# On the source of distributive inferences*

Paul Marty
*L-Università ta' Malta*

Sonia Ramotowska
*Heinrich-Heine Universität Düsseldorf*

Richard Breheny
*University College London*

Jacopo Romoli
*Heinrich-Heine Universität Düsseldorf*

Yasutada Sudo
*University College London*

**Abstract** This paper reports on two experiments investigating the relationship between DISTRIBUTIVE and NEGATED UNIVERSAL inferences arising from disjunction embedded within a universal quantifier. It has been claimed that DISTRIBUTIVE inferences can be derived independently from NEGATED UNIVERSAL inferences with *nominal*, but not with *modal* quantifiers (Booth 2022, Crnič et al. 2015, Fusco 2015, Sayre-McCord 1986). Experiment 1 tested this claim by comparing cases involving the determiner *every* and cases involving the modal *must*, where *must* expressed epistemic necessity. Experiment 2 followed up on Experiment 1 by testing the same two quantifiers, only this time the modal *must* expressed deontic necessity. The results from both experiments show that DISTRIBUTIVE inferences may arise independently of NEGATED UNIVERSAL inferences with both types of operator. While the findings for *every* essentially replicate those from Crnič et al. 2015, the findings for *must* are new and go against the aforementioned claim. Furthermore, the response time results from both experiments show that DISTRIBUTIVE inferences are associated with response delay effects in the opposite direction to those generally observed for regular scalar implicatures, raising a new challenge for (some versions of) the implicature-based account of these inferences. We discuss the prospects of non-implicature accounts such as Aloni 2022.

**Keywords:** distributive inferences, universal modals, scalar implicatures, neglect-zero, response delay effects, pragmatic reasoning

* We would like to thank Maria Aloni, Moysh Bar-Lev, Itai Bassi, Kyle Blumberg, Ivano Ciardelli, Luka Crnič, Marco Degano, Danny Fox, Melissa Fusco, Paloma Jeretič, Matt Mandelkern, Paolo Santorio, Florian Schwarz, Benjamin Spector, and the audience at the Amsterdam Colloquium 2022. This research was supported by the Leverhulme Trust Grant RPG-2018-425.

Consider a situation where there are three open boxes, each of which contains one or two balls. In this situation, an utterance of (1), where disjunction occurs in the scope of *every*, can give rise to the inferences in (1a) and (1b). Theses inferences are generally referred to as 'Distributive' inferences, henceforth D-inferences.

(1) Every box contains a yellow ball or a blue ball. $\forall x(\mathsf{A}x \vee \mathsf{B}x)$

  a. $\rightsquigarrow$ *Some box contains a yellow ball.* $\exists x\mathsf{A}x$

  b. $\rightsquigarrow$ *Some box contains a blue ball.* $\exists x\mathsf{B}x$

While the existence of these inferences is uncontroversial, their status and source are still debated. Traditionally, these inferences have been analysed as Scalar Implicatures (SIs), derived through negating the stronger alternatives to (1) in (2) (Fox 2007, Sauerland 2004 a.o.). The meaning of (1), together with the negations of (2a) and (2b), entail the D-inferences above.

(2) a. Every box contains a yellow ball. $\forall x\mathsf{A}x$

  b. Every box contains a blue ball. $\forall x\mathsf{B}x$

This account predicts that D-inferences should always arise in combination with the 'Negated Universal' inferences in (3a) and (3b), henceforth NU-inferences.

(3) a. $\rightsquigarrow$ *Not every box contains a yellow ball.* $\neg\forall x\mathsf{A}x$

  b. $\rightsquigarrow$ *Not every box contains a blue ball.* $\neg\forall x\mathsf{B}x$

Crnič et al. 2015, in a study that we review below, provide experimental evidence that partly disconfirms this prediction. In particular, their results suggest that, when disjunction is embedded under a *nominal* universal quantifier, D-inferences can arise independently of NU-inferences. Crnič et al. submit, however, that the relevant prediction is born out when disjunction is embedded under a *modal* universal quantifier. Specifically, the authors claim, based on their own judgments, that D-inferences cannot be observed independently of NU-inference with universal modals (see Booth 2022, Fusco 2015, Sayre-McCord 1986 for related claims).

