condition|Subject)`				
AD					
Target vs. **False**					
	(Intercept)	6.506	1.608	4.047	$<.001$
	ConditionFalse	−4.363	1.631	−2.676	$<.01$
Target vs. **True**					
	(Intercept)	7.259	1.632	4.447	$<.001$
	ConditionTrue	−16.858	2.875	−5.863	$<.001$
AR					
Target vs. **False**					
	(Intercept)	−1.059	0.630	−1.68	.09
	ConditionFalse	10.245	1.918	5.34	$<.001$
Target vs. **True**					
	(Intercept)	−1.092	0.639	−1.708	.08
	ConditionTrue	−4.768	1.824	−2.614	$<.01$
BETWEEN PICTURE TYPES	Responses to AR vs. AD pictures in the **Target** vs. **False** condition				
	`Response~Picture*Condition+(1+Picture|Subject)`				
	(Intercept)	3.815	0.460	8.277	$<.001$
	Picture	−4.304	0.527	−8.169	$<.001$
	Condition	−1.335	0.309	−4.316	$<.001$
	Sentence:Condition	4.914	0.440	11.160	$<.001$

Table 14 Outputs of the Generalized linear mixed models used to compare participants' responses to the control and target sentences in Experiment 2. Note: R pseudo-code shown at the top of the main sections.

		β	Std. Error	z-value	Pr(|z|)
WITHIN SENTENCE TYPE	`Response~Condition+(1+Condition|Subject)`				
SI					
Target vs. **False**					
	(Intercept)	−1.813	0.769	−2.356	< .05
	ConditionFalse	11.375	3.610	3.151	< .01
Target vs. **True**					
	(Intercept)	−1.900	0.824	−2.305	< .05
	ConditionTrue	−5.385	2.533	−2.126	< .05
ISI					
Target vs. **False**					
	(Intercept)	−0.182	0.414	−0.440	.6
	ConditionFalse	7.728	2.046	3.776	< .001
Target vs. **True**					
	(Intercept)	−0.189	0.415	−0.456	.6
	ConditionTrue	−2.773	0.848	−3.270	< .01
FC					
Target vs. **False**					
	(Intercept)	8.105	2.030	3.993	< .001
	ConditionFalse	2.368	4.280	0.553	.5
Target vs. **True**					
	(Intercept)	7.797	2.072	3.763	< .001
	ConditionTrue	−14.455	2.629	−5.499	< .001
NFC					
Target vs. **False**					
	(Intercept)	−0.058	0.443	−0.133	.89
	ConditionFalse	7.903	1.968	4.014	< .001
Target vs. **True**					
	(Intercept)	−0.050	0.440	−0.114	0.9
	ConditionTrue	−2.898	0.833	−3.47	< .001
BETWEEN SENTENCE TYPES	Responses to target sentences in the **Target** conditions				
	`Response~Inference*Polarity+(1+Inference|Subject)+(1|Item)`				
	(Intercept)	−1.140	0.388	−2.935	< .01
	InferenceFC	5.066	0.663	7.633	< .001
	PolarityNegative	0.935	0.356	2.624	< .01
	InferenceFC:PolarityNegative	−4.903	0.678	−7.231	< .001

Table 15 Outputs of the Generalized linear mixed models used to analyse participants' responses to the target sentences in Experiment 3. Note: R pseudo-code shown in the first line of every section.

References

Aloni, Maria. 2003. Free choice in modal contexts. *Proceedings of Sinn und Bedeutung 7* 25–37.

Aloni, Maria. 2007. Free cohice, modals, and imperatives. *Natural Language Semantics* 15(1). 65–94. http://dx.doi.org/10.1007/s11050-007-9010-2.

Aloni, Maria. 2018. FC disjunction in state-based semantics. Ms., University of Amsterdam.

Alonso-Ovalle, Luis. 2006. *Disjunction in alternative semantics*: University of Massachusetts, Amherst dissertation.

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* 28. 175–223. http://dx.doi.org/10.1007/s11050-020-09162-y.

Bates, Douglas, Martin Maechler & Ben Bolker. 2011. Package 'lme4'. *lme4: Linear mixed-effects models using S4 classes* .

Bott, Lewis & Ira Noveck. 2004. Some utterances are underinformative. *Journal of Memory and Language* 51(3). 437–457. http://dx.doi.org/10.1016/j.jml.2004.05.006.

