What makes an inference robust?*

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 Sentences involving disjunction give rise to IGNORANCE, DISTRIBUTIVE and FREE CHOICE inferences. These inferences display certain similarities with regular Scalar Implicatures (SIs) and some researchers have proposed to treat them as such. This proposal, however, faces an important challenge: experimental results have shown that all three inferences are generally more robust, faster to process, and easier to acquire than regular SIs. A common response to this challenge in the literature is to hypothesise that such discrepancies among different types of SIs stem from the type of alternatives used to derive them: in contrast to regular SIs, IGNORANCE, DISTRIBUTIVE and FREE CHOICE inferences are computed on the basis of sub-constituent alternatives, which are alternatives that are formed without lexical substitution. This paper reports on a series of experiments that tested this hypothesis by comparing positive, disjunctive sentences giving rise to the three inference types of inferences to variants of these sentences involving either negation and conjunction, or negation and disjunction, for which the implicature approach predicts similar inferences on the basis of the same type of alternatives. Our results reveal that, while the three inferences are indeed quite robust in the disjunctive cases, regardless of whether negation is present or not, the inferences that their negative, conjunctive variants give rise to are not. These findings are challenging for the idea that the type of alternatives involved in SI computation (i.e., sub-constituent vs. lexical) is a major factor responsible for differences in robustness. We outline two possible alternative explanations of our data. One supplements the implicature approach with an extra assumption about how relevance is calculated for disjunction; the other is based on a non-implicature approach to the cases involving disjunction.

Keywords: free choice, distributive inferences, ignorance, alternatives, implicature

* For very helpful discussion and feedback, we would like to thank Moysh Bar-Lev, Kyle Blumberg, Danny Fox, Simon Goldstein, Matt Mandelkern, Uli Sauerland, and the audience at the Scales, degrees, and implicature Workshop at Potsdam University. This research was supported by the Leverhulme Trust grant RPG-2018-425.

Contents

1 Introduction

Sentences involving matrix or embedded disjunctions like those in (1a)–(3a) have long been observed to give rise to the inferences indicated. The inference type in (1b) is generally described and referred to as an IGNORANCE inference (henceforth, II), the one in (2b) as a FREE CHOICE inference (henceforth, FC), and the one in (3b) as a DISTRIBUTIVE inference (henceforth, DI).[1]

(1) IGNORANCE INFERENCE (II)
  a. The box contains either a blue ball or a yellow ball.
  b. ⇝ *The speaker doesn't know whether the box contains a blue ball and she doesn't know whether it contains a yellow ball*

(2) FREE CHOICE (FC)
  a. It is possible that the box contains either a blue ball or a yellow ball.
  b. ⇝ *It's possible that the box contains a blue ball and it's possible that it contains a yellow ball*

(3) DISTRIBUTIVE INFERENCE (DI)
  a. It is certain that the box contains either a blue ball or a yellow ball.
  b. ⇝ *It's not certain that the box contains a blue ball and it's not certain that it contains a yellow ball*

These three inferences have all been observed to have important similarities with regular Scalar Implicatures (henceforth, SIs) like the one in (4b), which typically arises from a simple sentence containing a possibility modal like the sentence in (4a).

(4) SCALAR IMPLICATURE (SI)
  a. It is possible that the box contains a blue ball.
  b. ⇝ *It's not certain that the box contains a blue ball*

One of the central characteristics of SIs is that they tend not to arise in downward entailing contexts. Thus for instance, the SI in (4b) does not arise under negation, as exemplified in (5). Note that, if it did, the sentence would then mean that it is certain (or impossible, if the modal is read deontically; we'll focus on the epistemic reading here) that the box contains a blue ball. On this reading, (5) should be judged as true in situations where it is certain that the box contains a blue ball, against intuitions.

1 For ignorance inferences, see Gazdar 1979, Sauerland 2004, Fox 2007, Meyer 2013 among others; for free choice, see Kamp 1974, 1978 and much subsequent work; for distributive inferences, see Crnič, Chemla & Fox 2015, Chierchia 2013, Santorio & Romoli 2017 among others.

(5) It is not possible that the box contains a blue ball.
$\not\leadsto$ *It's not possible that the box contains a blue ball or it is certain that it does*

FC, DI and II also exhibit this property. To illustrate, consider the case of FC. If FC arose under negation, the sentence in (6) would be compatible with it being possible that the box contains a ball that is either blue or yellow. The most prominent interpretation of the sentence is yet one which suggests instead that the box contains neither a blue ball, nor a yellow one (a reading sometimes referred to as DOUBLE PROHIBITION).[2]

(6) It is not possible that the box contains either a blue ball or a yellow ball.
$\not\leadsto$ *It's not possible that the box contains a blue ball or it's not possible that it contains a yellow ball*

Similar observations hold for DI and II. Intuitively, the sentence in (7) conveys that it is not certain that the box contains a blue ball and that it is not certain that it contains a yellow one while the sentence in (8) conveys that the speaker believes that the box contains neither a blue ball nor a yellow one.

(7) It is not certain that the box contains either a blue ball or a yellow ball.

(8) The box does not contain either a blue ball or a yellow ball.

A prominent approach in the literature treats FC, DI and II as SIs. This approach has the advantage of immediately capturing the similarities between these inferences and regular SIs while maintaining standard approaches to modals and disjunction. An important challenge for this approach, however, comes from recent experimental evidence that FC, DI and II differ from regular SIs in terms of robustness, processing and acquisition. In particular, FC has been shown to be much more robust than regular SIs (Chemla 2009, Marty et al. 2021), to be derived faster (Chemla & Bott 2014, Van Tiel & Schaeken 2017), and to be acquired earlier (Tieu et al. 2016). While DI and II have been comparatively less studied, similar differences have been found between them and regular SIs in terms of processing speed and acquisition (Van Tiel & Schaeken 2017, Pagliarini et al. 2018). These differences constitute an important challenge for the implicature approach and, more generally, for any uniform account treating all these inference types alike.

In response to this challenge, it has been proposed that the discrepancies observed between FC, DI, and II, on one side, and regular SIs, on the other, stems from the nature of the alternatives involved in the derivation of these inferences. The proposal is based on the

2 Enguehard & Chemla (2021) argue that free choice inferences are more easily embeddable in downward entailing contexts than regular SIs. They suggest that this fact teaches us that these inferences have distinct distributions, but that it is nonetheless compatible with an implicature approach to free choice. These differences between SIs and free choice are tangential to our main concerns, so we will leave them aside.

observation that while both types of inferences require substitution of material present in the assertion, SI requires an additional step of accessing material from the lexicon. With this in mind, it has been hypothesised that SIs based on alternatives which do not require lexical access are more readily available, easier to process and acquired earlier. Once supplemented with this hypothesis about alternatives, the uniform treatment offered by the implicature approach to FC, DI and II can be reconciled with the differences observed in the experimental literature.

In this paper, we investigate the empirical adequacy of this hypothesis by looking at related sentences such as (9a)–(11a), which correspond to the negative counterparts of the examples in (1a)–(3a). As we discuss below, these sentences are predicted on the implicature approach to give rise to the inferences indicated right below them. The inference in (10b) is generally called NEGATIVE FREE CHOICE (Fox 2007, Chemla 2009, Marty et al. 2021). To our knowledge, the inferences in (9b) and (11b) have not been previously discussed; in analogy to FREE CHOICE VS. NEGATIVE FREE CHOICE, we will simply call them NEGATIVE IGNORANCE and NEGATIVE DISTRIBUTIVE inferences. As we will see, the implicature approach derives these inferences in the same way as their positive counterparts in (1b)–(3b), on the basis of the same type of non-lexical alternatives, thus predicting these inferences to behave like positive FC, DI and II, and unlike regular SIs.

(9) NEGATIVE IGNORANCE INFERENCE (NEGATIVE II)
   a. The box does not contain both a blue ball and a yellow ball.
   b. ⇝*The speaker doesn't know whether the box contains a blue ball and she doesn't know whether it contains a yellow ball*

(10) NEGATIVE FREE CHOICE (NEGATIVE FC)
   a. It is not certain that the box contains both a blue ball and a yellow ball.
   b. ⇝*It's not certain that the box contains a blue ball and it's not certain that it contains a yellow ball*

(11) NEGATIVE DISTRIBUTIVE INFERENCE (NEGATIVE DI)
   a. It is not possible that the box contains both a blue ball and a yellow ball.
   b. ⇝*It's possible that the box contains a blue ball and it's possible that it contains a yellow ball*

We report on a series of experiments in which we tested the above positive and negative versions of the three inference types of interests and compared them to regular SIs. For the negative cases, we further manipulated the position of negation across experiments, testing logically equivalent sentences involving either negation and conjunction or negation and disjunction. Concretely, for each of the three inference types, we tested the high negative cases above and two of their variants: their intermediate negative variant, where negation appears right below the modal, and their low negative variant, where negation is pushed further down below the embedded connective, as exemplified in (12) and (13) for negative FC.

(12) INTERMEDIATE NEGATIVE FC

a. It is possible that the box does not contain both a blue ball and a yellow ball.
b. ⇝*It's possible that the box does not contain a blue ball and it is possible that it does not contain a yellow ball*

(13) LOW NEGATIVE FC

a. It is possible that the box does not contain a blue ball or does not contain a yellow ball.
b. ⇝*It's possible that the box does not contain a blue ball and it is possible that it does not contain a yellow ball*

Our results reveal that low negative cases (e.g., (13a)) readily give rise to the expected inferences, similarly to their positive counterparts. By contrast, there is no evidence in our data for the existence of similar inferences in high negative cases (e.g., (10a)) or their intermediate negative variants (e.g., (12a), where negation is below the modal and scopes over conjunction. In sum, our results show that, while the three inferences are quite robust in all the disjunctive cases, whether or not negation is present, those of their negative (high or intermediate) conjunctive variants are not. These findings are challenging for the idea that the type of alternatives involved in SI computation (i.e., lexical vs. sub-constituent alternatives) plays a major role in explaining the differences in robustness previously observed between FC, DI, II and regular SIs. Moreover, we will discuss how an additional hypothesis about alternatives recently suggested by Bar-Lev (2018) and Bar-Lev & Fox (2020) fares with our results, as well as certain non-implicature approaches to FC, DI and II, which make different predictions for these additional cases.

We end the article by outlining two promising accounts of our results. The first one retains an implicature-based approach to FC, DI and II, but supplements this approach with an additional assumption about the calculation of relevance for disjunctive sentences. The other relies on a non-implicature approach to the cases involving disjunction, but retains an implicature approach for the other cases involving conjunction. Regardless of the direction one takes, we show that both accounts offer novel insights about these inferences by refining either the role relevance plays in their derivation or the nature of the alternatives they arise from.

The rest of the article is organised as follows. In the following section (Section 2), we describe in more detail the three inferences of interest and their derivation on the implicature approach before discussing the alternative-based hypothesis and its predictions for the negative cases. In Section 3, we report on our experiments. In Section 4, we discuss the novel challenge coming from our results and its consequences for current theories and, in Section 5, we sketch two possible accounts of our results. Section 6 concludes.

2 Background

In this section, we start by sketching the implicature approach to FC, DI and II. We then discuss the challenge coming from the observed differences between these inferences and regular SIs and as well as its response in the literature. We will outline the alternative-based hypothesis and the predictions it makes for the novel cases we are interested in.

2.1 The implicature approach to Free choice, Distributive, and Ignorance

Implicatures were first extensively discussed by Grice (1975), who also coined the term. From the start, the question of how implicatures exactly arise has been debated. This has been particularly the case for the so-called SCALAR IMPLICATURES like the one in (4), repeated from above. In a nutshell, the question is whether these inferences arise mainly from the pragmatic side of the semantics-pragmatics interface, as a result of the listener's reasoning over the speaker's communicative intentions, or whether they are actually part of the truth-conditions of a strengthened meaning of the sentences they arise from.[3]

(4) SCALAR IMPLICATURE (SI)

a. It is possible that the box contains a blue ball.

b. ⇝ *It's not certain that the box contains a blue ball*

In the following, we sketch a version of the latter approach and explain how it extends to FC, II and DI. We note, however, that this choice is just for concreteness and that the debate between the two approaches is not crucial for our present purposes.

2.1.1 Basic ingredients

On the more semantic approach, SIs are assumed to arise from the application of a silent exhaustification operator, generally referred to as EXH. In a nutshell, this operator combines with a sentence and returns the meaning of that sentence together with its implicatures. More precisely, EXH takes as arguments a sentence and a set of contextually relevant alternatives, and it returns the conjunction of that sentence with the negation of a subset of its relevant alternatives, namely these relevant alternatives that are 'innocently excludable'. In effect what EXH does is strengthening the meaning of the sentence as much as possible, while avoiding contradictions and arbitrary choices between alternatives. The meaning of EXH is given in (14) and the definition of innocent exclusion in (15), where '*C*' stands for the set of salient alternatives.

(14) $[\![\textsf{EXH}]\!](A)(p)(w) = p_w \wedge \forall q \in \textsf{IE}(p,A)[\neg q_w]$

3 See, e.g., Sauerland 2004, Chierchia, Fox & Spector 2012, Franke 2011, Chemla 2010, Geurts 2010.

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

With these definitions in place, we can now illustrate how regular SIs are derived on this approach. For these purposes, consider the possible parse of (4) in (16), which involves EXH. The alternatives for EXH's prejacent are given in (17).[4] In this case, there is only one maximal excludable subset which includes only one alternative, namely the labeled as $\Box$a below. Excluding this alternative gives rise to the intuitively correct implicature in (4b), i.e., *it is not certain that the box contains a blue ball.*

(16) EXH[It is possible that the box contains a blue ball]

(17) $\left\{ \begin{array}{ll} \text{It is possible that the box contains a blue ball} & \Diamond\mathsf{a} \\ \text{It is certain that the box contains a blue ball} & \Box\mathsf{a} \end{array} \right\}$

The meaning strengthening mechanism described here is very general and its application extends to a variety of simple and more complex cases involving regular SIs. In the following, we explain in turn how it can be used to derive DI, FC and II.

2.1.2 Distributive inferences

Consider again the genuine case of DI in (2), repeated below for convenience.

(2) DISTRIBUTIVE INFERENCE (DI)

a. It is certain that the box contains either a blue ball or a yellow ball.

b. ⇝ *It's not certain that the box contains a blue ball and it's not certain that it contains a yellow ball* $\neg\Box\mathsf{a} \wedge \neg\Box\mathsf{b}$

The literal meaning of (2a) is (schematically) $\Box(\mathsf{a} \vee \mathsf{b})$, and it does not entail the distributivity inference and is compatible with either of the disjuncts being certain. However, if (2) is parsed with EXH, as shown in (18), then the inference in (2b) arises as implicatures. To illustrate, assume that the prejacent of EXH, (2a), has the alternatives in (19).

(18) EXH[It is certain that the box contains a blue ball or a yellow ball]

(19) $\left\{ \begin{array}{ll} \text{It is certain that the box contains a blue ball or a yellow ball} & \Box(\mathsf{a} \vee \mathsf{b}) \\ \text{It is certain that the box contains a blue ball} & \Box\mathsf{a} \\ \text{It is certain that the box contains a yellow ball} & \Box\mathsf{b} \\ \text{It is certain that the box contains a blue ball and a yellow ball} & \Box(\mathsf{a} \wedge \mathsf{b}) \end{array} \right\}$

4 While the question of how alternatives are determined remains very debated, most accounts assume that the alternatives of (16) include those in (17); see Breheny et al. 2018 and references therein for discussion.

In this case, all the alternatives in (19) but the prejacent are innocently excludable. The exclusion of these alternatives, and in particular the exclusion of the two alternatives corresponding to the independent disjuncts, $\Box a$ and $\Box b$, yields the desired distributive meaning, as shown in (20).[5]

(20) ⟦EXH[It is certain that the box contains a blue ball or a yellow ball]⟧ =
$\Box(a \vee b) \wedge \neg\Box(a \wedge b) \wedge \boxed{\neg\Box a \wedge \neg\Box b}$

In sum, a theory of scalar implicatures can account for distributive inferences. We illustrated this result using a semantic approach to implicatures, but nothing hinges on this presentation choice. As one can verify, the key observation here is that distributive inferences can be derived as SIs as long as each disjunct in the scope of the modal (here, $\Box a$ and $\Box b$) is taken to be an alternative. We turn next to the case of free choice.

2.1.3 Free choice inferences

The exhaustification-based process deriving regular SIs does not immediately derive FC inferences, but it can be amended (and has been amended) in various ways to achieve these purposes (Fox 2007, Klinedinst 2007, Santorio & Romoli 2017, Bar-Lev 2018, Bar-Lev & Fox 2020, Chemla 2010: among others).

(1) FREE CHOICE (FC)

a. It is possible that the box contains either a blue ball or a yellow ball.

b. ⇝ *It's possible that the box contains a blue ball and it's possible that it contains a yellow ball* $\Diamond a \wedge \Diamond b$

Most prominently, Fox (2007) proposes that FC is a recursive or higher order implicature, arising through two successive applications of EXH. More recently, Bar-Lev (2018) and Bar-Lev & Fox (2020) put forward an amendment to the definition of EXH so that EXH not only excludes alternatives but also includes some others. We use the latter account for illustrative purposes but, here again, we note that nothing hinges on this choice.

In addition to conjoining the prejacent with the negation of its innocently excludable alternatives, the exhaustivity operator also conjoins the prejacent with a subset of other alternatives, those that are 'innocently includable'. This conceptualisation of EXH relies on

5 We note that it is not entirely clear what exactly constitutes 'distributive inference´. Should we think of them as implying that each disjunct is not certain, as indicated in (20) above, or should we phrase them in terms of each being possible as in (i).

(i) $\Diamond a \wedge \Diamond b$

Note that the inferences in (i) follow from the assertion and those in (20). Here and in our experiments, we do not distinguish between these two options, but see Crnič, Chemla & Fox 2015 and Bar-Lev & Fox 2020 for discussion.

the characterisation of innocently includable alternatives in (21). Concretely, innocently includable alternatives are those alternatives that are in all maximal subsets of alternatives that can be conjoined consistently with the assertion and with the negation of all innocently excludable alternatives. Following this characterisation, the definition of EXH is then amended as shown in (22): EXH conjoins the prejacent with the negation of all innocently excludable alternatives, just as before, and now also conjoins it with all the innocently includable alternatives.

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

(22) $$[\![\text{EXH}]\!](C)(p)(w) = p_w \wedge \forall q \in \mathsf{IE}(p,C)[\neg q_w] \wedge \forall r \in \mathsf{II}(p,C)[r_w]$$

Assuming the novel definition in (22), FC inferences can be derived via a single application of EXH. To illustrate, consider the parse of (1a) in (23), where EXH occurs at matrix level. We assume that the alternatives to EXH's prejacent are those in (24).

(23) EXH[It is possible that the box contains a blue ball or a yellow ball]

(24) $$\left\{ \begin{array}{ll} \text{It is possible that the box contains a blue ball or a yellow ball} & \Diamond(\mathsf{a} \vee \mathsf{b}) \\ \text{It is possible that the box contains a blue ball} & \Diamond\mathsf{a} \\ \text{It is possible that the box contains a yellow ball} & \Diamond\mathsf{b} \\ \text{It is possible that the box contains a blue ball and a yellow ball} & \Diamond(\mathsf{a} \wedge \mathsf{b}) \end{array} \right\}$$

The conjunctive alternative, namely $\Diamond(\mathsf{a} \wedge \mathsf{b})$, is the only innocently excludable alternative in this case. But there is now another way whereby the basic meaning of the assertion can be strengthened, i.e., by identifying and including the innocently includable alternatives. As Bar-Lev & Fox (2020) show, there is one and only one subset of includable alternatives here, $\{\Diamond(\mathsf{a} \vee \mathsf{b}), \Diamond\mathsf{a}, \Diamond\mathsf{b}\}$. Including each of these alternatives gives us the FC inference we were after, as shown in (25).