After reviewing the traditional approach and the challenge raised by Crnič et al.'s study in §1, we report in §2 on the results of two experiments comparing the availability and time course of D-inferences arising from disjunction under universal determiners and universal modals. The results show that (i) D-inferences can arise independently of NU-inferences with both types of quantifier and that (ii) not deriving D-inferences incurs a processing slowdown, unlike what is commonly observed for regular SIs. In §3, we present two alternative accounts for these findings, and discuss how they fare with respect to our data. Section §4 concludes.

## 1 Background

### 1.1 The traditional approach

The traditional approach to D-inferences derive these inferences through negating the alternatives corresponding to the embedded disjuncts. This approach can be implemented in a neo-Gricean or a grammatical account of SIs. For concreteness, we exemplify the gist of this approach using the latter.

On the grammatical account, SIs are derived through the use of an exhaustivity operator, EXH, which excludes all innocently excludable alternatives to its prejacent. The definition of EXH is given in (4); the auxiliary notion of Innocent Exclusion (IE) is given in (5), where '$C$' is the set of contextually relevant alternatives.

(4) $[\![\textsf{EXH}]\!](C)(p)(w) \Leftrightarrow p(w) \wedge \forall q \in \textsf{IE}(p,C)[\neg[\![q]\!](w)]$

(5) $\textsf{IE}(p,C) := \bigcap \left\{ C' \;\middle|\; \begin{array}{l} C' \text{ is a maximal subset of } C \text{ such that} \\ \exists w[p(w) \wedge \forall q \in C'[\neg[\![q]\!](w)]] \end{array} \right\}$

With this in mind, consider again the sentence in (1). This sentence has, among others, the universally quantified sentences in (6) as alternatives, both of which are structurally simpler, stronger and innocently excludable.

(6) $\left\{ \begin{array}{ll} \text{Every box contains a yellow ball.} & \forall x\, \textsf{A}x \\ \text{Every box contains a blue ball.} & \forall x\, \textsf{B}x \end{array} \right\}$

The result of exhaustifying the meaning of (1) on the basis of these alternatives delivers NU-inferences, through which the D-inferences of interest follow:

(7) $[\![\textsf{EXH}[\text{Every box contains a yellow ball or a blue ball}]]\!]$
$\Leftrightarrow \forall x(\textsf{A}x \vee \textsf{B}x) \wedge \neg\forall x(\textsf{A}x) \wedge \neg\forall x(\textsf{B}x)$
$\Rightarrow \exists x(\textsf{A}x) \wedge \exists x(\textsf{B}x)$

In sum, the traditional approach derives D-inferences through SIs in a straightforward way. Crucially, on this approach, D-inferences derivationally depend on NU-implicatures in that, for the former to arise, the latter must be derived. This, in turn, predicts that D-inferences cannot arise in the absence of NU-inferences.

### 1.2 Crnič et al. 2015 and the extension to modals

Crnič et al. 2015 report experimental data that challenges the key prediction of the traditional approach. In their study, Crnič et al. found that *every*-sentences like (1) were robustly judged as true when their D-inferences are true while their NU-inferences are false, e.g., in situations where all boxes contain a yellow ball and some of them also contain a blue one. By contrast, they found that these same sentences were rejected when their D-inferences and NU-inferences are both false, e.g., in situations where all boxes contain a yellow ball but none of them also contain a blue one.[1] These findings suggest that D-inferences can arise independently of NU-inferences, contrary to what is predicted by the traditional approach.[2]

Crnič et al. 2015 argue that their findings do not extend to analogous cases involving universal modals. To illustrate, imagine a situation where there are four boxes. We can see what's inside the first three boxes but not what's inside the last one, let us call it 'the mystery box'; we know that the mystery box has the same contents as one of the three open boxes, but we don't know which. Consider now the modal variant of (1) in (8), where disjunction appears this time in the scope of *must*. The corresponding D-inferences are given in (8a) and (8b). In this situation, these inferences are true only if at least one of the three open boxes contains a yellow ball and at least one of them contains a blue ball.

(8) The mystery box must contain a yellow ball or a blue ball. $\Box(\mathsf{A} \vee \mathsf{B})$

a. $\rightsquigarrow$ *The mystery box might contain a yellow ball.* $\Diamond\mathsf{A}$

b. $\rightsquigarrow$ *The mystery box might contain a blue ball.* $\Diamond\mathsf{B}$

Crnič et al. claim that, with universal modals, D-inferences cannot arise in the absence of NU-inferences (see Booth 2022, Fusco 2015, Sayre-McCord 1986 for related claims).[3] In other words, they claim that the D-inferences associated with a sentence like (8) can only arise with the NU-inferences in (9a) and (9b), which arise through negating the simpler universal alternatives to (8), $\Box\mathsf{A}$ and $\Box\mathsf{B}$.