Breheny, Richard. 2005. Exhaustivity, homogeneity, and definiteness. In Paul Dekker & Michael Franke (eds.), *Proceedings of the Fifteenth Amsterdam Colloquium*, 59–65. ILLC/Department of Philosophy, University of Amsterdam Grafisch Centrum Amsterdam.

Breheny, Richard, Nathan Klinedinst, Jacopo Romoli & Yasutada Sudo. 2018. The symmetry problem: current theories and prospects. *Natural Language Semantics* 26(2). 85–110. http://dx.doi.org/https://doi.org/10.1007/s11050-017-9141-z.

Champollion, Lucas, Anna Alsop & Ioana Grosu. 2019. Free choice disjunction as a rational speech act. In *Proceedings of SALT 29*, 238–257.

Chemla, Emmanuel. 2009a. Similarity: towards a unified account of scalar implicatures, free choice permission and presupposition projection. Ms., LSCP and MIT.

Chemla, Emmanuel. 2009b. Universal implicatures and free choice effects: Experimental data. *Semantics and Pragmatics* 2(2). 1–33. http://dx.doi.org/10.3765/sp.2.2.

Chemla, Emmanuel & Lewis Bott. 2014. Processing inferences at the semantics/pragmatics frontier: disjunctions and *free choice*. *Cognition* 130(3). 380–396. http://dx.doi.org/10.1016/j.cognition.2013.11.013.

Ciardelli, Ivano, Zhang Linmin & Lucas Champollion. 2018. Two switches in the theory of counterfactuals: A study of truth conditionality and minimal change. *Linguistic and Philosophy* 41(6). 577–621. http://dx.doi.org/10.1007/s10988-018-9232-4.

Cremers, Alexandre & Emmanuel Chemla. 2014. Direct and indirect scalar implicatures share the same processing signature. In Salvatore Pistoia Reda (ed.), *Pragmatics, Semantics and the Case of Scalar Implicatures*, 201–227. London: Palgrave McMillan.

Cremers, Alexandre, Elizabeth Coppock, Jakub Dotlacil & Floris Roelofsen. 2017. Modified numerals: two routes to ignorance. Unpublished ms.

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.

EARLY ACCESS

Palgrave.
Fox, Danny & Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19(1). 87–107. http://dx.doi.org/10.1007/s11050-010-9065-3.
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.
Fusco, Melissa. 2015. Deontic modality and the semantics of choice. *Philosophers' imprint* 15.
Fusco, Melissa. 2019. Sluicing on free choice. *Semantics & Pragmatics* 12. 1–20.
Gajewski, Jon. 2005. *Neg-raising: Polarity and presupposition*: MIT dissertation.
Geurts, Bart & Nausicaa Pouscoulous. 2009. Free choice for all: a response to emmanuel chemla. *Semantics and Pragmatics* 2(5). 1–10. http://dx.doi.org/10.3765/sp.2.5.
Goldstein, Simon. 2019. Free choice and homogeneity. *Semantics and Pragmatics* 12(23). 1–47. http://dx.doi.org/10.3765/sp.
Gotzner, Nicole & Jacopo Romoli. 2018. The scalar inferences of strong scalar terms under negative quantifiers and constraints on the theory of alternatives. *Journal of Semantics* 35(1). 95–126. http://dx.doi.org/10.1093/jos/ffx016.
Jaeger, T Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007.
Kamp, Hans. 1974. Free choice permission. *Proceedings of the Aristotelian Society* 74. 57–74. http://dx.doi.org/10.1093/aristotelian/74.1.57.
Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistic and Philosophy* 30(6). 669–690. http://dx.doi.org/10.1007/s10988-008-9029-y.
Klinedinst, Nathan. 2007. *Plurality and possibility*: UCLA dissertation.
Kratzer, Angelika & Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. In Yukio Otsu (ed.), *Proceedings of the Tokyo conference on psycholinguistics*, vol. 3, 1–25. Tokyo: Hituzi Syobo.
Križ, Manuel. 2015. *Aspects of homogeneity in the semantics of natural language*: University of Vienna dissertation.
Križ, Manuel & Emmanuel Chemla. 2015. Two methods to find truth value gaps and their application to the projection problem of homogeneity. *Natural Language Semantics* 23(3). 205–248. http://dx.doi.org/10.1007/s11050-015-9114-z.
Magri, Giorgio. 2011. Another argument for embedded scalar implicatures based on oddness in downward entailing environments. *Semantics & Pragmatics* 4(6). 1–51. http://dx.doi.org/10.3765/sp.4.6.
Magri, Giorgio. 2014. An account for the homogeneity effects triggered by plural definites and conjunction based on double strengthening. In Salvatore Pistoia Reda (ed.), *Pragmatics, semantics, and the case of scalar implicatures*, 99–145. Basingstoke: Palgrave Macmillan.
Marty, Paul, Emmanuel Chemla & Benjamin Spector. 2015. Phantom readings: The case of modified numerals. *Language, Cognition and Neuroscience* 30(4). 462–477. http://dx.doi.org/10.1080/23273798.2014.931592.