(25) $[\![$EXH[It is possible that the box contains a blue ball or a yellow ball]$]\!]$ =
$\Diamond(\mathsf{a} \vee \mathsf{b}) \wedge \neg\Diamond(\mathsf{a} \wedge \mathsf{b}) \wedge \boxed{\Diamond\mathsf{a} \wedge \Diamond\mathsf{b}}$

In sum, the implicature approach can be extended to capture FC as well. Crucially, despite some differences between the derivations of FC and DI, we can observe that there is a noticeable similarity in the type of alternatives that these inferences are assumed to be derived from: just like DI, FC is derived in reference to those alternatives that correspond to the independent disjuncts (here, $\Diamond\mathsf{a}$ and $\Diamond\mathsf{b}$). We will now see that similar observations hold of ignorance inferences.

2.1.4 Ignorance inferences

To complete our overview, consider again the case of II in (3), repeated below. To formalise ignorance inferences, we will use, here and throughout the article, the conventional notation 'Ka' to represent that, according to what the speaker believes, a is true.

(3) Ignorance inference (II)
  a. The box contains either a blue ball or a yellow ball.
  b. ⇝ *The speaker doesn't know whether the box contains a blue ball and she doesn't know whether it contains a yellow ball* $\neg K\mathsf{a} \wedge \neg K\neg\mathsf{a} \wedge \neg K\mathsf{b} \wedge \neg K\neg\mathsf{b}$

There are two major accounts of ignorance inferences in the literature, both of which assume a standard Boolean meaning for disjunction. The first account relies on a cooperativity principle akin to the Gricean Maxim of Quantity whereby what is actually said is compared to whatever else the speaker could have said that would have also been relevant (Fox 2007, Sauerland 2004, Gazdar 1979 among others). For a sentence like (3a), the comparison set would include the propositions corresponding to the disjuncts as well as their negation, leading to the conclusion that the speaker doesn't know which of the two disjuncts is true, thereby deriving ignorance. The second account relies on the idea that every utterance involves a silent modal operator, notated 'K', ranging over the speaker Doxastic's worlds. We outline this account here for concreteness. As Meyer (2013), Fox (2016), and Buccola & Haida (2019) show, applying exhaustification both below and above the matrix K-operator, as shown in (26), delivers the ignorance inferences of interest.

(26) exh[K[exh[the box contains a blue ball or a yellow ball]]]

First, note that the most embedded occurrence of exh operates on the set of alternatives in (27). This first exhaustification layer gives rise to the meaning $K[(\mathsf{a} \vee \mathsf{b}) \wedge \neg(\mathsf{a} \wedge \mathsf{b})]$.

(27)
$$\left\{ \begin{array}{ll} \text{the box contains a blue ball or a yellow ball} & (\mathsf{a} \vee \mathsf{b}) \\ \text{the box contains a blue ball} & \mathsf{a} \\ \text{the box contains a yellow ball} & \mathsf{b} \\ \text{the box contains a blue ball and a yellow ball} & (\mathsf{a} \wedge \mathsf{b}) \end{array} \right\}$$

Next, the uppermost occurrence of exh quantifies over the alternatives in (28). This second exhaustification layer gives rise to the final meaning in (29), which entails the ignorance inferences indicated in the box. Once again, it is critical to observe that the alternatives needed to derive the relevant inferences are not lexical, but sub-constituent alternatives, corresponding here to the (exhaustified) disjuncts appearing in the scope of K.

(28) $\left\{ \begin{array}{ll} \text{K[EXH[the box contains a blue ball or a yellow ball]]} & K[(\mathsf{a} \vee \mathsf{b}) \wedge \neg(\mathsf{a} \wedge \mathsf{b})] \\ \text{K[EXH[the box contains a blue ball]]} & K[\mathsf{a} \wedge \neg \mathsf{b}] \\ \text{K[EXH[the box contains a yellow ball]]} & K[\mathsf{b} \wedge \neg \mathsf{a}] \\ \text{K[EXH[the box contains a blue ball and a yellow ball]]} & K[\mathsf{a} \wedge \mathsf{b}] \end{array} \right\}$

(29) $[\![\text{EXH[K[EXH[the box contains a blue ball or a yellow ball]]]}]\!] =$
$K[(\mathsf{a} \vee \mathsf{b}) \wedge \neg K[\mathsf{a} \wedge \neg \mathsf{b}] \wedge \neg K[\mathsf{a} \wedge \neg \mathsf{b}] \wedge \neg K[\mathsf{a} \wedge \mathsf{b}] =$
$K[(\mathsf{a} \vee \mathsf{b}) \wedge \boxed{\neg K[\mathsf{a}] \wedge \neg K \neg \mathsf{a} \wedge \neg K[\mathsf{b}] \wedge \neg K \neg \mathsf{b}}$

To summarise, the implicature approach has been a prominent approach in the literature and it is shown to successfully account for the three inference types under investigation: DISTRIBUTIVE, FREE CHOICE and IGNORANCE inferences. In fact, these inferences, and FREE CHOICE in particular, are commonly taken as a fruitful testing ground for disentangling predictions of different theories of implicatures (Fox 2007, Geurts 2010, Franke 2011, Bar-Lev 2018, Chemla 2010, Marty & Romoli 2021 among others). As we shall now see, however, results from recent experimental studies have challenged the implicature approach to FC, DI and II based on certain discrepancies between them and regular SIs.

2.2 The challenge and the response

Experimental work has unveiled discrepancies in the processing and acquisition of FC, DI and II, compared to regular SIs. Arguably, FC offers a striking example of such discrepancies as it has been found to differ from regular SIs both in terms of processing and acquisition. First, FC has been shown to differ in its processing profile from regular SIs. For instance, Chemla & Bott 2014, building on Bott & Noveck 2004, found that responses based on FC interpretations are not slower than those based on the corresponding literal meanings, unlike what is usually found for regular SIs (see Van Tiel & Schaeken 2017 for similar results). Second, Tieu et al. (2016) found that 5-year-old children behave in an adult-like fashion with FC, unlike with regular SIs, indicating that the former type of inference is mastered earlier than the latter. Distributive and ignorance inferences have been much less studied, but similar differences in processing speed and acquisition have been found for them as well (Van Tiel & Schaeken 2017, Pagliarini et al. 2018). In sum, FC, DI and II all appear to be faster to process and easier to acquire than regular SIs. These discrepancies raise a challenge for any uniform approach treating these inferences like regular SIs.

Thus far, the response in the literature has been to relate the observed discrepancies to the nature of the alternatives involved in the derivation of these inferences. The gist of the idea is that the derivation of regular SIs involves alternatives which requires lexical substitution, while those of FC, DI and II do not. To illustrate, consider again the case in (4) repeated from above. As it is easy to see, in order to arrive at the expected result in (4b), speakers need to entertain the alternative in (30), which involves accessing the lexicon and substituting *possible* with *certain*.

(4) SCALAR IMPLICATURE (SI)

a. It is possible that the box contains a blue ball.

b. ⇝ *It's not certain that the box contains a blue ball*

(30) It is certain that the box contains a blue ball.

By contrast, none of the alternatives involved in the derivation of FC, DI or II require lexical substitution. Rather, the derivations of these inferences are all based on alternatives that are subconstituents of the asserted sentence. For instance, the distributive inference associated with (2a) is derived on the basis of the alternatives in blue in (31), corresponding to the disjuncts appearing in the scope of the modal.

(2) DISTRIBUTIVE INFERENCE (DI)

a. It is certain that the box contains either a blue ball or a yellow ball.

b. ⇝ *It's not certain that the box contains a blue ball and it's not certain that it contains a yellow ball* $\neg\Box a \wedge \neg\Box b$

(31) $$\left\{\begin{array}{ll} \text{It is certain that the box contains a blue ball or a yellow ball} & \Box(a \vee b) \\ \text{It is certain that the box contains a blue ball} & \Box a \\ \text{It is certain that the box contains a yellow ball} & \Box b \\ \text{It is certain that the box contains a blue ball and a yellow ball} & \Box(a \wedge b) \end{array}\right\}$$

Similar observations hold for FC, as evidenced below.

(1) FREE CHOICE (FC)

a. It is possible that the box contains either a blue ball or a yellow ball.

b. ⇝ *It's possible that the box contains a blue ball and it's possible that it contains a yellow ball* $\Diamond a \wedge \Diamond b$

(32) $$\left\{\begin{array}{ll} \text{It is possible that the box contains a blue ball or a yellow ball} & \Diamond(a \vee b) \\ \text{It is possible that the box contains a blue ball} & \Diamond a \\ \text{It is possible that the box contains a yellow ball} & \Diamond b \\ \text{It is possible that the box contains a blue ball and a yellow ball} & \Diamond(a \wedge b) \end{array}\right\}$$

The case of II is slightly more complex, as it involves two occurrences of the exhaustivity operator. Yet it is easy to verify that the alternatives needed in this case are also parts of the prejacent, as indicated below.

(33) EXH[K[EXH[the box contains a blue ball or a yellow ball]]]

(34)
$$\left\{\begin{array}{ll} \text{the box contains a blue ball or a yellow ball} & (\mathsf{a} \vee \mathsf{b}) \\ \text{the box contains a blue ball} & \mathsf{a} \\ \text{the box contains a yellow ball} & \mathsf{b} \\ \text{the box contains a blue ball and a yellow ball} & (\mathsf{a} \wedge \mathsf{b}) \end{array}\right\}$$

(35)
$$\left\{\begin{array}{ll} \text{K[}\textsc{exh}\text{[the box contains a blue ball or a yellow ball]]} & K[(\mathsf{a} \vee \mathsf{b}) \wedge \neg(\mathsf{a} \wedge \mathsf{b})] \\ \text{K[}\textsc{exh}\text{[the box contains a blue ball]]} & K[\mathsf{a} \wedge \neg \mathsf{b}] \\ \text{K[}\textsc{exh}\text{[the box contains a yellow ball]]} & K[\mathsf{b} \wedge \neg \mathsf{a}] \\ \text{K[}\textsc{exh}\text{[the box contains a blue ball and a yellow ball]]} & K[\mathsf{a} \wedge \mathsf{b})] \end{array}\right\}$$

In sum, the observation is that the kind of alternatives involved in the computation of regular SIs differ from that involved in the derivation of DI, FC and II: the former involve lexical substitutions while the latter don't. Based on this observation, researchers have hypothesized that appealing to lexical substitution when building alternatives is what slows down the processing of an SI and makes it harder to acquire. This hypothesis can be formulated as in (36) (Singh et al. 2016, Tieu et al. 2016, Pagliarini et al. 2018, Chemla & Bott 2014, Van Tiel & Schaeken 2017, Barner, Brooks & Bale 2011):

(36) **Alternative-based hypothesis**
Alternatives that do not involve lexical substitutions give rise to inferences that are faster to process and easier to acquire.

The Alternative-based hypothesis above has been primarily formulated in reference to processing and acquisition results. Yet another way in which FC, DI and II have been shown to differ from regular SIs is in their robustness (Chemla 2009, Van Tiel & Schaeken 2017, Marty et al. 2021). Thus, it is tempting to refine the original hypothesis along the lines of (37) so as to account for this difference as well and relate all of these properties to one single source having to do with the nature of the alternatives involved in the derivation of the inferences in question. In particular, we note that, although the cognitive cost of a certain inference is in principle independent from its robustness, it is reasonable to assume that both measures are positively correlated to some extent. It could be so for instance if (i) the lower the processing cost associated with a given inference, the more likely it is that this inference will be accessed by speakers, and (ii) when speakers perceive both the strong and weak reading of a given sentence, they preferentially resolve the ambiguity at hand by favoring the stronger over the weaker reading. These assumptions permit to relate ease of processing and robustness by linking the likelihood that a given inference be accessed by speakers, which partly depends on its processing cost (among other factors), to the likelihood that its corresponding reading be reported by speakers, which depends in turns on its accessibility.

(37) **Alternative-based hypothesis (extended version)**
Alternatives that do not involve lexical substitutions give rise to inferences that are more robust, faster to process, and easier to acquire.

The extended version of the Alternative-based hypothesis permits to reconcile the implicature approach with the results reported in the experimental literature. This hypothesis also makes novel predictions which we now turn to.

2.3 The negative cases: novel predictions

Once supplemented with the hypothesis in (37), the implicature approach predicts that inferences based on subconstituent alternatives should behave more like FC, DI and II, and unlike regular SIs. One straightforward way to test this general prediction is to look at the negative counterparts of (1a)–(3a) in (9)–(11) and their predicted inferences.

(10) NEGATIVE FREE CHOICE (HIGH NEGATIVE FC) $\neg\Box(\mathsf{a} \wedge \mathsf{b})$

a. It is not certain that the box contains both a blue ball and a yellow ball.
b. ⇝*It's not certain that the box contains a blue ball and it's not certain that it contains a yellow ball*

(11) NEGATIVE DISTRIBUTIVE INFERENCE (HIGH NEGATIVE DI) $\neg\Diamond(\mathsf{a} \wedge \mathsf{b})$

a. It is not possible that the box contains both a blue ball and a yellow ball.
b. ⇝*It's possible that the box contains a blue ball and it's possible that it contains a yellow ball*

(9) NEGATIVE IGNORANCE INFERENCE (HIGH NEGATIVE II) $\neg(\mathsf{a} \wedge \mathsf{b})$

a. The box does not contain both a blue ball and a yellow ball.
b. ⇝*The speaker doesn't know whether the box contains a blue ball and she doesn't know whether it contains a yellow ball*

These cases are particularly interesting for our purposes because the implicature approach derives negative FC, negative DI and negative II in the same way as their positive counterparts, on the basis of the same type of non-lexical alternatives. To illustrate, consider first the case of negative FC in (10). The literal meaning of this sentence, as predicted by standard approaches to modals and conjunction, is equivalent to $\neg\Box\mathsf{a} \vee \neg\Box\mathsf{b}$ and it is thus compatible with one of a and b being certain. However, in the same way as before, this literal meaning can be exhaustified as shown in (38), using the alternatives in (39) (where the critical alternatives are indicated in blue, as before).

(38) EXH[It is not certain that the box contains both a blue ball and a yellow ball]

(39) $\left\{\begin{array}{ll} \text{It is not certain that the box contains both a blue ball and a yellow ball} & \neg\Box(a \wedge b) \\ \text{It is not certain that the box contains a blue ball} & \neg\Box a \\ \text{It is not certain that the box contains a yellow ball} & \neg\Box b \\ \text{It is not certain that the box contains either a blue ball or a yellow ball} & \neg\Box(a \vee b) \end{array}\right\}$

In a similar way as for positive FC, $\neg\Box(a \vee b)$ is the only innocently excludable alternative while the other three are innocently includable, i.e., $\text{II} = \{\neg\Box(a \wedge b), \neg\Box a, \neg\Box b\}$. By including these alternatives, we obtain the negative free choice meaning of the sentence.

(40) $[\![\text{EXH}[\text{It is not certain that the box contains both a blue ball and a yellow ball}]]\!] =$
$\neg\Box(a \wedge b) \wedge \Box(a \vee b) \wedge \boxed{\neg\Box a \wedge \neg\Box b}$

Importantly, note that the alternatives over which negative FC is derived are parts of the asserted sentence and do not involve any lexical substitution. The same can be shown for the negative distributive inference in (11) and the negative ignorance inference in (9): the former is derived from the alternatives in (41) and the latter from those in (42) and (43).

(41) $\left\{\begin{array}{ll} \text{It is not possible that the box contains both a blue ball and a yellow ball} & \neg\Diamond(a \wedge b) \\ \text{It is not possible that the box contains a blue ball} & \neg\Diamond a \\ \text{It is not possible that the box contains a yellow ball} & \neg\Diamond b \\ \text{It is not possible that the box contains either a blue ball or a yellow ball} & \neg\Diamond(a \vee b) \end{array}\right\}$

(42) $\left\{\begin{array}{ll} \text{the box does not contain both a blue ball and a yellow ball} & \neg(a \wedge b) \\ \text{the box does not contain a blue ball} & \neg a \\ \text{the box does not contain a yellow ball} & \neg b \\ \text{the box does not contain either a blue ball or a yellow ball} & \neg(a \vee b) \end{array}\right\}$

(43) $\left\{\begin{array}{ll} \text{K[EXH[the box does not contain both a blue ball and a yellow ball]]} & K[\neg(a \wedge b) \wedge (a \vee b)] \\ \text{K[EXH[the box does not contain a blue ball]]} & K[\neg a \wedge b] \\ \text{K[EXH[the box does not contain a yellow ball]]} & K[\neg b \wedge a] \\ \text{K[EXH[the box does not contain either a blue ball or a yellow ball]]} & K[\neg(a \vee b)] \end{array}\right\}$

Finally, since the implicature approach works on logical relations between the asserted sentence and its alternatives, it predicts the availability of similar inferences for logically equivalent sentences. In particular, it predicts the same inferences for the sentence in (9) and its de-Morgan equivalent in (44), which we dub 'low' negative ignorance.

(44) LOW NEGATIVE II $\neg a \vee \neg b$

a. The box doesn't contain a blue ball or doesn't contain a yellow ball

b. $\rightsquigarrow$ *The speaker doesn't know whether the box contains a blue ball and she doesn't known whether it contains a yellow ball*

The same goes for (10) and (11) with respect to the equivalent 'intermediate' negative cases in (45a) and (46a), where negation appears just below the modal, and the corresponding 'low' negative versions in (47a) and (48a), where negation is embedded further down.

(45) INTERMEDIATE NEGATIVE FC $\Diamond\neg(a \wedge b)$

a. It is possible that the box does not contain both a blue ball and a yellow ball.

b. $\rightsquigarrow$ *It's possible that the box does not contain a blue ball and it's possible that it does not contain a yellow ball*

(46) INTERMEDIATE NEGATIVE DI $\Box\neg(a \wedge b)$

a. It is certain that the box does not contain both a blue ball and a yellow ball.

b. $\rightsquigarrow$ *It's certain that the box does not contain a blue ball and it's certain that it does not contain a yellow ball*

(47) LOW NEGATIVE FC $\Diamond(\neg a \vee \neg b)$

a. It is possible that the box either does not contain a blue ball or it does not contain a yellow ball.

b. $\rightsquigarrow$ *It's possible that the box does not contain a blue ball and it's possible that it does not contain a yellow ball*

(48) LOW NEGATIVE DI $\Box(\neg a \vee \neg b)$

a. It is certain that the box either does not contain a blue ball or it does not contain a yellow ball.

b. $\rightsquigarrow$ *It's certain that the box does not contain a blue ball and it's certain that it does not contain a yellow ball*

Table 1 summarises all the cases discussed so far, together with their predicted inferences, across the different sentence types and environments.

As discussed, the implicature approach predicts the negative inferences described in this section to be derived in the same way and on the basis of the same type of alternatives as their positive counterparts. Coupled with the Alternative-based hypothesis, this approach predicts these inferences will behave like their positive counterparts, and unlike regular implicatures, in terms of strength, processing, and acquisition. Moreover, these predictions do not change across the different versions of the negative cases we have described, i.e., whether negation appears high in the sentence, at an intermediate position below the modal, or within each disjunct. These predictions are summarised in Table 2. In the next section, we report on a series of experiments testing these predictions.