(9) a. $\rightsquigarrow$ *The mystery box doesn't have to contain a yellow ball.* $\neg\Box\mathsf{A}$

b. $\rightsquigarrow$ *The mystery box doesn't have to contain a blue ball.* $\neg\Box\mathsf{B}$

If this claim is right, then D-inferences should go hand-in-hand with NU-inferences for disjunction under universal modals, as predicted by the traditional approach. Given the previous findings from Crnič et al. 2015, one would thus expect *every*-sentences like (1) and *must*-sentences like (8) to be judged differently in the critical situations where their D-inferences are true while their NU-inferences are false: simplifying a bit, the former should be accepted while the latter should be rejected.

1 The sentences tested in Crnič et al. 2015 involved boxes and letters, rather than boxes and balls. The examples in (1) and (8) are used for consistency with the stimuli in our experiments.

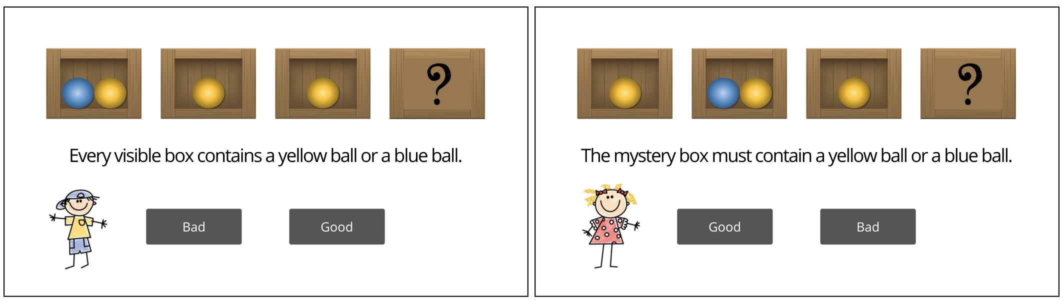
2 Crnič et al. (2015) propose a different version of the implicature approach. We will not review this proposal here, but see Bar-Lev & Fox 2023 for a critical discussion.

3 Crnič et al. discussed a different example, however analogous to (8). See footnote 7 for discussion.

## 1.3 The present study

We carried out two acceptability experiments comparing the availability and time course of D-inferences arising from disjunction under nominal and modal quantifiers. Experiment 1 tested *every*-sentences like (1) and *must*-sentences like (8), where *must* expressed epistemic necessity. Experiment 2 followed up on Experiment 1 by testing the same two quantifiers with the same design but, this time, the modal *must* expressed deontic necessity. Example items are given in Figure 1.

**Experiment 1 – *every* vs. epistemic *must***




**Experiment 2 – *every* vs. deontic *must***

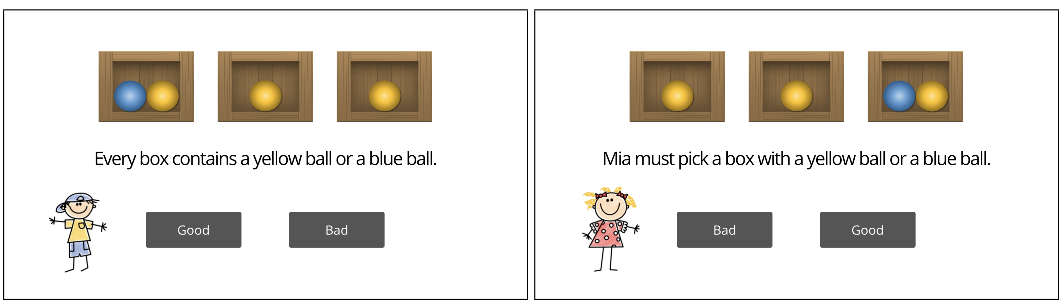



**Figure 1** Example items illustrating the items' layout in Exp.1 (top) and Exp.2 (bottom). These examples correspond to A-AB-A instances of the TARGET-1 conditions for the nominal (left) and modal (right) cases.