Marty, Paul, Jacopo Romoli, Yasutada Sudo & Richard Breheny. 2020. Processing implicatures: a comparison between direct and indirect SIs. Presentation at Experiments in Linguistic Meaning, Upenn.

Meyer, Marie-Christine. 2016. *An Apple or a Pear: Free Choice Disjunction*. Ms., MS. Hebrew Uniersity and ZAS.

R Development Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Romoli, Jacopo, Paolo Santorio & Eva Wittenberg. 2020. Alternatives in Counterfactuals: What is right and what is not. Unpublished ms. University of Bergen, University of Maryland, and University of California, San Diego.

Rothschild, Daniel & Stephen Yablo. 2018. Permissive updates. MS UCL and MIT.

Santorio, Paolo & Jacopo Romoli. 2017. Probability and implicatures: A unified account of the scalar effects of disjunction under modals. *Semantics & Pragmatics* 10(13). 1–61. http://dx.doi.org/10.3765/sp.10.13.

Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27(3). 367–391. http://dx.doi.org/10.1023/B:LING.0000023378.71748.db.

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.

Simons, Mandy. 2005. Dividing things up: The semantics of or and the modal/or interaction. *Natural Language Semantics* 13. 271–316. http://dx.doi.org/10.1007/s11050-004-2900-7.

Singh, Raj, Ken Wexler, Andrea Astle-Rahim, Deepthi Kamawar & Danny Fox. 2016. Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics* 24(4). 305–352. http://dx.doi.org/10.1007/s11050-016-9126-3.

Szabolcsi, Anna & Bill Haddican. 2004. Conjunction meets negation: A study of cross-linguistic variation. *Journal of Semantics* 21(3). 219–250. http://dx.doi.org/10.1093/jos/21.3.219.

van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina & Bart Geurts. 2016. Scalar diversity. *Journal of Semantics* 33(1). 137–175. http://dx.doi.org/10.1093/jos/ffu017.

Tieu, Lyn, Cory Bill & Jacopo Romoli. 2019a. Homogeneity or implicature: an experimental investigation of free choice. In *Proceedings of SALT 29*, 706–726. https://semanticsarchive.net/Archive/DgzMjIyN/homogeneity.

Tieu, Lyn, Manuel Križ & Emmanuel Chemla. 2019b. Children's acquisition of homogeneity in plural definite descriptions. *Frontiers in Psychology* http://dx.doi.org/10.3389/fpsyg.2019.02329. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02329/full?&utm_source=Email_to_authors_&utm_medium=Email&utm_content=T1_11.5e1_author&utm_campaign=Email_publication&field=&journalName=Frontiers_in_Psychology&id=463805.

Tieu, Lyn, Jacopo Romoli, Peng Zhou & Stephen Crain. 2016. Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics* 33(2). 269–298. http://dx.doi.org/10.1093/jos/ffv00110.1093/jos/ffv001.

von Wright, Georg Henrik. 1968. *An essay in deontic logic and the general theory of action*. Amsterdam: North-Holland Publishing Company.

Willer, Malte. 2017. Widening free choice. In Alexandre Cremers, Thom van Gessel & Floris Roelofsen (eds.), *Proceedings of the 21st Amsterdam Colloquium*, 511–520.

Zimmerman, Thomas Ede. 2000. Free choice disjunction and epistemic possibility. *Natural Language Semantics* 8(255–290). http://dx.doi.org/10.1023/A:1011255819284.