3 Experiments

Some of the cases we are interested in have been investigated in two previous studies. The first one is Chemla 2009, which compared positive and negative FC in simple and quan-

Environment	Sentence type FC	DI	II	SI
POSITIVE	$\Diamond(a \vee b)$	$\Box(a \vee b)$	$(a \vee b)$	$\Diamond a$
Inference:	$\Diamond a \wedge \Diamond b$	$\neg\Box a \wedge \neg\Box b$	$Ia \wedge Ib$	$\neg\Box a$
NEGATIVE				
High	$\neg\Box(a \wedge b)$	$\neg\Diamond(a \wedge b)$	$\neg(a \wedge b)$	$\neg\Box a$
Intermediate	$\Diamond\neg(a \wedge b)$	$\Box\neg(a \wedge b)$		
Low	$\Diamond(\neg a \vee \neg b)$	$\Box(\neg a \vee \neg b)$	$(\neg a \vee \neg b)$	$\Diamond\neg a$
Inference:	$\neg\Box b \wedge \neg\Box a$	$\Diamond a \wedge \Diamond b$	$Ia \wedge Ib$	$\Diamond a$

Table 1 Overview of the cases investigated in this study, alongside with their predicted inferences. We use the conventional notation Ia to indicate that the speaker is ignorant as to whether a is true (i.e., $Ia \equiv \neg Ka \wedge \neg K\neg a$). Note that the positive and the low negative cases involve disjunction, while the other cases involve conjunction.

Environment	Sentence type FC	DI	II	SI
POSITIVE	✓	✓	✓	✓
NEGATIVE				
High	✓	✓	✓	✓
Intermediate	✓	✓		
Low	✓	✓	✓	✓

Table 2 Predictions of the implicature approach, supplemented with the Alternative-based hypothesis in (37). A checkmark indicates that the relevant inference is predicted to be available; a dark green checkmark ✓ indicates that a robust inference is predicted, while a light green one ✓ stands for a weaker/regular inference.

tificational environments. The second is Marty et al. 2021, which tested positive and negative FC against various baselines and compared them to positive and negative SIs. Both of these studies found a clear difference in inference strength between positive FC and high negative FC cases involving deontic modalities (e.g., *allowed/required*). The present study expands on these initial investigations in two main ways. First, it extends the empirical scope by looking at the effect of sentence polarity across four inferences types – FC, II, DI and SI – and by testing other modals expressing epistemic possibility and necessity (i.e., *possible/certain, might/must*). Second, the position of negation in the negative cases was manipulated across experiments so as to cover the whole range of constructions presented in Table 1 – high, intermediate and low negative cases. By adding these novel cases to the set of comparison points, our goal was to investigate the effect of sentence polarity on inference strength in a more systematic fashion and, ultimately, reach a more complete and accurate description of this effect. As we discussed, the resulting picture should allow us to asses the generality of this phenomenon, evaluate the challenge it constitutes for the Extended Alternative-based hypothesis and, finally, test the predictions of certain non-implicature approaches to FC, DI and II.

In the following sections, we report on the four experiments we carried out to achieve these purposes. All four experiments involved a sentence-picture acceptability task where participants were presented with sentence-picture items like the two examples in Fig. 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 on one of two response buttons, labelled 'Good' and 'Bad', respectively. In the critical conditions, the test sentences were paired with pictures that make them false if the target inference is derived, but true if not, as illustrated in Fig. 1 for positive FC vs. high negative FC. The linking hypothesis was that the rate of acceptance ('Good' responses) observed in these conditions inform us about the robustness of the target inference: the more robust a given inference is (FC, DI, II or SI), the less participants should select the 'Good' response option in these conditions and, consequently, the lower the acceptance rate should be.

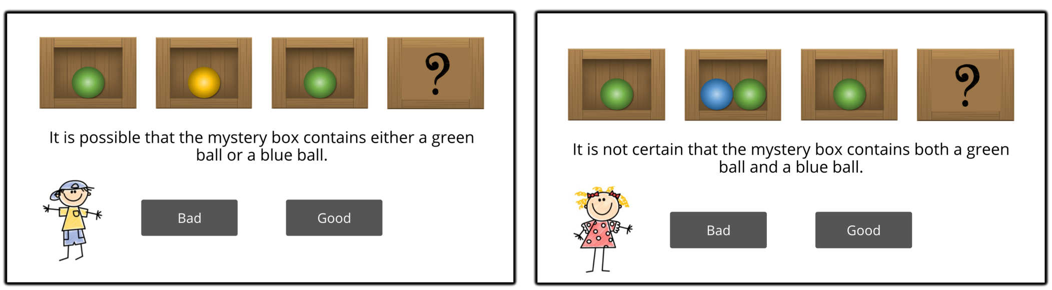


Figure 1 Examples of sentence-picture items used in our experiments. These examples correspond to target trials for the positive and negative FC sentences in Experiment 1. Participants were instructed that the rightmost box, called *the mystery box*, always had the same contents as one of the three open boxes (see Section 3.1.3 and Appendix A.1).

For the sake of comparison, all four experiments were designed in a similar fashion and differed from one another along one unique dimension, the level of negation (high, intermediate, low). The level of negation was achieved by manipulating a combination of the structure (position of negation) or the lexical contents (choice of modals and connectives) of the sentences presented to participants. Experiment 1 tested the predictions from the

Alternative-based hypothesis (see Table 2) by investigating the positive and high negative versions of FC, DI and II, and by comparing them to positive and negative SI. Experiment 2 followed up on Experiment 1 by further investigating the intermediate negative cases for FC and DI, where negation is pushed below the modal, as well as the low negative case of II. Experiment 3 was based on the same logic as Experiment 2, but tested a different set of modals. Finally, Experiment 4 tested the low negative cases for FC, DI and II.

3.1 Experiment 1: High negation

3.1.1 Participants

70 participants (average age 34.6 yrs; 41 female) were recruited online using Prolific (country of residence: UK; country of birth: UK; first language: English; minimum prior approval rate: 90%; vision: see colours normally). Participants were paid £1.75 for their participation and average completion time was about 12 minutes (£8.87/hr). All participants gave written informed consent to the processing of their information for the purposes of this study, which was approved by the institutions's Research Ethics Commitee. 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.

3.1.2 Materials

Each trial involved a sentence presented below a picture (see Figure 1). Sentences were constructed using one of the eight frames given in Table 3. There were four positive sentences, one for each of the four inference types of interest (FC, DI, II and SI), and four related negative sentences, obtained from their positive counterparts by adding negation at matrix level and by replacing the embedded modals and connectives with their scalemates (i.e., *possible*→*certain*, *certain*→*possible*, *or*→*and*). The [A] and [B] terms were two different color adjectives among four possible options: *green*, *blue*, *yellow* or *grey*.

Every picture displayed a quadruplet of boxes horizontally arranged. Each quadruplet was made of three open boxes, containing one or two balls, and a closed one, placed at the rightmost position, marked with the symbol "?" and referred to as *the mystery box*. The contents of the open boxes on each picture were determined in reference to the sentence it was paired with, and their contents were experimentally manipulated so as to create true, false and target pictures for each sentence type, as described and illustrated in Table 4. Target pictures were designed so as to make the relevant sentence false if the inference of interest (i.e., FC, DI, II or SI) is present, but true if it is absent; by contrast, true and false pictures were designed so as to make the relevant sentence respectively true and false, independently of the inferences under scrutiny.[6]

6 In designing the pictures, further precautions were taken, on a case-by-case basis, to prevent other commonly observed implicatures from affecting participants' judgments. In particular, for the pos-

		Schematic description of the target sentences
POSITIVE		
	FC	It is possible that the mystery box contains either a [A] ball or a [B] ball.
	DI	It is certain that the mystery box contains either a [A] ball or a [B] ball.
	II	The mystery box contains either a [A] ball or a [B] ball.
	SI	It is possible that the mystery box contains a [A] ball.
NEGATIVE		
	FC	It is not certain that the mystery box contains both a [A] ball and a [B] ball.
	DI	It is not possible that the mystery box contains both a [A] ball and a [B] ball.
	II	The mystery box does not contain both a [A] ball and a [B] ball.
	SI	It is not certain that the mystery box contains a [A] ball.

Table 3 Schematic description of the sentences tested Experiment 1 by sentence polarity and inference type. [A] and [B] correspond to different colour adjectives; for a more concrete illustration, you may read [A] as *green* and [B] as *blue*.

Each sentence type was paired with all three picture types, giving rise to the eponymous True, False and Target conditions. Each condition was iterated three times, leading to a total of 72 test trials. For each trial, the color adjectives used in the sentence was picked at random from our list of color terms, with replacement across trials. The color of the balls on the accompanying picture was determined according to the relevant sentence and the relevant condition: the colors of A-balls and B-balls always matched the color terms used in the sentence; the colors of the C-balls and D-balls were randomly chosen from our list by excluding the color(s) of the matching balls.

3.1.3 Procedure

The experiment was run as an online survey using Gorilla Experiment Builder (Anwyl-Irvine et al. 2020). Participants were introduced in the instructions to two characters, Sam and Mia, and they were presented with a short cover story. The cover story was as follows (see Appendix A.1 for details): Sam and Mia are looking at quadruplets of boxes containing balls of various colors. For each quadruplet, they can only see what's inside the first three boxes. However, they know that the fourth box always has the same contents as one of the three open boxes and, therefore, they can make certain inferences about what's inside this mystery box. Participants were shown two example quadruplets where the contents

itive and negative FC-sentences, all the pictures were designed so as to be compatible with the regular direct and indirect SIs that may arise from these sentences, i.e., the inference that *it's not certain that (A or B)* for the positive form and the inference that *it's possible that (A and B)* for the negative one.

		Description of the pictures by experimental condition		
		TRUE	FALSE	TARGET
POSITIVE				
FC	◊(A ∨ B)	A C B ?	D C D ?	A C A ?
DI	□(A ∨ B)	A AB B ?	A C B ?	A A A ?
II	(A ∨ B)	A B A ?	D C D ?	A A A ?
SI	◊A	A D C ?	D D D ?	A A A ?
NEGATIVE				
FC	¬□(A ∧ B)	A AB B ?	AB AB AB ?	A AB A ?
DI	¬◊(A ∧ B)	A B A ?	A AB B ?	A A A ?
II	¬(A ∧ B)	A B A ?	AB AB AB ?	A A A ?
SI	¬□A	A D A ?	A A A ?	D D D ?

Table 4 Schematic description and illustration of the picture types used in Experiment 1. The color of the A-balls and B-balls always matched the color adjectives used in the sentence (e.g., *green* and *blue*) while the color of the C-balls and D-balls never did (e.g., *grey* and *yellow*). Picture types are illustrated here using the following color assignment: A=green, B=blue, C=yellow and D=grey.

of the mystery box were progressively revealed, allowing them to verify that its contents were indeed identical to those of one of the open boxes. Participants were told that they would see many quadruplets like these, each of which would be followed by an utterance from either Sam or Mia about what the mystery box contains, and that their task was to decide if this utterance is or is not a good description of what's inside the mystery box. They were instructed to click on 'Good' if they consider the utterance a good description of the picture they see and otherwise to click on 'Bad'.

Following the instructions, participants started the survey with a short training devised to consolidate their understanding of the cover story and instructions (see Appendix A.2 for details). After the training phase, the survey continued with a block of 72 test trials.

Trials were presented in random order, with a 1000 ms interstimulus interval. Participants reported their judgments by clicking with the mouse one of two response buttons labelled 'Good' and 'Bad', respectively. The position of the response buttons (i.e., on the left or on the right) was counterbalanced across participants. Items remained on the screen until participants gave their response. At the end of the survey, participants were asked to fill out a short demographic questionnaire.

3.1.4 Data treatment and analyses

Responses from 5 participants were excluded prior to analyses because their overall performance in the True and False trials did not reach the threshold of 70% accuracy we had pre-established. In total, 360 out of 5,040 responses were removed through this procedure (7% of the data). Data were analyzed by modeling response-type likelihood using logit mixed effects regression models (Jaeger 2008). Analyses were conducted using the lme4 (Bates, Maechler & Bolker 2011, Bates et al. 2014) and emmeans (Lenth et al. 2018) libraries for the R statistics program (R Core Team 2013).

We planned our analyses to address the following three questions. First, are the classical instances of DI and II more robust than direct SIs, as previously observed for FC? Second, do these contrasts extend to the negative forms of these inferences? That is, are the negative variants of FC, DI, and II more robust than indirect SIs? Third, are these inferences as robust in their negative than in their positive forms? To address the first two questions, we assessed whether participants' responses in the True and in the Target conditions differ as a function of the inference type for both the positive and the negative sentences. The models included Condition (2 levels: True, Target), Inference type (4 levels: SI, FC, DI, II) and their interaction as fixed effects, with Subject as a random effect and a by-Subject random slope for Condition. Condition was coded with sum contrasts and Inference type was coded with treatment contrasts using SI as a reference level, thus allowing us to use the regression output to compare SI to all three other inference types. To address the third question, we assessed, for each of the four inference types, whether participants' responses in the True and in the Target conditions differ as a function of the sentence polarity. The models included Condition (2 levels: True, Target; sum-contrasts), Sentence polarity (2 levels: Positive, Negative; sum-contrasts) and their interaction as fixed effects, with the same random effect structures as the models previously described.

3.1.5 Results

Figure 2 shows the mean rates of acceptance for the target sentences by sentence polarity, inference type and experimental condition. In the following, we report on the two sets of analyses that we carried out (see 3.1.4 above for discussion).

In the first analysis, the model for positive sentences yielded a main effect of Condition ($\chi^2(1) = 31.08$, $p < .001$), Inference type ($\chi^2(3) = 91.24$, $p < .001$) and a significant interaction between the two ($\chi^2(3) = 82.31$, $p < .001$). The difference between True and

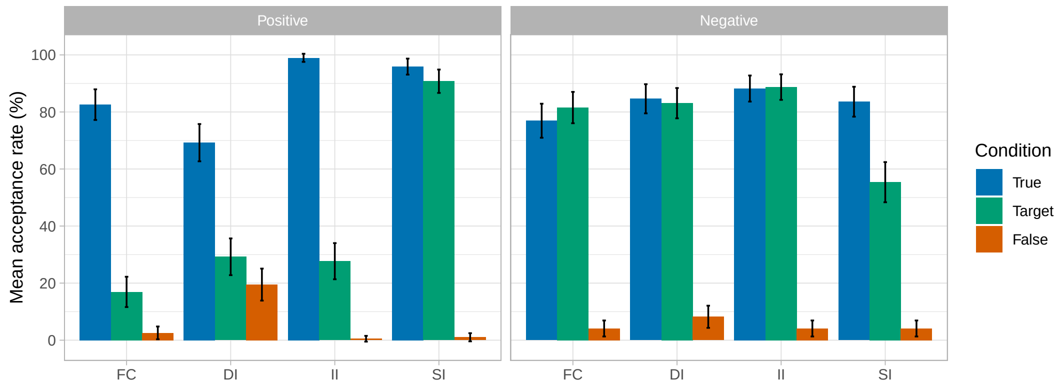


Figure 2 Mean proportion of 'Good' responses to each sentence type in Experiment 1 by sentence polarity (Positive, Negative), inference type (FC, DI, II, SI) and experimental condition (TRUE, TARGET, FALSE). Error bars represent 95% CIs.

TARGET conditions was significantly smaller for SI than for FC ($\beta = 3.30$, $p < .001$), DI ($\beta = 2.12$, $p < .001$) and II ($\beta = 4.21$, $p < .001$). The model for negative sentences yielded the same main results as the previous one; that is, there was a main effect of Condition ($\chi^2(1) = 12.07$, $p < .001$), Inference type ($\chi^2(3) = 49.52$, $p < .001$) and a significant interaction between them ($\chi^2(3) = 28.79$, $p < .001$). This time, however, the difference between TRUE and TARGET conditions was significantly larger for SI than for FC ($\beta = -0.91$, $p < .001$), DI ($\beta = -0.67$, $p < .001$) and II ($\beta = -0.74$, $p < .001$).

Turning to the second analysis, the models yielded a main effect of Polarity for each inference type (Negative>Positive; all χ^2s$>$ 28, all ps$<$ .001), a main effect of Condition for FC, DI and SI (TRUE>TARGET; all χ^2s$>$ 14, all ps$<$ .001; II: $\chi^2(1) = 1.65$, $p = .19$) and a signification interaction between both factors for FC, DI and II (all χ^2s$>$ 34, all ps$<$ .001; SI: $\chi^2(1) = 2.69$, $p = .1$) such that, for these three inference types, the difference between TRUE and TARGET conditions was significantly larger for the positive sentences than for their negative counterparts. Effects of Condition were investigated further in a post-hoc analysis comparing the estimated marginal means for the TRUE and TARGET conditions in the positive and in the negative cases (p-value adjusted for multiple comparisons). Results revealed that, in the positive cases, there was a significant contrast between TRUE and TARGET conditions for all inference types except SI; in the negative cases, the situation was reversed in that there was a significant contrast between both conditions only for SI.

3.1.6 Discussion

The contribution of the results of this first experiment is twofold. First, the results reproduce previous findings from the literature regarding the strength of positive and negative

FC compared to direct and indirect SIs. Specifically, our results replicate the main findings from Marty et al. 2021 (Experiment 3; see also Chemla 2009) in showing that positive instances of FC are more robust than positive SI whereas high negative instances of FC are less robust than negative SI. It bears pointing out that the relevant contrasts are observed here for a different scale (i.e., ⟨possible, certain⟩ instead of ⟨allow, require⟩) using different materials and task, suggesting that these contrasts are not peculiar to specific test items. In addition, as discussed in Marty et al. 2021, the finding of an interaction between sentence polarity and inference type for the comparison at hand tells us that the discrepancy observed between positive and negative FC cannot simply result from a general difference in robustness between the implicatures triggered by the use of the weaker scalar term (i.e., *possible*) and that of its (negated) stronger scale-mate (i.e., *not certain*).

Next, and more interestingly, our results bring something new to the debate: they show that the difference in strength observed between positive vs. negative free choice extends to distributive and ignorance inferences. This is clearly established in our results by the fact that the patterns of responses for FC, DI and II, together with their modulations across sentence polarity, display very strong similarities on all comparisons of interest. Specifically, positive FC, DI and II sentences were all strongly rejected by the participants in their TARGET conditions (all Ms$< 30\%$), unlike positive SI sentences, which were strongly accepted ($M = 90\%$); on the other hand, their high negative variants were all strongly accepted in their corresponding TARGET conditions (all Ms$> 80\%$), unlike negative SI sentences, which gave rise to an intermediate acceptance rate ($M = 55\%$). In sum, the negative variants of FC, DI and II tested in this experiment were all found to be far less robust than their positive counterparts but also slightly less robust than negative SI. These findings are not in line with the predictions of the alternative-based Hypothesis (see Table 2) insofar as the expected robustness of high negative FC, DI and II is concerned.

Before we turn to our second experiment, let us discuss two more points that we believe deserve further attention. The first one pertains to the general prediction from the implicature approach that negative FC, DI and II sentences have the potential to give rise to the eponymous inferences. As we explained, this prediction follows from the fact that, everything else being equal, the mechanisms giving rise to FC, DI and II in the positive cases should apply to their negative variants in a similar way, predicting the existence of comparable inferences. Given our linking hypothesis, this means that one could have expected high negative FC, DI and II sentences to give rise in their TARGET conditions to acceptance rates somewhat intermediate between those observed in their TRUE and FALSE conditions, as we found for the positive cases. This expectation, however, was not borne out: participants accepted these sentence to the same extent in their TRUE and TARGET conditions. Thus, there is no evidence for the existence of negative FC, DI or II in our data. While there is currently no experimental evidence supporting the existence of negative DI or II, Marty et al. 2021 have recently offered such evidence for high negative FC. The fact that we did not detect the presence of this inference in our study could indicate that some features of the task we devised made it more challenging for participants to engage in scalar reasoning, hence leaving more room for literal interpretations. In support of this line of explanation, we note for instance that the acceptance rates for positive FC and positive SI in our critical conditions (around 23% and 90%, respectively) were higher than those found in Marty et al. 2021's studies (around 7% and 75%, respectively), evidencing that participants derived these inferences to a lesser extent in our study.[7] It is important to emphasise that this potential task-related effect, if present, should be general and, consequently, does not affect in any way our conclusion regarding the discrepancies between positive and negative cases. However, these considerations may invite future research investigating the existence of not-yet-attested scalar inferences, like negative DI and negative II, to devise acceptability tasks that are tailored for these purposes.

The second point concerns an aspect of our results which we have not yet commented on. Specifically, we observe that the mean acceptance rate to the positive DI sentences in the True conditions was slightly lower than expected ($M = 69$, 95% CI[75,62]) in comparison to the other True positive conditions (all Ms> 82). As we did not design our experiment to probe for such differences, it is difficult to tell whether this result should be given any theoretical importance in the context of our study. Nonetheless, we point out that this result could indicate that participants sometimes computed embedded SIs for these sentences, e.g., by applying the exhaustivity operator to each disjunct; that is, participants may have parsed $\Box(\mathsf{A} \vee \mathsf{B})$ as $\Box(\textsc{exh}(\mathsf{A}) \vee \textsc{exh}(\mathsf{B}))$, generating the additional inference $\Box((\mathsf{A} \wedge \neg\mathsf{B}) \vee (\mathsf{B} \wedge \neg\mathsf{A}))$. Crucially, this inference was false in the True picture conditions for positive DI, as one of the three open boxes made $(\mathsf{A} \wedge \mathsf{B})$ possible, while it is was true in the corresponding Target picture conditions. While such embedded SIs may have also been derived for positive FC and DI sentences, it is worth emphasizing that these SIs were, by contrast, always true in the True and Target picture conditions associated with these sentences. Thus, the potential presence of embedded SIs could explain the circumscribed variations observed among the positive True conditions.[8]

In the following two sections, we report on two follow-up experiments which aimed to evaluate the generality of the main findings from this first study, first by moving the position of negation in the negative FC, DI and SI sentences below the modal (Experiment 2) and then by testing these same constructions with another pair of modals (Experiment 3).

7 It is possible, for instance, that the reasoning process required to infer the possible contents of the mystery box heavily drew on participants' working memory resources, affecting their pragmatic reasoning in a way similar to a dual memory task (De Neys & Schaeken 2007, Marty & Chemla 2013, van Tiel, Pankratz & Sun 2019, van Tiel et al. 2019, Marty et al. 2020). For the time being, we simply note that the average increase in mean acceptance rates observed for positive FC and positive SI in our study, compared to Marty et al. 2021's Experiment 3, is in line with the increase generally induced by higher memory load in dual task studies.

8 Similar discrepancies were also found in Exp.2 and Exp.3, but not in Exp.4, where the structure of the DI sentences was slightly different.

3.2 Experiment 2: Intermediate negation

The materials and method used in this experiment were the same as in Exp.1 with only one exception: the high negative sentences involving the modals *possible* and *certain* were modified so as to obtain their intermediate negative variants. Concretely, this was done by replacing the negated modals in these sentences with their (non-negated) scalemates (i.e.,*not possible*→*certain*, *not certain*→*possible*) and by moving the negation further down, at the level of the main verb of the embedded clause (i.e.,*contain*→*not contain*), as shown in Table 5. The positive cases and the negative II cases were the same as in Exp.1.

		Schematic description of the sentences
POSITIVE		
	FC	It is possible that the mystery box contains either a [A] ball or a [B] ball.
	DI	It is certain that the mystery box contains either a [A] ball or a [B] ball.
	II	The mystery box contains either a [A] ball or a [B] ball.
	SI	It is possible that the mystery box contains a [A] ball.
NEGATIVE		
	FC	It is possible that the mystery box does not contain both a [A] ball and a [B] ball.
	DI	It is certain that the mystery box does not contain both a [A] ball and a [B] ball.
	II	The mystery box does not contain both a [A] ball and a [B] ball.
	SI	It is possible that the mystery box does not contain a [A] ball.

Table 5 Schematic description of the sentences tested Experiment 2 by sentence polarity and inference type, where [A] and [B] correspond to different colour adjectives (e.g., [A]=green and [B]=blue).

The rest of the design was identical to that of Exp.1 in all regards (see 3.1.2 for details). Thus, Table 3 also serves as a summary of the sentence-picture combinations giving rise to the True, False and Target conditions in this novel experiment. As in Exp.1, each condition was iterated three times, leading to a total of 72 test trials. The procedure to pseudo-randomly choose the colors of the matching and non-matching balls was also the same as in Exp.1.

3.2.1 Participants

70 new participants (average age 33.4 yrs; 51 female) were recruited online through Prolific using the same pre-screening criteria as in Exp.1. Participants were paid £1.75 for their participation and average completion time was about 11 minutes (£9.49/hr). The consent and data collection procedures were the same as in Exp.1.

3.2.2 Procedure

The procedure was identical to the one used in Exp.1. We refer the reader to Section 3.1.3 for a general description, to Appendix A.1 for the instructions and to Appendix A.2 for an overview of the training phase.

3.2.3 Data treatment and analyses

Data treatment was the same as for Exp.1. Responses from 5 participants were excluded prior to analyses because their overall performance to the TRUE and FALSE trials did not reach the pre-established threshold of 70% accuracy. In total, 360 out of 5,040 responses were removed through this procedure (7% of the data; as in Exp.1). The data were analysed using the data analysis pipelines from Exp.1 in accordance with the two sets of analyses we had planned (see 3.1.4 for details and discussion).

3.2.4 Results

Figure 3 shows the mean acceptance rates to each sentence type by sentence polarity, inference type and experimental condition. As mentioned above, the two main analyses we carried out were the same as for Exp.1. In the first analysis, the model for the positive sentences yielded a main effect of Condition ($\chi^2(1) = 63.03$, $p < .001$), Inference type ($\chi^2(3) = 160.76$, $p < .001$) and a significant interaction between both factors ($\chi^2(3) = 59.81$, $p < .001$). The difference between TRUE and TARGET conditions was significantly lower for SI than for FC ($\beta = 1.87$, $p < .001$) and II ($\beta = 2.43$, $p < .001$); in contrast to what we found in Exp.1, however, the interaction involving SI vs. DI did not reach significance ($\beta = 0.62$, $p = .15$).[9] The model for negative sentences yielded the same general results: there was a main effect of Condition ($\chi^2(1) = 6.56$, $p < .05$), Inference type ($\chi^2(3) = 40.47$, $p < .001$) and a significant interaction between them ($\chi^2(3) = 13.49$, $p < .005$). As in Exp.1, the difference between TRUE and TARGET conditions in the negative cases was significantly larger for SI than for FC ($\beta = -0.64$, $p < .005$), DI ($\beta = -0.64$, $p < .05$) and II ($\beta = -0.72$, $p < .005$). The results delivered by the second analysis were similar to those reported in Exp.1: there was a main effect of Polarity for each inference type (Negative>Positive; all χ^2s> 25, ps$< .001$), a main effect of Condition for FC, DI and SI (TRUE>TARGET; χ^2s> 18, ps$< .001$; II: $\chi^2(1) = 2.11$, $p = .14$) and a significant interaction

9 A visual inspection of the results in Fig. 2 and Fig. 3 suggests that the discrepancy reported here is mainly driven by a decrease in acceptability for the positive DI sentences in the TRUE conditions in Exp.2. This subtle variation between both experiments is immaterial for our purposes. We note, however, that one explanation for this variation could be that participants derived embedded SIs for positive DI sentences and even more so in Exp.2 (see our Discussion in 3.1.6).

between both factors for FC, DI and II (χ^2s> 9, $ps< .005$; SI: $\chi^2(1) = 1.05$, $p = .30$).[10] The post-hoc analysis investigating the effects of Condition within inference type also delivered the same results as in Exp.1: in the positive cases, the estimated marginal means for the TRUE and TARGET conditions were significantly different for all inference types but SI whereas, in the negative cases, there was no such contrasts except for SI.

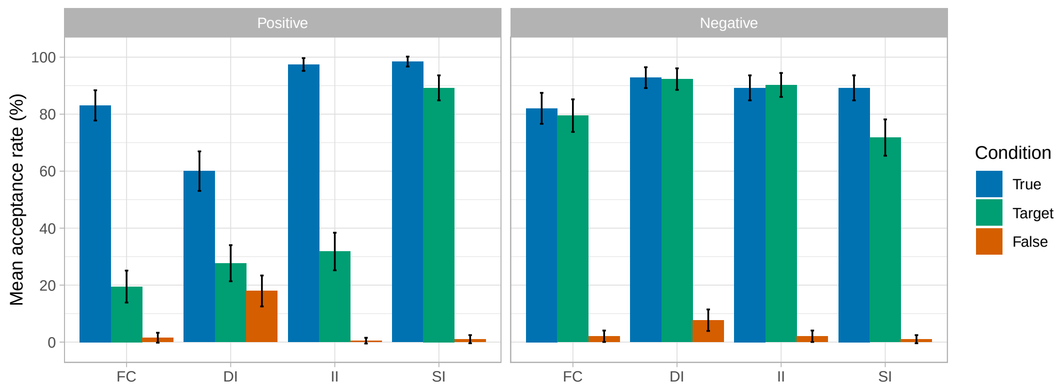


Figure 3 Mean proportion of 'Good' responses to each sentence type in Experiment 2 by sentence polarity (Positive, Negative), inference type (FC, DI, II, SI) and experimental condition (TRUE, TARGET, FALSE). Error bars represent 95% CIs.

3.2.5 Discussion

Results from Exp.1 showed that the contrasts between positive and negative FC reported in the previous literature extend to epistemic modals and, crucially, to two other inference types, namely DI and II. Following up on these findings, we carried out Exp.2 to investigate whether the contrasts unveiled in Exp.1 reproduce with variants of the negative cases where negation appears below, rather than above the modal. For these purposes, we designed Exp.2 as a minimal variant of Exp.1 by replacing the high negative FC, DI and SI sentences with the variants of interest, i.e., intermediate negative FC, intermediate negative DI and low negative SI. The patterns of responses we found for these novel cases, as well as their relationships to their corresponding positive forms, were similar to those we found in Exp.1 for the high negative cases. Specifically, the acceptance rates for intermediate negative FC and DI sentences in their TARGET conditions were found to be (i) substantially higher than those for their positive forms in the same conditions, (ii) slightly higher than that for the low negative SI sentences in the same conditions, and (iii) similar

10 The model for FC failed to converge, resulting in unreliable estimates and standard errors. To overcome this issue, we rerun this model with a simpler random effect structure obtained from the original one by removing the by-Subject random slope for Condition.

to those obtained for these same sentences in their True conditions. These results refute further predictions of the Extended Alternative-based Hypothesis (see Table 2) in establishing that the contrasts from Exp.1 reproduce in full with negative cases where negation appears instead in the complement clause of the modal. Pursuing our goal to establish the generality of this phenomenon, our next experiment aimed to test whether the novel contrasts from Exp.2 reproduce with other epistemic modals.

3.3 Experiment 3: Intermediate negation with other modals

Using the same method and pictures as in Exp.1-2, this experiment tested novel sentences involving the modals *might* and *must* in place of *possible* and *certain*, as shown in Table 6. Sentences were constructed after those investigated in Exp.2 so that, in the negative FC, DI and SI sentences, negation occurred below the modal (see Table 5 for comparisons).

		Schematic description of the sentences
POSITIVE		
	FC	There might be either a [A] ball or a [B] ball.
	DI	There must be either a [A] ball or a [B] ball.
	II	There is either a [A] ball or a [B] ball.
	SI	There might be a [A] ball.
NEGATIVE		
	FC	There might not be both a [A] ball and a [B] ball.
	DI	There must not be both a [A] ball and a [B] ball.
	II	There isn't both a [A] ball and a [B] ball.
	SI	There might not be a [A] ball.

Table 6 Schematic description of the sentences tested Experiment 3, where [A] and [B] correspond to different colour adjectives (e.g., [A]='green' and [B]='blue').

The rest of the design was identical to that of Exp.1–2 (see 3.1.2 for details). Thus, here again, Table 3 serves as a summary of the experimental conditions. As before, each condition was iterated three times, leading to a total of 72 test trials. The procedure to pseudo-randomly choose the colors of the various balls was also the same as in Exp.1–2.

3.3.1 Participants

70 new participants (average age 32.9 yrs; 54 female) were recruited online through Prolific using the same pre-screening criteria as in Exp.1–2. Participants were paid £1.75 for their

participation and average completion time was about 11 minutes (£10.15/hr). The consent and data collection procedures were the same as in Exp.1–2.

3.3.2 Procedure

The procedure was identical to the one used in Exp.1–2 (see 3.1.3 for details and Appendix A.1 for the instructions). The sentences presented to participants during the training phase were adjusted to our present purposes (see Appendix A.2 for details).

3.3.3 Data treatment and analyses

Data treatment was the same as for Exp.1–2. Responses from 4 participants were excluded prior to analyses because their overall performance to the TRUE and FALSE trials did not reach the pre-established threshold of 70% accuracy. In total, 288 out of 5,040 responses were removed through this procedure (about 6% of the data; as in Exp.1–2). The data were analysed using the data analysis pipelines from Exp.1-2 in accordance with the two sets of analyses we already described (see 3.1.4).

3.3.4 Results

Figure 4 shows the mean acceptance rates to each sentence type by sentence polarity, inference type and experimental condition. The results delivered by our two analyses were largely similar to those reported in Exp.1–2. In the first analysis, the model for positive sentences yielded a main effect of Condition ($\chi^2(1) = 108.33$, $p < .001$), Inference type ($\chi^2(3) = 214.60$, $p < .001$) and a significant interaction between both factors ($\chi^2(3) = 46.51$, $p < .001$). The difference between TRUE and TARGET conditions was significantly lower for SI than for FC ($\beta = 0.88$, $p < .05$) and II ($\beta = 2.57$, $p < .001$); the interaction involving SI vs. DI did not reach significance ($\beta = 0.57$, $p = .1$), as in Exp.2 (see footnote 9 for discussion). Next, for the negative sentences, there was a main effect of Condition ($\chi^2(1) = 15.80$, $p < .001$), Inference type ($\chi^2(3) = 32.12$, $p < .001$) and a significant interaction between them ($\chi^2(3) = 59.62$, $p < .001$). As in Exp.1–2, the difference between TRUE and TARGET conditions was significantly larger for SI than for FC ($\beta = -1.30$, $p < .001$), DI ($\beta = -0.97$, $p < .001$) and II ($\beta = -1.36$, $p < .001$). Turning to the second analysis, there was main effect of Polarity for each inference type (Negative>Positive; all χ^2s> 26, ps$< .001$), a main effect of Condition for FC, DI and SI (TRUE>TARGET; all χ^2s> 13, ps$< .001$; II: $\chi^1 = 1.77$, $p = .18$), and a significant interaction between both factors for FC, SI and II (all χ^2s> 31, ps$< .001$; SI: $\chi^1 = 1.18$, $p = .27$).[11] Finally, the post-hoc analysis revealed that, in the positive cases, the estimated marginal means for the TRUE

11 The model for FC failed to converge, as in the data analyses for Exp.2. We rerun this model with a simpler random effect structure following the procedure described in fn 10.

vs. Target conditions were significantly different for all inference types whereas, in the negative cases, there was no such contrasts except for SI.

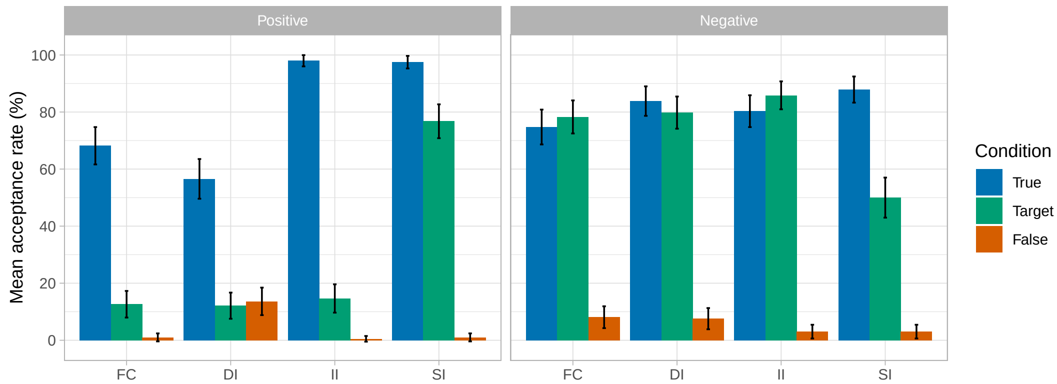


Figure 4 Mean proportion of 'Good' responses to each sentence type in Experiment 3 by sentence polarity (Positive, Negative), inference type (FC, DI, II, SI) and experimental condition (True, Target, False). Error bars represent 95% CIs.

3.3.5 Discussion

Results from Exp.2 showed that the contrasts found in Exp.1 reproduce with variants of the high negative cases where negation appears right below the modal. The present results generalise these findings by establishing that the contrasts observed in the latter cases reproduce with related expletive constructions involving *might* and *must*. These findings confirm and consolidate those from Exp.2.

In addition, we observe that, for the four positive cases and the negative SI cases, the novel sentences we tested generally gave rise, in the critical conditions, to lower acceptance rates than in Exp.2. While our data do not permit us to identify precisely the driving force behind these variations, they suggest that several factors might be at play. For instance, for positive FC and DI, the variations observed between Exp.2 and Exp.3 do not seem specific to the Target conditions; rather, a similar trend is observed in the True conditions, suggesting that the corresponding sentences were overall more rejected in Exp.3 than in Exp.2. This line of explanation, however, does not extend to the other cases, i.e., positive II, positive SI and negative SI, where those differences were only found in the Target conditions. The variations observed in these cases support the idea that the sentences tested in Exp.3 gave rise to more IIs and SIs than those tested in Exp.2. One reason for that could be that the use of shorter and simpler sentences in Exp.3 alleviated the load on participant's cognitive resources during the experiment, allowing them to engage more in scalar inferencing (see the end of Section 3.1.6 for discussion). Whether or not this explanation is on the right track, the fact that no such decrease in acceptability was found for negative FC, DI and II offers further evidence that these putative inferences, if they exist, are far less robust than their positive instances.

3.4 Experiment 4: Low negation

This fourth experiment investigated the remaining negative cases not yet tested, the low negative ones, where negation is embedded as low as possible (see Table 1). The low negative cases for FC, DI and II were built after their intermediate negative variants from Exp.2 by placing a negation in each sub-clause of the complex embedded clause and by replacing the embedded connectives with their scalemate (i.e., *and*→*or*). The linguistics contents of these sentences were further adjusted to ensure that the resulting sentences sound as natural as possible to native speakers, e.g., by making all the syntactic subjects phonologically explicit. For the sake of parallelism, similar adjustments were applied to their positive counterparts. The low negative SI sentences were the same as in Exp.2. The resulting sentence types are shown in Table 7.

		Schematic description of the sentences
POSITIVE		
	FC	It is possible that either the mystery box contains a [A] ball or it contains a [B] ball.
	DI	It is certain that either the mystery box contains a [A] ball or it contains a [B] ball.
	II	Either the mystery box contains a [A] ball or it contains a [B] ball.
	SI	It is possible that the mystery box contains a [A] ball.
NEGATIVE		
	FC	It is possible that either the mystery box does not contain a [A] ball or it does not contain a [B] ball.
	DI	It is certain that either the mystery box does not contain a [A] ball or it does not contain a [B] ball.
	II	Either the mystery box does not contain a [A] ball or it does not contain a [B] ball.
	SI	It is possible that the mystery box does not contain a [A] ball.

Table 7 Schematic description of the sentences tested Experiment 4, where [A] and [B] correspond to different colour adjectives (e.g., [A]='green' and [B]='blue').

Pictures and experimental conditions were the same as in Exp.1–3 (see Table 3 for an overview of the sentence-picture combinations). As before, each condition was iterated three times, for a total of 72 test trials. The procedure to pseudo-randomly choose the colors of the various balls was also the same as in Exp.1–3.

3.4.1 Participants

70 new participants (average age 31.2 yrs; 47 female) were recruited online through Prolific using the same pre-screening criteria as in Exp.1–3. Participants were paid £1.75 for their

participation and average completion time was about 12 minutes (£8.93/hr). The consent and data collection procedures were the same as in Exp.1–3.

3.4.2 Procedure

The procedure and instructions were identical to the ones for Exp.1–3. We refer the reader to Appendix A.2 for an overview of the training phase.

3.4.3 Data treatment and analyses

Data treatment was the same as for Exp.1–3. Responses from 5 participants were excluded prior to analyses because their overall performance to the True and False trials did not reach the pre-established threshold of 70% accuracy. In total, 360 out of 5,040 responses were removed through this procedure (7% of the data; as in Exp.1–3). The data were analysed using the data analysis pipelines from Exp.1–3 (see 3.1.4 for description).

3.4.4 Results

Figure 5 shows the mean acceptance rates to each sentence type by sentence polarity, inference type and experimental condition. As in Exp.1–3, in the first analysis, the model for positive sentences and the one for negative sentences both yielded a main effect of Condition (χ^2s$>$ 74, ps$<$ .001), Inference type (χ^2s$>$ 17, ps$<$ .001) and a significant interaction between both factors (χ^2s$>$ 50, ps$<$ .001). In contrast to what we found in Exp.1–3, however, the direction of the contrasts between SI and the other inferences types was the same across the board: for both the positive and the negative sentences, the difference between True and Target conditions was significantly smaller for SI than for FC (positive: $\beta = 4.33$, $p < .001$; negative: $\beta = 0.62$, $p < .01$), DI (positive: $\beta = 3.66$, $p < .001$; negative: $\beta = 1.83$, $p < .001$) and II (positive: $\beta = 5.18$, $p < .001$; negative: $\beta = 2.05$, $p < .001$).

The second set of analyses also yielded slightly different results compared to Exp.1–3. The effect of Polarity was significant for DI, II and SI (all βs$>$ 20, ps$<$ .001), but not for FC ($\beta = .001$, $p = .96$), indicating that responses to FC sentences were overall the same in their positive and low negative versions. The effect of Condition was significant for FC, SI as well as II (all βs$>$ 10, ps$<$ .005), but not for SI ($\beta = 1.13$, $p = .28$), indicating that response to positive and negative SI sentences were overall the same in their True and Target conditions. As before, however, the interaction between Polarity and Condition was significant for FC, DI and II (all βs$>$ 4, ps$<$ .05), but not for SI ($\beta = 0.05$, $p = .81$). The post-hoc analysis largely confirmed the general results from both analyses in showing that the contrasts between True and Target conditions were significant across the board for all inference types, except for SI.

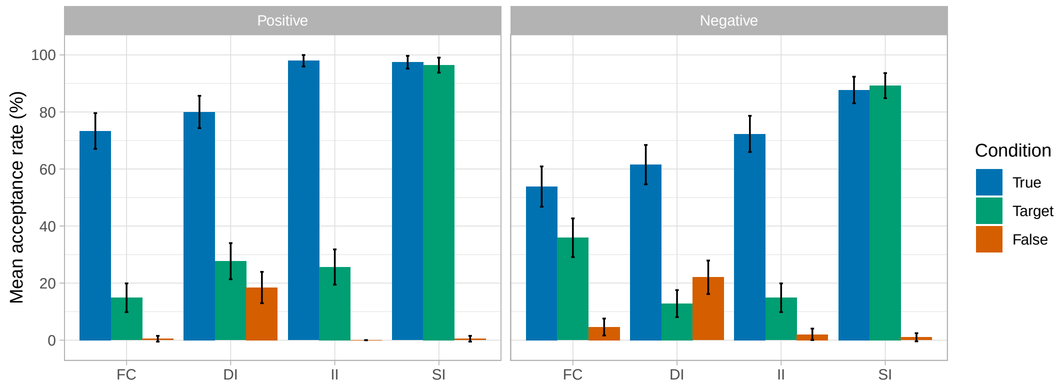


Figure 5 Mean proportion of 'Good' responses to each sentence type in Experiment 3 by sentence polarity (Positive, Negative), inference type (FC, DI, II, SI) and experimental condition (True, Target, False). Error bars represent 95% CIs.

3.4.5 Discussion

Results from Exp.1–3 showed strong and systematic differences in the availability of FC, DI and II between the positive and negative cases tested in these experiments. The results of the present experiment stand in stark contrast with these previous results: the patterns of responses for the low negative sentences were all found to display the same general behavior as those for their positive versions across all inference types. Specifically, just like their positive versions, low negative FC, DI and II sentences were all found to be more rejected in their Target conditions ($12\% < Ms < 35\%$) than in their True conditions ($53\% < Ms < 73\%$). These results evidence that the low negative cases for FC, DI and II gave rise to the eponymous inferences. Crucially, we note that these inferences were detected using the same task and method as in Exp.1–3 and in the context of an experiment where, by contrast, there appears to be no sign of regular SIs, whether direct or indirect.[12]

Let us turn next to more intricate questions: how robust are the FC, DI and II inferences arising from the low negative cases? And how do they fare compared to those arising from the positive ones? Taken at face value, the relatively low acceptance rates observed in the Target conditions suggest that these inferences are quite strong. In particular, we observe that participants did not fully accept the low negative FC, DI and II sentences even in situations that made them clearly true, i.e., in the True conditions; furthermore, the acceptance rates for these sentences in their True conditions were substantially lower than those for their positive versions in the corresponding conditions. We take these results to

12 The fact that we didn't find any sign of regular SIs in this experiment is surprising as the SI sentences were the same as in Exp.2. The source of these variations remains unclear to us at the moment. We note however that these variations invite us to interpret with caution the contrasts in inference strength between SI and the other inference types that we observed in Exp.4.

indicate that participants' judgements for these sentences were affected by considerations beyond the evaluation of their truth and felicity. One such consideration could relate for instance to the naturalness of these sentences. Compared to the negative sentences tested in Exp.1–3, the low negative sentences were lengthier (no phonological reduction) as well as structurally more complex (two negations instead of one), making them less natural than their high and intermediate variants. Thus, it is conceivable that some participants found these sentences quite cumbersome and used the "Bad" response option to report not only a judgment of falsity, but also a perceived defect of naturalness. Setting this possible explanation aside, the fact that, at a general level, participant's acceptance of the low negative FC, DI and II sentences were negatively impacted by independent factors has some consequences for the interpretation of our results. Essentially, it tells us that, for these sentences, the acceptance rates observed in the Target conditions should not be taken at face value and should not be used in isolation to evaluate the strength of the relevant inferences. We note however that, regardless of the factor at play here, its effect should be general and affect participant's responses to these sentences in a similar way across experimental conditions. Consequently, the results of the comparisons that we carried out, across sentence polarity and across inference type, can still be used to approximate the strength of the inferences at hand. With these considerations in mind, we conclude that our results suggest that low negative FC, DI and II are not as robust as their positive counterparts while they are clearly more robust than their high and intermediate variants and certainly more robust than high and low negative SI (see fn. 12 for refinements).

4 General Discussion

On the implicature approach, the derivation of FC, II and DI, in both their positive and their (high, intermediate, or low) negative instances are all based on alternatives that do not involve lexical access substitution. Our results are challenging for this uniform approach if it is supplemented with the alternative-based in (37), repeated below.

(37) **Alternative-based hypothesis (extended version)**
Alternatives that do not involve lexical substitutions give rise to inferences that are more robust, faster to process and easier to acquire.

According to (37), inferences involving non-lexical alternatives should be, among other things, more robust. The challenge for this hypothesis comes from the finding that the availability of FC, DI and II substantially differs across the various positive and negative environments we tested. Specifically, we found no sign of these inferences in the high and intermediate negatives cases, where negation scopes over conjunction, while we found that these inferences were quite robust for the logically equivalent low negative cases, where negation appears in the scope of disjunction. The challenge is strengthened by the fact that no such a difference was found between positive and negative regular SIs; in fact, if anything, this last comparison shows a trend in the opposite direction. Our findings are summarised in Table 8, alongside the predictions of the implicature approach supplemented with the hypothesis in (37). As it is clear from the table, the predictions of this approach are not empirically born out for the high and intermediate negative cases. It is worth emphasising that the problem here is not *per se* that we found no evidence for the relevant inferences in these cases. As we discussed, our data are compatible with these inferences being extremely weak and hard to detect. Rather, the problem comes from the differences observed between these cases and their positive and low negative counterparts, which were all computed at quite high rates across our experiments.

	Predictions				**Findings**			
	FC	DI	II	SI	FC	DI	II	SI
POSITIVE	✓	✓	✓	✓	✓	✓	✓	✓
NEGATIVE								
High	✓	✓	✓	✓	✗	✗	✗	✓
Intermediate	✓	✓			✗	✗		
Low	✓	✓	✓	✓	✓	✓	✓	✓

Table 8 Summary of the predictions of the implicature approach supplemented with the hypothesis about alternatives in (37) (on the left), together with the actual results we found in our experiments (on the right). We use ✗ to indicate that we found no evidence of the inference.

Before going on, let us briefly consider another alternative-related hypothesis recently suggested by Bar-Lev (2018) and Bar-Lev & Fox (2020) (see also Schulz 2019 and Romoli, Santorio & Wittenberg 2020). The hypothesis in question is given in (49).

(49) **The EXH-NEG scalemate hypothesis**
EXH and negation are lexical alternatives; as a consequence, substituting one for the other gives rise to additional alternatives.

The idea behind the formulation of (49) is that a sentence involving negation (or EXH) has more lexical alternatives than usually assumed. Specifically, such a sentence is hypothesized to have additional lexical alternatives derived by replacing negation with EXH (and vice-versa). It is crucial to observe that, since these additional alternatives enter SI computation, their presence may affect the calculation of the sets of 'innocently excludable' and 'innocently includable' alternatives and, in turn, block certain scalar inferences that would be expected otherwise to go through. The question for us is thus whether this hypothesis could account for the various contrasts we found in our experiments. As we discuss in detail in Appendix B, this hypothesis can account for the difference between positive and high negative cases for FC and DI. Furthermore, it correctly predicts that these inference types should be robust in the low negative cases. However, as it stands, it makes incorrect predictions for the intermediate cases, as it predicts similarly robust inferences in these cases as in the low negative cases. In addition, it is not clear it can help with the corresponding difference we found for II, and it incorrectly extends to regular SIs, predicting that negative SIs should be generally weaker than positive SIs. In sum, while this hypothesis can explain some of our results, it is far from providing a full account of them. What we need is an account that can distinguish (i) between the positive and negative cases of FC, DI and II, without extending to regular SIs, and (ii) between the high and intermediate negative cases on one side, and the low negative cases on the other. In the next section, we outline two promising directions to account for our results: one based on an additional hypothesis about relevance, and the other based on a non-implicature approach to (the positive version of) FC, DI and II.

5 Two promising directions

5.1 A relevance-based account

We start by sketching another route for the implicature approach. The proposal is meant to account for the difference between the cases involving disjunction, whether positive or negative, and the other negative cases involving conjunction. It is based on two main ingredients – a notion of relevance and a stipulation about how disjunction makes its disjuncts relevant – which we will illustrate in turn. We then point to part of the approach, which would need to be developed.

As commonly assumed, we take implicature computation to be constrained by relevance. Specifically, we assume that the implicature associated with a given alternative is derived only if that alternative is relevant. One common way to characterise the set of relevant alternatives to a given utterance is to define them as those alternatives that appropriately 'address', in some sense to be defined, the understood Question Under Discussion (QUD). For concreteness, we can formalise the notion of relevant proposition as in (50), according to which a proposition is relevant if it is equivalent to a cell or a union of cells in such partition (Groenendijk & Stokhof 1984, Marty 2017, Romoli 2012, Marty & Romoli 2021: among others).

(50) **Relevance**: A proposition p is relevant in a context c given a partition Q of c iff for any cell $q \in Q$ and any $w, w' \in q$, $p(w) = p(w')$.

The second ingredient comes from the observation that disjunctive statements are generally perceived as felicituous only if their disjuncts can be construed as relevant alternatives to one another (for discussion, see Marty & Romoli 2021, Simons 2001, Fox 2007, Singh 2008, Fox & Katzir 2011). This felicity condition would be evidenced by examples like (51), from Simons 2001, where the two disjuncts are not discourse-related in any obvious way and the resulting disjunction sounds quite odd.

(51) ??There is dirt in the fuel line or it is raining in Tel-Aviv.

In light of the definition we gave in (50), one way to understand this felicity condition on disjunction is to assume that a disjunctive statement is likely to be understood as an answer to a QUD to which its disjuncts are also possible answers. To put it differently, a disjunctive statement is preferentially understood as an answer to a question which makes both its disjuncts relevant because its disjuncts are preferentially construed as relevant alternatives to one another. This interpretive preference can be formulated as follows:

(52) **Relevance for Disjunctions**
A disjunction is likely understood as an answer to an active QUD which is also addressed by its single disjuncts. That is, whenever a disjunction is relevant, its disjuncts are likely relevant as well.

Note that it is sufficient for our purposes that the principle in (52) be probabilistic in nature, rather than absolute. In fact, as Marty & Romoli (2021) and Simons (2001) discuss, there are situations where the preference stated in (52) can be counteracted. It is so for instance if a disjunctive statement is understood as an answer to a 'Yes-No', polar question. For instance, in a conversation like (53), where a doctor is asking the patient about the symptoms of a particular disease, the partition associated with A's question has only two cells: one in which B haven't had either of persistent dry cough or high temperature, and one in which she has had at least one of them, possibly both. Given this partition, one can verify that the whole disjunction is relevant, but neither of the independent disjuncts is.[13]

(53) A: Have you had a persistent dry cough or high temperature?
B: I have (had a persistent dry cough or high temperature).

With these two ingredients in place, we can now go back to our cases. The general result, on this approach, is that positive and low negative cases involving disjunction are expected to give rise to more robust inferences than their high and intermediate negative variants involving conjunction, consistent with our data.

5.1.1 Positive and low negative cases

The positive and low negative cases involve disjunction. In these cases, the principle in (52) makes it so that the alternatives corresponding to the disjuncts – which are central to the derivation of FC, DI and II – are likely relevant and, as a result, the corresponding

13 Other cases where this preference has been found to be counteracted are examples like (i), which Simons (2001) calls 'monkey's uncle' disjunctions.

(i) Either Jane won, or I'm a monkey's uncle.

inferences, be it FC, DI, or II, should be readily derived. To illustrate, consider for instance the case of ignorance inferences. According to (52), a simple disjunctive sentence like (54) is likely to be understood as an answer to a QUD which makes its disjuncts relevant, e.g., *What does the box contain?*. Therefore, the ignorance inferences associated with an utterance of (54) will tend to be computed.

(54) The box contains a blue ball or a yellow ball.
  a. The box contains a blue ball.
  b. The box contains a yellow ball.

Similarly, the low negative variant of (54) in (55) will be understood as an answer to a QUD which makes its disjuncts relevant (e.g., *What does the box does not contain?*).

(55) The box doesn't contain a blue ball or it doesn't contain a yellow ball.
  a. The box doesn't contain a blue ball.
  b. The box doesn't contain a yellow ball.

Finally, since the principle in (52) applies to disjunctions in general, it naturally extends to the corresponding cases giving rise to FC and DI. We should note, however, that for this embedded case one would have to say more about how the constraint on disjunction and its disjuncts is extended to include the embedding environments (e.g. the embedding modals in our cases of FC and DI).[14]

5.1.2 High and intermediate negative cases

The principle in (52) is silent about cases that do not involve disjunction. In particular, it leaves open the possibility that sentences like (56) be preferentially understood as answers to QUDs which do not make their independent conjuncts relevant (e.g., *Did the box contain a blue and a yellow ball?*). Hence, the relevance-based account we sketched does not make predictions regarding the strength of the inferences associated with such sentences. Similar observations hold for the FC and II cases involving embedded conjunction.

(56) The box doesn't contain a blue ball and a yellow ball.
  a. The box doesn't contain a blue ball.
  b. The box doesn't contain a yellow ball.

In sum, the principle in (52) only states that the disjuncts of a disjunctive sentence are preferentially understood as relevant alternatives, predicting the implicatures derived from these alternatives to be quite robust. It is thus compatible with the inferences arising from conjunctive sentences being comparatively much weaker.

14 For attempts along these lines and discussion see Marty 2017, Marty & Romoli 2021.

5.1.3 Regular SIs

The principle in (52) is also silent about regular SIs as long as disjunction is not involved. Hence, it is compatible with regular SIs being less robust than FC, DI and II, as suggested by our results. It bears pointing out, however, that the present account predicts that the regular 'exclusivity' SI associated with disjunctive sentences should in principle be more robust than the regular 'inclusivity' SI associated with their negated, conjunctive variants. Concretely, it predicts that the exclusive SI associated with (57a), derived from the conjunctive alternative in (57b), should be more robust than the inclusive SI associated with (58a) and derived from the disjunctive alternative in (58b).

(57) Exclusive SI

a. The box contains either a blue ball or a yellow ball.
b. The box contains a blue ball and a yellow ball.
c. ⇝ *The box doesn't contain both a blue ball and a yellow ball*

(58) Inclusive SI

a. The box doesn't contain both a blue ball and a yellow ball
b. The box contains either a blue ball or a yellow ball.
c. ⇝ *The box contains either a blue ball or a yellow ball*

This prediction obtains because, according to the principle in (52), if (57a) is relevant, then its disjuncts - namely *the box contains a blue ball* and *the box contains a yellow ball* – are likely relevant as well. Since the general assumption in the literature is that relevance is closed under conjunction (Katzir 2007, Fox & Katzir 2011), it is natural to think that the conjunctive alternative in (57a) should also be perceived as relevant. While we did not test the exclusive and inclusive SIs of disjunction, we suspect that the expected contrast doesn't hold, but we leave a more detailed investigation of this question for future work.[15]

5.1.4 Summary of the relevance-based account

The relevance-based account relies on an assumption about the calculation of relevance for disjunction which, as we explained, captures previous observations regarding certain felicity conditions that disjunctive statements are subject to. The predictions of this accounts are summarised in Table 9. Crucially, this account predicts FC, DI and II to be robust whenever disjunction is involved, regardless of whether a modal or negation is present as well, consistent with our results. All in all, we think that this direction is more promising than the one based on the distinction between lexical and sub-constituent alternatives, or the one based on the idea that negation and EXH are lexical alternatives.

15 The results from Marty et al. 2020 suggest that, if anything, the contrast goes the other way round: inclusive SIs like (58c) are more robust than exclusive ones like (57c).

	Predictions				Findings			
	FC	DI	II	SI	FC	DI	II	SI
POSITIVE	✓	✓	✓	✓	✓	✓	✓	✓
NEGATIVE								
High	✓	✓	✓	✓	✗	✗	✗	✓
Intermediate	✓	✓			✗	✗		
Low	✓	✓	✓	✓	✓	✓	✓	✓

Table 9 Summary of the predictions of the implicature approach supplemented with the relevance assumption for disjunction in (52) (on the left), together with the actual results we found in our experiments (on the right). We use ✗ to indicate that we found no evidence of the inference.

In the following, we outline another direction to account for our data, a hybrid approach combining a non-implicature account of the positive and low negative cases, with an implicature approach for the other ones.

5.2 A non-implicature approach

The implicature approach to FC, DI and II is not the only option in the literature. In particular, a number of authors have put forward a way to derive FC as a plain semantic entailment (Aloni 2003, 2007, 2018, Goldstein 2019, Rothschild & Yablo 2018, Simons 2005, Willer 2017, Zimmerman 2000). An appealing feature of this approach is that, since semantic entailments are generally robust, it immediately explains the robustness of FC. Furthermore, many of these theories are equipped with an optional mechanism to cancel FC, which is useful for capturing the observation that FC exhibits more nuanced robustness than what would be expected for an obligatory semantic entailment. Another notable feature of this approach is that many of its implementations are couched in a richer semantic system than classical semantics. The reason is that, in order to derive FC as an entailment, the possibility modal has to be able to access the meaning of each disjunct in a compositional fashion. Typically, Inquisitive Semantics (or more broadly, Alternative Semantics) is employed for this purpose. One feature of this framework that is particularly relevant for our discussion is that negation has the effect of (re)creating classical meaning. As we shall see, this feature proves useful for understanding the contrasts unveiled by our experimental results between, on one side, positive and low negative cases and, on the other, high and intermediate negative cases.

In addition to free choice, a recent work by Cremers et al. (2017) has put forward a theory of ignorance inferences that uses the same type of non-classical semantic system and that is therefore compatible with the entailment approach to FC. Although this theory does not treat ignorance inferences as plain entailments, attributing them instead to a pragmatic principle, it shares certain features with the entailment approach to free choice that are useful in accounting for our experimental results, as we will discuss in this section. Finally, as we will demonstrate below, we can also develop an entailment approach to distributive inferences using the same semantic system (cf. Simons 2005).

Building on the previous works cited above, we present below a concrete implementation of the non-implicature approach to FC, DI and II, and discuss its predictions for the cases involving negation and conjunction that were tested in our experiments. After pointing out some shortcomings of this theory, we will discuss the possibility of augmenting it with scalar implicatures, following the suggestion in Marty et al. 2021.

5.2.1 Alternative possibilities

The entailment approach to FC is often couched in Inquisitive Semantics (Ciardelli, Linmin & Champollion 2018: among others) or related frameworks. This is because these richer semantic systems make it possible to define a notion of alternatives that erive FC as an entailment. One should be keep in mind here that these alternatives are theoretically distinct from, but can co-exist with the alternatives that are used in scalar implicature computation. In order to avoid confusion, let us call the former *alternative possibilities*, and the latter *formal alternatives*. In our implementation, alternative possibilities are sets of possible worlds, while formal alternatives are linguistic expressions.

A key feature of these systems is that disjunction typically creates a set of alternative possibilities corresponding to their disjuncts. In order to formalise this idea, sentence meaning is consistently lifted to a set of propositions. That is, an atomic sentence is taken to denote a singleton set, as shown in (59).[16] For clarity, we will use boldface when we speak of a proposition qua a set of possible worlds.

(59) $[\![\text{p}]\!] = \{\mathbf{p}\}$

A disjunctive sentence simply denotes the union of the denotations of the disjuncts:

(60) $[\![\phi \text{ or } \psi]\!] = [\![\phi]\!] \cup [\![\psi]\!]$

5.2.2 Free choice

As a number of authors have pointed out (Aloni 2003, 2007, Simons 2005, Willer 2017, Goldstein 2019), a framework like this allows one to bake in FC in the meaning of the possibility modal as in (61). For better readability, we write $\Diamond\mathbf{p}$ for the proposition (qua a set

16 This is different from the basic version of Inquisitive Semantics, known as InqB, where sentence denotations are always downward closed sets. This is because, as far as FC and other inferences are concerned, only the maximal elements (known as 'alternatives') in such sets matter.

of possible worlds) that the proposition **p** is possible, that is, $\Diamond\mathbf{p} := \{w \mid \exists w'[wRw' \wedge w' \in \mathbf{p}\}$, where R is the relevant accessibility relation.

(61) $[\![\text{possible}(\phi)]\!] = \{\bigcap_{\mathbf{p}\in[\![\phi]\!]} \Diamond\mathbf{p}\}$

By way of illustration, let us take a simple case with two atomic sentences, a and b, as disjuncts. The resulting set contains a single proposition that **a** and **b** are both possible:

(62) $[\![\text{possible(a or b)}]\!] = \{\bigcap_{\mathbf{p}\in[\![\text{a or b}]\!]} \Diamond\mathbf{p}\} = \{\Diamond\mathbf{a} \cap \Diamond\mathbf{b}\}$

In contrast to the implicature approach to FC, an entailment approach needs to say something about how Double Prohibition comes about. That is, it needs to account for the fact that the meaning of (63) is stronger than the negation of its FC reading.

(63) It is not possible that the mystery box contains a blue ball or a yellow ball.

In this system, negation is usually defined as shown in (64). As one can verify, the semantics of negation does not help derive Double Prohibition. While one could try a different denotation for negation, there are reasons to maintain the definition in (64) and derive Double Prohibition in some other way. As we will later see, the denotation in (64) is in fact very useful for explaining some of our experimental results.

(64) $[\![\text{not } \phi]\!] = \{\overline{\bigcup[\![\phi]\!]}\}$

In order to account for Double Prohibition, one can follow Goldstein 2019 for instance and assume that the possibility modal has a homogeneity presupposition that either all alternatives of its prejacent are possible or none of them is. On this assumption, a negated possibility modal embedding disjunction is now defined and true when neither of the disjunct is possible (see Goldstein 2019 for details), hence deriving Double Prohibition.

5.2.3 Distributive inferences

As far as we are aware, there is no fully developed entailment approach to distributive inferences in the currently literature (cf. Simons 2005). However, the account above can easily be extended to derive these inferences as plain entailments as well. Specifically, we can bake in DI in the meaning of the universal modal, as shown in (65). For readability, we write $\Box\mathbf{p}$ for the proposition that **p** is necessary, that is, $\Box\mathbf{p} := \{w \mid \forall w'[wRw' \rightarrow w' \in \mathbf{p}]\}$. We assume that grand intersection with an empty domain of propositions returns the set of all possible worlds, W.

(65) $[\![\text{certain}(\phi)]\!] = \{(\Box\bigcup[\![\phi]\!]) \cap (\bigcap_{\mathbf{p}\in[\![\phi]\!]\setminus\{\bigcup[\![\phi]\!]\}} \overline{\Box\mathbf{p}})\}$

Let us unpack this denotation. The first bit, $\Box\bigcup[\![\phi]\!]$, is essentially the classical meaning of the necessity modal which requires that, in every accessible possible world, ϕ be true one way or another. The second bit, $\bigcap_{\mathbf{p}\in[\![\phi]\!]\setminus\{\bigcup[\![\phi]\!]\}} \overline{\Box\mathbf{p}}$, is the distributive inference, which amounts to the conjunction of statements of the form '**p** is not certain', for each proper alternative possibility **p**. Note that we only quantify over proper alternative possibility here since, when ϕ is a singleton for example, there shouldn't be any distributivity inference. Note also that, as in the case of FC, this meaning will derive something too weak for the negation of a necessity statement. For instance, a sentence like (66) would be true if it was certain that the box contained, say, a blue ball.

(66) It is not certain that the mystery box contains either a blue ball or a yellow ball.

One way of solving this issue is to elaborate on Goldstein (2019) and postulate a presupposition in the meaning of the necessity modal that strengthens the negative sentence appropriately. In this case, the required presupposition would be that either the necessity statement is true – i.e., the proposition $(\Box\bigcup[\![\phi]\!]) \cap (\bigcap_{\mathbf{p}\in[\![\phi]\!]\setminus\{\bigcup[\![\phi]\!]\}} \neg\Box\mathbf{p})$ holds – or the proposition $\overline{\Box\bigcup[\![\phi]\!]}$ is true. The latter disjunct strengthens the meaning of a negated necessity sentence like (66), deriving its intuitive meaning.

5.2.4 Ignorance inferences

For ignorance inferences, we follow a proposal by Coppock & Brochhagen 2013 (see also Cremers et al. 2017). This theory assumes the same semantic system as the one we are developing here and derives ignorance inferences using an independent pragmatic principle. Specifically, Coppock & Brochhagen 2013 make use of the following maxim of *inquisitive sincerity*.

(67) **Inquisitive sincerity**
Don't utter an inquisitive sentence if you already know how to resolve the issue that it expresses.

The maxim in (67) is concerned with the issue (or inquisitive content) expressed by a given sentence, that is, with the issue that is resolved by the alternative possibilities that this sentence introduces. A sentence ϕ is called inquisitive just in case the issue it expresses is not trivially resolved by the information conveyed by ϕ itself, i.e., just in case $[\![\phi]\!]$ contains more than one maximal set of possible worlds. Typically, a disjunctive sentence like *The box contains a blue ball or a yellow ball* is inquisitive in this technical sense because (i) it generates multiple alternatives, call them **blue** and **yellow**, and (ii) the information it conveys is not sufficient to determine which of these propositions contains the actual world. Thus, if a speaker utters such a sentence, assuming that the sincerity maxim is in force between the interlocutors, we can infer that this speaker does not know which of

blue or **yellow** is true. These inferences, together with the semantics in (60), derive the ignorance inferences associated with disjunctive sentences.

One noteworthy consequence of this theory is that it correctly predicts that negated disjunction should not rise to ignorance inferences. The reason is that, as per the definition in (64), negation gives rise to singleton sets of alternative possibilities and, as a result, a negative statement is always non-inquisitive, trivialising the sincerity maxim in (67). Finally, let us emphasise that, on this theory, ignorance inferences are neither semantic entailments, nor scalar implicatures. Therefore, one can make auxiliary assumptions about the robustness of such inferences so as to reflect our experimental results.

5.2.5 High negation

So far, we have seen how the theory derives FC and DI as plain entailments and II as pragmatic inferences. We have also seen that when negation is applied to such sentences, the results are essentially as in classical semantics. Specifically, for FC and DI, this is achieved by the presuppositions of modals; for II, this is due to the fact that negation is generally assumed to give rise to a single alternative possibility, no matter how many alternative possibilities its argument has. This property of negation is actually important for us. Let us now go through what the theory predicts for sentences containing negation over conjunction. Let us first consider the high negative cases for FC. Take (68), for example.

(68) It is not certain that the mystery box contains both a blue ball and a yellow ball.

In order to identify the predicted meaning for this sentence, we first have to define the meaning of conjunction. The standard way is to use point-wise intersection as follows:

(69) $[\![\phi \text{ and } \psi]\!] = \{\mathbf{p} \cap \mathbf{q} \,|\, \mathbf{p} \in [\![\phi]\!] \wedge \mathbf{q} \in [\![\psi]\!]\}$

Note in passing that, according to this definition, conjunction does not create new alternative possibilities; however, they can inherit alternative possibilities in case the conjuncts already contain alternatives. This is a good feature of the account, given that a conjunction of disjunctive statements (e.g. *It is possible that the mystery box contains a blue or yellow ball and a green or red ball*) generally give rise to FC, DI and II.

Now, consider a sentence of the form *certain(ϕ and ψ)*. As discussed, this sentence will have a presupposition that either it is necessary that both ϕ and ψ plus any distributive inference it may have, or the conjunction of ϕ and ψ is not necessary. Keeping this in mind, take the negation of this sentence, *not certain(ϕ and ψ)*, the denotation of which is given in (70).

(70) $[\![\text{not(certain(}\phi \text{ and } \psi\text{))}]\!] = \{\overline{(\Box \bigcup [\![\phi \text{ and } \phi]\!]) \cap (\bigcap_{\mathbf{p} \in [\![\phi \text{ and } \phi]\!] \setminus \{\bigcup [\![\phi \text{ and } \phi]\!]\}} \overline{\Box \mathbf{p}})}\}$

Due to presupposition projection, this sentence should inherit the same presupposition as its positive counterpart. Factoring this presupposition in, the sole proposition in the above set can be simplified to: $\overline{(\Box \bigcup [\![\phi \text{ and } \phi]\!])}$. Importantly, note that this proposition does *not* entail negative free choice since it comprises all worlds in which it is not the case that both ϕ and ψ are certain and, at the same time, includes worlds where either one of them is certain, if the other one is not certain. Hence, the present system predicts that positive sentences of the form *possible(ϕ or ψ)* should give rise to robust FC inferences, while negative sentences of the form *not(certain(ϕ and ψ))* should not. In light of our experimental results, this prediction appears to be borne out.[17] For similar reasons, the system does not derive a negative distributive inference as an entailment. Consider (71).

(71) It is not possible that the mystery box contains both a blue ball and a yellow ball.

This sentence is predicted to be true even if one of the most embedded conjuncts is impossible. To see this more clearly, consider a sentence of the form *not(possible(ϕ and ψ))*, which will denote the singleton set $\{\overline{\bigcap_{\mathbf{p} \in [\![\phi \text{ and } \psi]\!]} \Diamond \mathbf{p}}\}$. Recall now that we follow Goldstein (2019) and assume that a possibility modal has a homogeneity presupposition. When this presupposition is taken into consideration, this denotation gets strengthened to $\{\bigcap_{\mathbf{p} \in [\![\phi \text{ and } \psi]\!]} \overline{\Diamond \mathbf{p}}\}$, i.e., none of the alternative possibilities are possible. Since each alternative possibility $\mathbf{p}$ is a set of possible worlds where ϕ and ψ are both true, this amounts to say that it's impossible that ϕ and ψ be simultaneously true, which does not entail that ϕ or ψ is individually impossible. Therefore, in this system, negative distributive inferences are not derivable as plain entailments.

Finally, as we have already remarked, negative statements are generally non-inquisitive, and non-inquisitive sentences trivially satisfy the sincerity maxim. Hence, when negation takes scope over conjunction, as in (72), no ignorance inference is predicted.

(72) The mystery box does not contain both a blue ball and a yellow ball.

5.2.6 Intermediate negation

Let us now turn to cases where negation takes scope below the modal but above the connective, starting with the intermediate negative case for FC.

(73) It is possible that the box does not contain both a blue ball and a yellow ball.

Recall that negation always returns a singleton alternative possibility and that, with a single alternative possibility, the possibility modal remains essentially classical since, in the absence of multiple alternative possibilities, the additional meaning component that

17 As we discuss below, the present theory can still be integrated with an implicature approach to derive negative free choice.

normally derives FC has nothing to add. Consequently, the predicted meaning of (73) is simply classical, meaning that there is no FC inference in the semantics of this sentence.

Let us now turn to DI for a sentence with negation in an intermediate position, as in (74).

(74) It is certain that the box does not contain both a blue ball and a yellow ball.

Due to the negation that sits immediately below it, the necessity modal only has one alternative possibility to operate on which means that the meaning stays essentially classical. Therefore, the sentence is predicted to have no distributivity inference as entailment.

5.2.7 Low negation

Let us now turn to the cases where negation appears below the connective. For these sentences, the system predicts that, since disjunction is involved and takes scope over negation, the inferences should re-emerge. Consider first the case of free choice in (75).

(75) It is possible that the mystery box either does not contain a blue ball or it does not contain a yellow ball.

The prediction here is the same as for the classical cases of FC. That is, the sentence is predicted to entail that the alternative possibilities denoted by the disjunction are both possible. Concretely, (75) is predicted to entail that it is possible that the box doesn't contain a blue ball and that it is possible that it doesn't contain a yellow ball.

The same applies to the low negative cases for DI. The necessity modal combines with the alternatives of the disjunctive prejacent and requires each of them to be not certain. The resulting meaning for a sentence like (76) is thus as follows: it is certain that the box doesn't contain a blue ball or that it doesn't contain a yellow ball, and it is not certain that it doesn't contain a blue ball and it is not certain that it doesn't contain a yellow ball (i.e., it is possible that it contains one and it is possible that it contains the other).

(76) It is certain that the mystery box either does not contain a blue ball or it does not contain a yellow ball.

Finally, low negative cases of ignorance like (77) generate similar alternative possibilities than their positive counterparts. Therefore, the sincerity maxim applies to these cases in the same way as before, giving rise to the expected ignorance inferences: the speaker doesn't know whether the box contains a blue ball and she doesn't know whether it contains a yellow ball.

(77) The mystery box does not contain a blue ball or it does not contain a yellow ball.

	Predictions				Findings			
	FC	DI	II	SI	FC	DI	II	SI
POSITIVE	✓ (green)	✓ (green)	✓ (green)		✓ (green)	✓ (green)	✓ (green)	✓ (yellow)
NEGATIVE								
High	✗	✗	✗		✗	✗	✗	✓ (yellow)
Intermediate	✗	✗			✗	✗		
Low	✓ (green)	✓ (green)	✓ (green)		✓ (green)	✓ (green)	✓ (green)	✓ (yellow)

Table 10 Predictions of the non-implicature approach (left), alongside our results again (right). Note that this approach is silent about SIs.

5.2.8 Summary of the non-implicature approach

The non-implicature approach to FC, DI and II developed above derives all three types of inferences for the cases involving disjunction, whether negation is absent or takes scope below disjunction. However, for cases involving negated conjunction, none of these inference types are derived as plain entailments. These predictions are largely in line with our experimental results, as summarised in Table 10.

Note however that, in order to have a full account of our results, we still need to say something about scalar implicatures. Fortunately, as previously discussed in Marty et al. 2021, the non-implicature approach can be combined with a regular theory of scalar implicatures to achieve this purpose. The predictions of the resulting approach are summarized in Table 11. As far as we can see, this hybrid approach covers all of the inferences that we investigated in our experiments.

	Predictions				Findings			
	FC	DI	II	SI	FC	DI	II	SI
POSITIVE	✓ (green)	✓ (green)	✓ (green)	✓ (yellow)	✓ (green)	✓ (green)	✓ (green)	✓ (yellow)
NEGATIVE								
High	✓ (yellow)	✓ (yellow)	✓ (yellow)	✓ (yellow)	✗	✗	✗	✓ (yellow)
Intermediate	✓ (yellow)	✓ (yellow)			✗	✗		
Low	✓ (green)	✓ (green)	✓ (green)	✓ (yellow)	✓ (green)	✓ (green)	✓ (green)	✓ (yellow)

Table 11 Predictions of the hybrid approach (on the left), alongside our results (on the right). Note that this approach does <u>not</u> adopt the hypotheses about alternatives in (37) and (49).

There are yet two aspects of our experimental results that remain unexplained by the hybrid approach, neither of which is actually specific to this approach *per se*. First, we consistently found negative SIs to be stronger than their positive counterparts, in line with comparable observations already reported in the literature (Marty et al. 2020). This observation is not accounted for by the hybrid approach and it remains, at this point, an open problem for all theories of scalar implicatures. Second, once we have a way to derive scalar implicatures, we leave room for FC, DI and II to be derived as scalar implicatures in the high and intermediate negative cases. As pointed out earlier for the relevance-based approach, this possibility is not incompatible with our results: the fact that we found no evidence for these negative inferences in our results could simply indicate that these inferences are very weak and, thus, hard to detect.[18] One concern, however, is that, in Exp.3, where both positive and negative SIs were found to quite robust, we still didn't find any evidence for the aforementioned inferences. This observation remains therefore a potential issue, albeit one that is common to virtually every theory adopting the standard mechanisms for deriving scalar implicatures.

Finally, the conceptualisation of the hybrid approach comes with the empirical question of how the two types of alternatives it deals with, formal alternatives and alternative possibilities, interact with one another, if at all. Consider, for instance the modalised disjunctive sentence in (78), where a scalar term (here, *some*) is embedded in one of the disjuncts.

(78) John might have met some of the students this morning or prepared next week's teaching.

This sentence appears to have a reading conveying that *John might have met some but not all of the students* and that *he might have prepared next week's teaching*, that is, a reading including both FC and the regular 'not-all' SI associated with *some*. Is this 'not-all' SI computed locally here? Do the formal alternatives associated with *some* interact with the alternative possibilities introduced by the disjunction? And, finally, what about the formal conjunctive alternative to disjunction, from which one can derive the exclusivity implicature that *it's not possible that John both met some of the students and prepared next week's teaching*? We shall leave these questions for future work exploring the underpinnings of the hybrid approach and its consequences.

6 Conclusion

Sentences involving simple or embedded disjunction, like those in (1)-(3), have long been observed to give rise to robust IGNORANCE, FREE CHOICE, and DISTRIBUTIVE INFERENCES. These inferences all display strong similarities with regular SIs and, for this reason, a prominent take in the literature is to treat them as such.

18 This reasoning holds if we do not adopt the extended alternative-based hypothesis in (37) which, as we discussed, wrongly predicts these inferences to be as robust as their positive counterparts.

(1) IGNORANCE INFERENCE (II)
   a. The box contains either a blue ball or a yellow ball.
   b. ⇝ *The speaker doesn't know whether the box contains a blue ball and she doesn't know whether it contains a yellow ball*

(2) FREE CHOICE INFERENCE (FC)
   a. It is possible that the box contains either a blue ball or a yellow ball.
   b. ⇝ *It's possible that the box contains a blue ball and it's possible that it contains a yellow ball*

(3) DISTRIBUTIVE INFERENCE (DI)
   a. It is certain that the box contains either a blue ball or a yellow ball.
   b. ⇝ *It's not certain that the box contains a blue ball and it's not certain that it contains a yellow ball*

The implicature approach to FC, DI and II has been yet recently challenged by experimental findings showing a variety of differences between these inferences and regular SI. In particular, all three inference types have been found to be more robust, faster to process, and easier to acquire than regular SIs. The general response in the literature has been to explain these differences in reference to the type of alternatives involved in the derivation of these inferences. This hypothesis, repeated below, relies on the fact that, on the implicature approach, the derivation of regular SIs involves alternatives which are derived by lexical substitution, while those of FC, DI and II involve alternatives which are already part of the asserted sentence.

(37) **Alternative-based hypothesis (extended version)**
Alternatives that do not involve lexical substitutions give rise to inferences that are more robust, faster to process and easier to acquire.

In this article, we investigated this hypothesis by looking at other, related constructions for which the implicature approach predicts similar ignorance, free choice and distributive inferences on the basis of the same type of non-lexical alternatives. These constructions, exemplified in (9)-(11), are variants of the positive cases above which involve negated conjunctions.

(9) NEGATIVE IGNORANCE INFERENCE (HIGH NEGATIVE II)
   a. The box does not contain both a blue ball and a yellow ball.
   b. ⇝ *The speaker doesn't know whether the box contains a blue ball and she doesn't know whether it contains a yellow ball*

(10) NEGATIVE FREE CHOICE (HIGH NEGATIVE FC)
   a. It is not certain that the box contains both a blue ball and a yellow ball.

b. ⇝*It's not certain that the box contains a blue ball and it's not certain that it contains a yellow ball*

(11) Negative Distributive Inference (high negative DI)

a. It is not possible that the box contains both a blue ball and a yellow ball.
b. ⇝*It's possible that the box contains a blue ball and it's possible that it contains a yellow ball*

We systematically compared the positive and negative instances of these inference types and compared them to positive and negative regular SIs. We further manipulated the position of negation to create additional logically equivalent cases such as (12) and (13), for free choice as well as for the distributive and ignorance cases.

(12) Intermediate Negative FC

a. It is possible that the box does not contain both a blue ball and a yellow ball.
b. ⇝*It's possible that the box does not contain a blue ball and it is possible that it does not contain a yellow ball*

(13) Low Negative FC

a. It is possible that the box does not contain a blue ball or does not contain a yellow ball.
b. ⇝*It's possible that the box does not contain a blue ball and it is possible that it does not contain a yellow ball*

Our results revealed large differences in robustness between the different cases. Specifically, we found that FC, DI and II were quite robust in cases involving disjunction, whether positive or negative. On the other hand, we found no trace of these inferences in cases involving negated conjunction, regardless of the position of negation. Furthermore, we found no corresponding difference between regular SIs and their negative counterparts. As we discussed, these findings are challenging for the idea that the implicature approach should be supplemented with the hypothesis in (37). This is because it predicts all of them to be similarly robust. In addition, we discuss how another hypothesis about the alternatives of negation, which was recently suggested in the literature, does not help much in reconciling this approach with our results.

We outlined two promising directions to account for our results. The first one retains an implicature-based approach to FC, DI and II, but supplements it with an additional assumption about the calculation of relevance for disjunctive sentences. The second relies instead on a non-implicature approach to these inferences, and may be combined with an implicature approach to derive regular SIs. Both approaches predict robust inferences for the positive and low negative cases – those cases involving disjunction – but not for the other negative cases involving conjunction, consistent with our results.

We end this discussion with two final remarks. First, while our results challenge the hypothesis in (37), we should stress that they do not exclude the one in (36), which restricts its scope of application to processing and acquisition. In other words, it could still be that the reason why these inferences are faster to process and easier to acquire than regular SIs is because their alternatives do not involve lexical substitutions. However, we take our results to teach us that their robustness does not stem from this property.

(36) **Alternative-based hypothesis**
Alternatives that do not involve lexical substitutions give rise to inferences that are faster to process and easier to acquire.

In the face of it, one may wonder why (36) would hold while (37) doesn't. In our view, this could suggest that the relationship between the salience of an alternative and the robustness of the corresponding inference is only indirect. We take this analytical option to point to a psycholinguistic model of alternative-based inferences where the properties of an alternative, like its salience in virtue of being subconstituents of the asserted sentence, can reduce the processing cost and acquisition complexity of the resulting inference, but where the actual computation of this inference depends on further considerations such as the evaluation of whether or not this inference is intended in the first place. This model is very much in line with the relevance-based approach outline above, where the decision to compute scalar implicatures is mediated by the evaluation of the alternatives' relevance, even if these alternatives are active enough. It is less clear to us, at this stage, how the hybrid approach would fit with such a model. Be as it may, before developing this model further, the next natural step would be to investigate the processing and acquisition of the negative versions of FC, DI and II, and test whether they are processed similarly fast and acquired similarly early as their positive counterparts, as the hypothesis in (36) predicts.

Finally, we briefly note that the present study has close connections to some recent work on counterfactuals embedding disjunctive and negated conjunctive sentences (see in particular Romoli, Santorio & Wittenberg 2020, Ciardelli, Linmin & Champollion 2018, Schulz 2019). In fact, one of the hypotheses we discussed, the one in (49), has been developed in response to the finding that disjunctive and negated conjunctive sentences embedded in counterfactuals are interpreted differently by speakers (despite being logically equivalent). Our results, through the differences we found between disjunctions and negated conjunctions, echo the results reported in that literature. An important next step is now comparing in detail the results between the simple and modalised cases we investigated and those involving counterfactuals, as well as testing further the predictions of each approach extended to both of these linguistic environments.

A Training trials and Instructions in Exp.1-4

A.1 Instructions for Exp.1-4

General Instructions

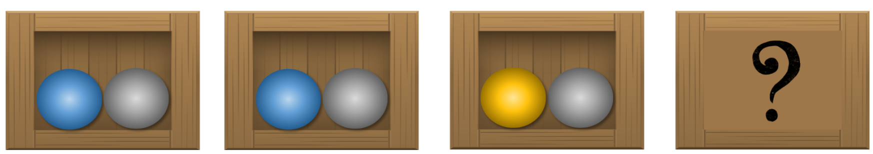
In this study, we will ask for your intuitions about certain kinds of sentences in English. These sentences will be uttered by two characters, Sam and Mia. Here they are:


Sam


Mia

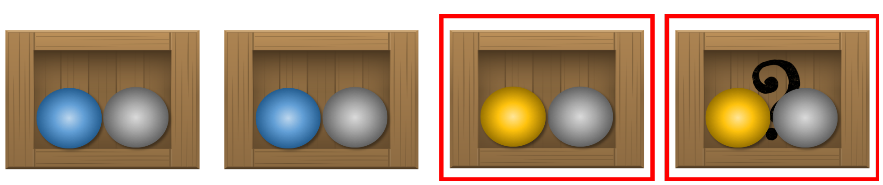
What's inside the mystery box?

Sam and Mia are looking at four boxes containing balls of various colors. Each time, they can only see what's inside the first three boxes, as exemplified below. However, they know that **the fourth box always has the same contents as one of the three open boxes**, and therefore they can make certain inferences about what's inside this mystery box.

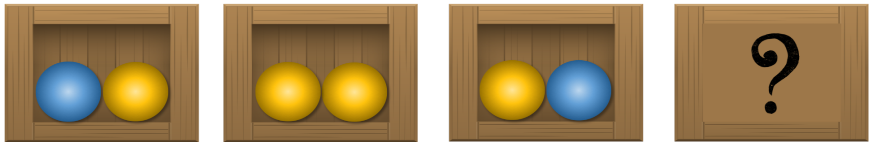


In this case, we can infer for instance that the mystery box must contain a grey ball, and that it can also contain either a blue ball or a yellow ball. Similarly, we can infer that it cannot contain a green ball, and that it cannot contain both a blue ball and a yellow ball.

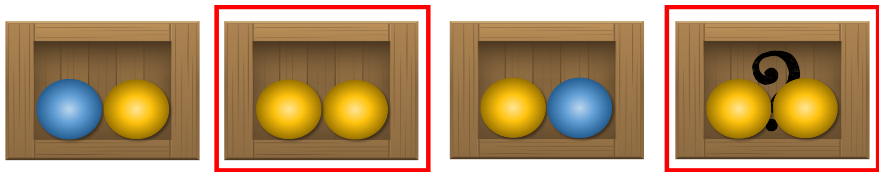
In this case, the contents of the mystery box were identical to those of the third box.



In this second example, we can infer for instance that the mystery box must contain at least one yellow ball. Similarly, we can infer that it need not contain both a blue ball and a yellow one, and that it need not contain a blue ball, etc.



In this case, the contents of the mystery box were identical to those of the second box.



You will see many quadruplets like these ones, each of which will be followed by an utterance from either Sam or Mia about what the mystery box contains. Your task is to decide if this utterance is or is not a good description of what's inside the mystery box. You will click on 'Good' if you consider the utterance a good description of the mystery box's contents; otherwise click on 'Bad'.

Training Phase

You will start with a short training to get you familiar with the study. During this training, you will receive **feedback** on your responses. If you answered correctly, your will see a **green smiley face**; otherwise, you will see a **red frowning face** and be asked to try again. Use this feedback wisely to improve your answers.

We are interested in your spontaneous responses, so please don't think too long before answering.

Test Phase

As in the training, Sam and Mia will produce utterances about what the mystery box contains and you will decide whether these utterances are appropriate descriptions of the pictures you see. From now on, **you will no longer receive feedback on your responses**.

Recall that we are interested in your spontaneous responses, so don't think too long before answering.

A.2 Training trials in Exp.1-4

All participants in our experiments started by completing a first block of training trials. The purpose of the training phase was to ensure that participants got the instructions right and, specifically, that they understood how the possible contents of the mystery box could be inferred from the contents of the open boxes. The training sentences were distinct from the target sentences tested in each experiment, but they involved similar lexical materials. In Exp.1, 2 and 4, participants were presented with two positive sentences of the form

$\Box(\mathsf{A} \wedge \mathsf{B})$ and $\Box(\mathsf{A})$, and two high negative sentences of the form $\neg\Diamond(\mathsf{A} \vee \mathsf{B})$ and $\neg\Diamond(\mathsf{A})$, involving the modals *certain* and *possible*, as shown in (79).

(79) a. It is certain that the mystery box contains both a [A] ball and a [B] ball.
b. It is certain that the mystery box contains a [A] ball.
c. It is not possible that the mystery box contains either a [A] ball or a [B] ball.
d. It is not possible that the mystery box contains a [A] ball.

In Exp.3, participants were presented with logically equivalent *there*-sentences involving the modals *must* and *can*, as shown in (80)

(80) a. There must be both a [A] ball and a [B] ball.
b. There must be a [A] ball.
c. There can't be either a [A] ball or a [B] ball.
d. There can't be a [A] ball.

Training sentences were presented with pictures that made them either clearly true or clearly false. Thus for instance, in the 'Good' training trials, the sentence types in (79b)-(80b) were paired with pictures on which all the boxes contained an A-ball; in the 'Bad' training trials, they were paired with pictures on which all but one box contained an A-ball. Each trial type was repeated 3 times. The result was a set of 24 training trials designed so that the proportion of expected 'Good' and 'Bad' responses was well-balanced within and between sentence type. Throughout the training phase, participants received feedback on the accuracy of their responses (see Appendix A.1 above).

B Predictions from the two hypotheses about alternatives

In the following, we outline the predictions of the implicature approach supplemented with the two hypotheses about alternatives formulated in (37) and (49). We start with some good predictions and then move to more problematic ones.

B.1 Good predictions

The EXH-NEG scalemate hypothesis in (49) has the effect of generating extra alternatives when negation is involved. In the high negative cases, once these extra alternatives are factored in, no inference is predicted to arise, in line with our results. To illustrate, consider again how the implicature approach derives FC for simple positive cases like (81), assuming the parse in (82) and the set of formal alternatives in (83).

(81) It is possible that the box contains A or B. $\Diamond(\mathsf{a} \vee \mathsf{b})$

(82) EXH[It is possible A or B]

(83) $$\left\{\begin{array}{ll} \text{It is possible A or B} & \Diamond(a \lor b) \\ \text{It is possible A} & \Diamond a \\ \text{It is possible B} & \Diamond b \\ \text{It is possible A and B} & \Diamond(a \land b) \end{array}\right\}$$

The only innocently excludable alternative is $\Diamond(a \land b)$, while the includable alternatives are: $\{\Diamond(a \lor b), \Diamond a, \Diamond b\}$. Thus, after exclusion and inclusion, we obtain the FC meaning of the sentence in (84): it is possible that the box contains a blue ball, it is also possible that it contains a yellow ball, but it is not possible that it contains both.

(84) EXH[It is possible A or B] = $\Diamond(a \lor b) \land \boxed{\Diamond a \land \Diamond b} \land \neg\Diamond(a \land b)$

The approach extends to NFC cases like (85), assuming the alternatives in (86). The only innocently excludable alternative is $\neg\Box(a \lor b)$, while the includable ones are: $\{\neg\Box(a \land b), \neg\Box a, \neg\Box b\}$. Thus, after exclusion and inclusion, we obtain the NFC meaning of the sentence in (87).

(85) EXH[It is not certain that A and B]

(86) $$\left\{\begin{array}{ll} \text{not[it is certain that A and B]} & \neg\Box(a \land b) \\ \text{not[it is certain that A]} & \neg\Box a \\ \text{not[it is certain that B]} & \neg\Box b \\ \text{not[it is certain that A or B]} & \neg\Box(a \lor b) \end{array}\right\}$$

(87) EXH[it is not certain that a and b] = $\neg\Box(a \land b) \land \boxed{\neg\Box a \land \neg\Box b} \land \Box(a \lor b)$

So far, this is the standard predictions of the implicature approach. Crucially, however, once we adopt the hypothesis in (49), the set of formal alternatives expands and NFC is no longer predicted:

(88) $$\left\{\begin{array}{ll} \text{not[it is certain that A and B]} & \neg\Box(a \land b) \\ \text{not[it is certain that A]} & \neg\Box a \\ \text{not[it is certain that B]} & \neg\Box b \\ \text{not[it is certain that A or B]} & \neg\Box(a \lor b) \\ \text{EXH[it is certain that A and B]} & \Box(a \land b) \\ \text{EXH[it is certain that A]} & \Box a \land \neg\Box b \\ \text{EXH[it is certain that B]} & \Box b \land \neg\Box a \\ \text{EXH[it is certain that A or B]} & \Box(a \lor b) \land \neg\Box(a \land b) \end{array}\right\}$$

(89) EXH[It is not certain A and B] = $\neg\Box[a \land b]$

As the reader can verify, the presence of the additional alternatives EXH[$\Box$a] and EXH[$\Box$b] make it so that the alternatives $\neg\Box$a and $\neg\Box$b are no longer innocently includable. Consequently, on this view, the predicted meaning does not entail negative free choice. In sum, the assumption in (49) blocks NFC and, in so doing, it can account for the contrasts in robustness between FC and NFC that we found in our results.

The same reasoning applies to negative distributive inferences. When a sentence like (90) is exhaustified with respect to the alternatives in (91), a distributive inference results from the negation of the innocently excludable alternatives $\neg\Diamond$a and $\neg\Diamond$b.

(90) EXH[It is not possible that A and B] = $\neg\Diamond(a \wedge b) \wedge \boxed{\Diamond a \wedge \Diamond b} \wedge \Diamond(a \vee b)$

(91)
$$\left\{\begin{array}{ll} \text{not[it is possible that A and B]} & \neg\Diamond(a \wedge b) \\ \text{not[it is possible that A]} & \neg\Diamond a \\ \text{not[it is possible that B]} & \neg\Diamond b \\ \text{not[it is possible that A or B]} & \neg\Diamond(a \vee b) \end{array}\right\}$$

However, when we consider the additional alternatives predicted by (49), the alternatives $\neg\Diamond$a and $\neg\Diamond$b become non-excludable (and non-includable) and, therefore, no distributive inference is predicted. As in the case of free choice, this could account for the relative weakness of the negative distributive case with respect to its positive counterpart.

(92)
$$\left\{\begin{array}{ll} \text{not[it is possible that A and B]} & \neg\Diamond(a \wedge b) \\ \text{not[it is possible that A]} & \neg\Diamond a \\ \text{not[it is possible that B]} & \neg\Diamond b \\ \text{not[it is possible that A or B]} & \neg\Diamond(a \vee b) \\ \text{EXH[it is possible that A and B]} & \Diamond(a \wedge b) \\ \text{EXH[it is possible that A]} & \Diamond a \wedge \neg\Diamond b \\ \text{EXH[it is possible that B]} & \Diamond b \wedge \neg\Diamond a \\ \text{EXH[it is possible that A or B]} & \Diamond(a \vee b) \wedge \neg\Diamond(a \wedge b) \end{array}\right\}$$

In sum, the EXH-NEG scalemate hypothesis allows us to account for the contrasts between positive vs. high negative cases of FC and DI. It is not clear, however, whether it can help with the corresponding intermediate and low cases, as well as with II more generally. In addition, it makes wrong predictions for positive vs. negative regular SIs, as we shall see in more detail below.

B.2 Problematic predictions

The implicature approach supplemented with the hypothesis in (49) can account for the differences we found between the positive and negative cases of FC and DI in Exp.1. However, when we move to the version of these cases tested in Exp.2-3, this hypothesis makes incorrect predictions. Recall that, in Exp.2, we tested sentences like (12) and (46a).

(12) INTERMEDIATE NEGATIVE FC $\Diamond\neg(a \wedge b)$
It is possible that the box does not contain both a blue ball and a yellow ball.
$\rightsquigarrow$*It's possible that the box does not contain a blue ball and it's possible that it does not contain a yellow ball*

(46a) INTERMEDIATE NEGATIVE DI $\Box\neg(a \wedge b)$
It is certain that the box does not contain both a blue ball and a yellow ball.
$\rightsquigarrow$*It's certain that the box does not contain a blue ball and it's certain that it does not contain a yellow ball*

As discussed, the implicature approach predicts that logically equivalent sentences should have the same inferences, hence no difference is expected here. However, once we adopt the hypothesis in (49), it turns out that the position of negation in these constructions matter for the overall result. In particular, for (12) and (46a), one can show that the addition of EXH-alternatives lead to strengthen, rather than block the relevant inferences, in contrast to what we saw for the high negative cases. To illustrate, consider the expanded set of alternatives for a sentence like (12):

(93)
$$\left\{\begin{array}{ll} \text{[It is possible that not [A and B]]} & \Diamond\neg(a \wedge b) \\ \text{[It is possible that not [A]]} & \Diamond\neg a \\ \text{[It is possible that not [B]]} & \Diamond\neg b \\ \text{[It is possible that not [A or B]]} & \Diamond\neg(a \vee b) \\ \text{[It is possible that } \textsc{exh}\text{[A and B]]} & \Diamond(a \wedge b) \\ \text{[It is possible that } \textsc{exh}\text{[A]]]} & \Diamond[a \wedge \neg b] \\ \text{[It is possible that } \textsc{exh}\text{[B]]} & \Diamond[b \wedge \neg a] \\ \text{[It is possible that } \textsc{exh}\text{[A or B]]} & \Diamond[(a \vee b) \wedge \neg(a \wedge b)] \end{array}\right\}$$

An exhaustification process is now also expected to also take place below the possibility modal, where negation appears. As it turns out, all the alternatives above are innocently includable and, therefore, negative free choice is predicted; in fact, as shown below, a stronger version of this inference is predicted.

(94) $\textsc{exh}\text{[It is possible not(A and B)]} = \Diamond\neg(a \wedge b) \wedge \boxed{\Diamond\neg a \wedge \Diamond\neg b} \wedge \Diamond\neg(a \vee b)$

Similarly, in the low negative cases, the account predicts that FC should re-emerge and, in fact, that it should re-emerge in a stronger form than just negative FC as the additional alternatives $\Diamond\textsc{exh}[a]$ and $\Diamond\textsc{exh}[b]$ can also be included:

(95)
$$\left\{\begin{array}{lc}
\text{[It is possible that not [A] or not [B]]} & \Diamond(\neg a \vee \neg b) \\
\text{[It is possible that not [A]]} & \Diamond\neg a \\
\text{[It is possible that not [B]]} & \Diamond\neg b \\
\text{[It is possible that not [A] and not [B]]} & \Diamond(\neg a \wedge \neg b) \\
\text{[It is possible that \textsc{exh}[A] or \textsc{exh} [B]]} & \Diamond((a \wedge \neg b) \vee (b \wedge \neg a)) \\
\text{[It is possible that \textsc{exh}[A]]} & \Diamond[a \wedge \neg b] \\
\text{[It is possible that \textsc{exh}[B]]} & \Diamond[b \wedge \neg a] \\
\text{[It is possible that \textsc{exh} [A] and \textsc{exh} [B]]} & \bot
\end{array}\right\}$$

(96) $\textsc{exh}$[It is possible (not A or not B)] =
$$\Diamond(\neg a \vee \neg b) \wedge \boxed{\Diamond\neg a \wedge \Diamond\neg b} \wedge \Diamond[a \wedge \neg b] \wedge \Diamond[b \wedge \neg a]$$

In sum, this approach predicts that the intermediate and low negative cases should pattern like the positive ones and differently from the high negative ones. Our results show that, while the low negative cases do pattern like the positive ones, there is no robust inferences associated with the intermediate negative cases, contrary to these predictions.[19]

19 For FC, we can actually improve on these predictions, if we also consider replacing the existential modal with its universal counterpart. In particular, one can show that, if we consider the following expanded set of alternatives for the intermediate cases, no FC inference is predicted anymore, in line with our results.

(i)
$$\left\{\begin{array}{lc}
\text{[It is possible that not [A and B]]} & \Diamond\neg(a \wedge b) \\
\text{[It is possible that not [A]]} & \Diamond\neg a \\
\text{[It is possible that not [B]]} & \Diamond\neg b \\
\text{[It is possible that not [A or B]]} & \Diamond\neg(a \vee b) \\
\text{[It is possible that \textsc{exh}[A and B]]} & \Diamond(a \wedge b) \\
\text{[It is possible that \textsc{exh}[A]]]} & \Diamond[a \wedge \neg b] \\
\text{[It is possible that \textsc{exh}[B]]} & \Diamond[b \wedge \neg a] \\
\text{[It is possible that \textsc{exh}[A or B]]} & \Diamond[(a \vee b) \wedge \neg(a \wedge b)] \\
\text{[It is certain that not [A and B]]} & \Box\neg(a \wedge b) \\
\text{[It is certain that not [A]]} & \Box\neg a \\
\text{[It is certain that not [B]]} & \Box\neg b \\
\text{[It is certain that not [A or B]]} & \Box\neg(a \vee b) \\
\text{[It is certain that \textsc{exh}[A and B]]} & \Box(a \wedge b) \\
\text{[It is certain that \textsc{exh}[A]]]} & \Box[a \wedge \neg b] \\
\text{[It is certain that \textsc{exh}[B]]} & \Box[b \wedge \neg a] \\
\text{[It is certain that \textsc{exh}[A or B]]} & \Box[(a \vee b) \wedge \neg(a \wedge b)]
\end{array}\right\}$$

(ii) $\textsc{exh}$[It is possible not(A and B)] = $\Diamond\neg(a \wedge b)$

Importantly, we note that the corresponding expanded set for the low negative cases still gives rise to the negative FC inference, as expected.

B.3 More problematic predictions

Recall that, on the implicature approach, the negative sentence in (97) is compared to the alternatives in (98), deriving the ignorance inferences that the speaker doesn't know whether the box contains a blue ball and doesn't know whether it contains a yellow ball.

(97) The box doesn't contain both a blue ball and a yellow ball $\neg(\mathsf{a} \wedge \mathsf{b})$

(98)
$$\left\{ \begin{array}{ll} \text{not[the box contains a blue ball]} & \neg \mathsf{a} \\ \text{not[the box contains a yellow ball]} & \neg \mathsf{b} \end{array} \right\}$$

With the assumption in (49), the set of alternatives in (98) is enlarged as in (99), from which we also derive the implicatures that the speaker doesn't know whether the box *only contains a blue ball* and doesn't know whether it *only contains a yellow ball*. Since these inferences are already entailed by those we derived before, nothing is added here.

(iii)
$$\left\{ \begin{array}{ll}
\text{[It is possible that not [A] or not [B]]} & \Diamond(\neg a \vee \neg b) \\
\text{[It is possible that not [A]]} & \Diamond\neg a \\
\text{[It is possible that not [B]]} & \Diamond\neg b \\
\text{[It is possible that not [A] and not [B]]} & \Diamond(\neg a \wedge \neg b) \\
\text{[It is possible that EXH[A] or EXH [B]]} & \Diamond((a \wedge \neg b) \vee (b \wedge \neg a)) \\
\text{[It is possible that EXH[A]]} & \Diamond[a \wedge \neg b] \\
\text{[It is possible that EXH[B]]} & \Diamond[b \wedge \neg a] \\
\text{[It is possible that EXH [A] and EXH [B]]} & \bot \\
\text{[It is certain that not [A] or not [B]]} & \Diamond(\neg a \vee \neg b) \\
\text{[It is certain that not [A]]} & \Box\neg a \\
\text{[It is certain that not [B]]} & \Box\neg b \\
\text{[It is certain that not [A] and not [B]]} & \Box(\neg a \wedge \neg b) \\
\text{[It is certain that EXH[A] or EXH [B]]} & \Box((a \wedge \neg b) \vee (b \wedge \neg a)) \\
\text{[It is certain that EXH[A]]} & \Box[a \wedge \neg b] \\
\text{[It is certain that EXH[B]]} & \Box[b \wedge \neg a] \\
\text{[It is certain that EXH [A] and EXH [B]]} & \bot
\end{array} \right\}$$

(iv) EXH[It is possible (not A or not B)] = $\Diamond(\neg a \vee \neg b) \wedge \boxed{\Diamond\neg a \wedge \Diamond\neg b} \wedge \Diamond[a \wedge \neg b] \wedge \Diamond[b \wedge \neg a]$

In other words, allowing multiple replacements makes the hypothesis in (49) consistent with our results for negative FC across all three cases (high, intermediate, and low). Extending this approach to DI is however more complicated for at least two reasons. First, as Fox (2007) and Bar-Lev & Fox (2020) discuss, when we allow multiple replacements, we can no longer account for distributive inferences across all cases, unless we consider LFs with recursive exhaustification. Second, Crnič, Chemla & Fox (2015) and Bar-Lev & Fox (2020) pointed out that, if we allow multiple replacements across the board, we can no longer account for certain differences between cases involving modals and cases involving their corresponding nominal quantifiers. For these reasons, we will not explore how to extend this approach to the DI cases; rather, we will focus on more problematic predictions having to do with II and regular SIs.

Hence, the assumption in (49) does not make any difference for II, and so the differences between positive and negative II is left unaccounted for.[20]

(99)
$$\left\{\begin{array}{ll} \text{not[the box contains a blue ball]} & \neg\text{a} \\ \text{not[the box contains a yellow ball]} & \neg\text{b} \\ \textsc{exh}\text{[the box contains a blue ball]} & \textsc{exh}[\text{a}] \\ \textsc{exh}\text{[the box contains a yellow ball]} & \textsc{exh}[\text{b}] \end{array}\right\}$$

Moreover, as discussed in Marty et al. 2021 and Romoli, Santorio & Wittenberg 2020, when we move to regular SIs, this approach makes wrong predictions altogether. The reason is that (49) makes it so that negative regular SIs should arise less often than their positive counterparts. To illustrate, consider the example in (100).

(100) It is not certain that the box contains a blue ball.
⇝ *It is possible that the box contains a blue ball*

This inference can be derived by exhaustifying the meaning of (100) against the formal alternatives in (101). As usual, the alternative $\neg\Diamond$a is excludable, and its exclusion gives rise to the observed implicature, i.e., $\Diamond$a.

(101)
$$\left\{\begin{array}{ll} \text{not[it is certain that the box contains a blue ball]} & \neg\Box\text{a} \\ \text{not[it is possible that the box contains a blue ball]} & \neg\Diamond\text{a} \end{array}\right\}$$

The problem now is that the EXH-NEG scalemate hypothesis brings in more alternatives which block the derivation of this inference. For illustration, consider the alternatives that we would obtain on this hypothesis:

(102)
$$\left\{\begin{array}{ll} \text{not[it is certain that the box contains a blue ball]} & \neg\Box\text{a} \\ \text{not[it is possible that the box contains a blue ball]} & \neg\Diamond\text{a} \\ \textsc{exh}\text{[it is certain that the box contains a blue ball]} & \textsc{exh}[\Box\text{a}] \\ \textsc{exh}\text{[it is possible that the box contains a blue ball]} & \textsc{exh}[\Diamond\text{a}] \end{array}\right\}$$

The presence of the alternative $\textsc{exh}[\Diamond[\text{a}]]$, equivalent to $\Diamond\text{a} \wedge \neg\Box\text{a}$, makes the alternative $\neg\Diamond$a non-excludable (neither of them is includable either, given the presence of the other). The result is that the negative implicature above is no longer derived.[21]

20 We note that, under the grammatical approach, the predictions for II would improve if we were to allow multiple replacements and recursive exhaustification, as discussed for the DI cases. This move, however, would require, among other things, to postulate an existential lexical alternative to the matrix K operator; we leave this avenue of research for future work.

21 Bar-Lev & Fox (2020: fn. 59) discuss this issue. They suggest that the problem can be alleviated if we take into account the sensitivity of SIs to the focus structure of the sentence. Following Fox &

To sum up this overview, the assumption that negation and EXH are alternatives allows to account for the difference between positive and negative FC and DI, but it doesn't help with the corresponding difference with II and it incorrectly extends to regular SIs, predicting negative SIs to be weaker than their positive counterparts.

References

Aloni, Maria. 2003. Free choice in modal contexts. In *Proceedings of sinn und bedeutung 7*, 25–37.

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

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

Anwyl-Irvine, A. L., J. Massonie, A. Flitton, N. Z. Kirkham & J. K. Evershed. 2020. Gorilla in our midst: an online behavioral experiment builder. *Behavior Research Methods* 52. 388–407.

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. https://doi.org/10.1007/s11050-020-09162-y.

Barner, David, N. Brooks & Alan Bale. 2011. Accessing the unsaid: the role of scalar alternatives in children's pragmatic inferences. *Cognition* 188. 87–96.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2014. Fitting linear mixed-effects models using lme4.

Bates, Douglas, Martin Maechler & Ben Bolker. 2011. *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. https://doi.org/10.1016/j.jml.2004.05.006.

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

Buccola, Brian & Andreas Haida. 2019. Obligatory irrelevance and the computation of ignorance inferences. *Journal of Semantics* 36(4). 583–616.

Katzir (2011), among others, they argue that only alternatives of items within the focus constituent of a sentence are considered for implicature computation. On this view, in a sentence like (100), negation can be replaced by EXH just in case it is part of the focus constituent. This leaves room for deriving negative SIs when negation is not part of the focused constituent of the sentence. However, as Romoli, Santorio & Wittenberg (2020) and Marty et al. (2021) discuss, we still observe robust negative SIs when negation is included in the focus constituent through an explicit Question Under Discussion (e.g., *What's up with you?*).

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

Chemla, Emmanuel. 2010. Similarity: towards a unified account of scalar implicatures, free choice permission and presupposition projection. Ms., ENS.

Chemla, Emmanuel & Lewis Bott. 2014. Processing inferences at the semantics/pragmatics frontier: disjunctions and free choice. *Cognition* 130(3). 380–396.

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

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

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. https://doi.org/10.1007/s10988-018-9232-4.

Coppock, Liz & Thomas Brochhagen. 2013. Raising and resolving issues with scalar modifiers. *Semantics and Pragmatics* 3(6). 1–57.

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

Crnič, Luka, Emmanuel Chemla & Danny Fox. 2015. Scalar implicatures of embedded disjunction. *Natural Language Semantics* 23(4). 271–305. https://doi.org/10.1007/s11050-015-9116-x.

De Neys, Wim & Walter Schaeken. 2007. When people are more logical under cognitive load: dual task impact on scalar implicature. *Experimental psychology* 54(2). 128–133.

Enguehard, Émile & Emmanuel Chemla. 2021. Connectedness as a constraint on exhaustification. *Linguistics and Philosophy* 44. 79–112. https://doi.org/10.1007/s10988-019-09286-3.

Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and implicature in compositional semantics*, 71–120. Palgrave.

Fox, Danny. 2016. On why ignorance might not be part of literal meaning. Commentary on Marie-Christine Meyer at the MIT Workshop on Exhaustivity, Sept. 2016.

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

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

Gazdar, Gerald. 1979. *Pragmatics: implicature, presupposition, and logical form*. New York: Academic Press.

Geurts, Bart. 2010. *Quantity implicatures*. Cambridge University Press.

Goldstein, Simon. 2019. Free choice and homogeneity. *Semantics and Pragmatics* 12(23). 1–47. https://doi.org/10.3765/sp.

Grice, Paul. 1975. Logic and conversation. in The Logic of Grammar, D. Davidson and G. Harman (eds), Encino, CA: Dickenson, 64-75.

Groenendijk, Jeroen & Martin Stokhof. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. University of Amsterdam dissertation.

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. https://doi.org/10.1016/j.jml.2007.11.007.

Kamp, Hans. 1974. Free choice permission. *Proceedings of the Aristotelian Society* 74. 57–74.

Kamp, Hans. 1978. Semantics versus pragmatics. In Franz Guenther & Siegfred Schmidt (eds.), *Formal semantics and pragmatics for natural languages*, 255–287. Dordrecht.

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

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

Lenth, Russell, Henrik Singmann, Jonathon Love, Paul Buerkner & Maxime Herve. 2018. *Emmeans: Estimated marginal means, aka least-squares means*.

Marty, Paul. 2017. *Implicatures in the DP domain*. MIT dissertation.

Marty, Paul & Jacopo Romoli. 2021. Presupposed free choice and the theory of scalar implicatures. *Linguistics and Philosophy*.

Marty, Paul, Jacopo Romoli, Yasu Sudo, Bob van Tiel & Richard Breheny. 2020. Processing implicatures: a comparison between direct and indirect sis. *Talk presented at the first Experiments on Linguistic Meaning (ELM) Conference*. https://campuspro-uploads.s3-us-west-2.amazonaws.com/2a1da9b9-a2cb-4a63-b175-3c919aa79aca/1fc946ff-35a6-427f-b473-efae316fc584/ELM-Directvs.IndirectSIs-Martyetal.(2020).pdf.

Marty, Paul, Jacopo Romoli, Yasutada Sudo & Richard Breheny. 2021. Negative Free Choice. Unpublished ms., UCL and University of Bergen.

Marty, Paul Pierre & Emmanuel Chemla. 2013. Scalar implicatures: working memory and a comparison with only. *Frontiers in psychology* 4. 403.

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

Pagliarini, Elena, Cory Bill, Lyn Tieu & Stephen Crain. 2018. On children's variable success with scalar inferences: insights from disjunction in the scope of a universal quantifier. *Cognition* 178. 178–192. https://semanticsarchive.net/Archive/mE0YjdkN/PagliariniBillRomoliTieuCrain_DistributiveInferencesAcq.pdf.

R Core Team. 2013. *R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.*

Romoli, Jacopo. 2012. *Soft but strong: neg-raising, soft triggers, and exhaustification.* Harvard University dissertation. https://dash.harvard.edu/bitstream/handle/1/9909638/Romoli_gsas.harvard_0084L_10566.pdf?sequence=3.

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).

Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27(3). 367–391.

Schulz, Katrin. 2019. The similarity approach strikes back: negation in counterfactuals. In *Proceedings of sinn und bedeutung 22*, vol. 2, 343–360. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/110.

Simons, Mandy. 2001. Disjunction and alternativeness. *Linguistics and Philosophy* 24. 597–619.

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

Singh, Raj. 2008. On the interpreation of disjunction: asymmetric, incremental and eager for inconsistency. *Linguistics and Philosophy* 31. 245–260.

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.

van Tiel, Bob, Paul Marty, Elizabeth Pankratz & Chao Sun. 2019. Scalar inferences and cognitive load. In *Proceedings of sinn und bedeutung*, vol. 23, 427–442.

van Tiel, Bob, Elizabeth Pankratz & Chao Sun. 2019. Scales and scalarity: processing scalar inferences. *Journal of Memory and Language* 105. 93–107.

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.

Van Tiel, Bob & Walter Schaeken. 2017. Processing conversational implicatures: alternatives and counterfactual reasoning. *Cognitive Science* 105. 93–107.

Willer, Malte. 2017. Widening free choice. In *Proceedings from the amsterdam colloquium 2017*.

Zimmerman, Thomas Ede. 2000. Free choice disjunction and epistemic possibility. *Natural Language Semantics* 8(255–290).