The target conditions in both experiments were constructed in a similar way to the critical conditions in Crnič et al.'s study. In the TARGET-1 conditions, the test sentences were paired with pictures that make their D-inferences true but their NU-inferences false; in the TARGET-2 conditions, they were paired with pictures that make both inference types false. We hypothesized that, if D-inferences derivationally depend on NU-inferences, no difference in participants' responses should be observed between both target conditions. On the other hand, if D-inferences can arise independently from NU-inferences, participants should reject the test sentences to a greater extent in the TARGET-2 than in the TARGET-1 conditions.

The study had two goals. The first was to test whether *must*-sentences differ from *every*-sentences in their ability to give rise to D-inferences independently of NU-inferences, as claimed by Crnič et al. and others. For these purposes, we compared responses in the TARGET-1 and TARGET-2 conditions within and across quantifier type. Based on Crnič et al.'s findings, we expected the *every*-sentences to be far less accepted in the TARGET-2 than in the TARGET-1 conditions. We were interested in testing whether a contrast of a similar magnitude is found for the *must*-sentences.

The second goal was to explore parallels with SI derivation in other paradigms by measuring Response Times (RTs), a measure that Crnič et al. did not look at. Among other things, we were interested in finding out whether participants take significantly longer to reject than accept the test sentences in the conditions where their D-inferences are false, i.e., in the TARGET-2 conditions. There is converging evidence that responses based on SIs take more time than responses based on the corresponding meaning without the SI (a.o., Noveck & Posada 2003, Bott & Noveck 2004, Breheny et al. 2006, Chevallier et al. 2008, Huang & Snedeker 2009, Bott et al. 2012, Tomlinson et al. 2013, Cremers & Chemla 2014, Chemla & Bott 2014a, van Tiel et al. 2019a, Van Tiel & Pankratz 2021, van Tiel et al. 2019b). This delay effect is one of the most replicated effects in judgement studies on SIs and it is often thought to be an important marker of this sort of meaning-strengthening operations. Thus, finding out that deriving D-inferences incurs a sizeable slowdown would support an implicature-based approach to these inferences.

## 2 Experiments

### 2.1 Participants

For each experiment, 100 native speakers of English were recruited online through Prolific (Palan & Schitter 2018) using the same pre-screen criteria (first language: English, nationality: UK/US, birth country: UK/US, approval rate: $\geq 90\%$). The recruitment was set up so that each participant could only take part in one of the experiments. Participants were paid £2.20. Average completion time was about 13 minutes. All participants gave written informed consent. Data were collected and stored in accordance with the provisions of Data Protection Act 2018. The study was approved by the Institution's Research Ethics Committee.

## 2.2 Materials and Design

Both experiments were based on the materials and method from Marty et al. (2023: Experiments 4–6) (see also Degano et al. 2023). Each item involved a sentence displayed just below a set of boxes horizontally aligned and right above the picture of one of two characters (see Figure 1). Sentences were constructed using the sentence frames in Table 1. The color adjectives, indicated by the [A] and [B] terms in Table 1, were picked at random from a list of four color terms – *yellow*, *blue*, *green* and *gray* – with replacement across items. The [name] term was the name of one of the two characters, *Mia* or *Sam*. The materials and cover story used in each experiment were adapted to the specific flavor of *must* that was targeted.

Experiment 1 built on Marty et al. (2023)'s implementation of Noveck (2001)'s mystery box paradigm to investigate epistemic modality. The test sentences were *every*-sentences like (1) and *must*-sentences like (8), in which *must* expressed epistemic necessity. Every item displayed a set of four boxes, each of which was made of three open boxes, containing one or two balls, and a covered box, marked with the symbol '?'. Participants were instructed that the characters could see what's inside the first three boxes – the *visible* boxes – but not what's inside the covered one – the *mystery* box. They were also instructed that the characters had been taught the rule that the mystery box always has the same contents as one of the three open boxes. For each item, participant had to decide whether the character's utterance was a good description of the relevant box(es) given the information available to them and the rule that they had learned.

Experiment 2 was built on Experiment 1 by adapting the materials and cover story from Experiment 1 to extend the investigation to deontic modality. The test sentences were *every*-sentences similar to those tested in Experiment 1 and novel *must*-sentences in which *must* expressed deontic necessity. The box display was similar to the one used in Experiment 1, except that it only involves three boxes, all open. In the instructions, participants were told that the two characters were playing games and that, depending on the game, one of them either had to describe what's inside the boxes or had to pick one of the boxes. The first game scenario was used for the NOMINAL cases and the second for the MODAL cases. Depending on the game scenario, participants had to decide whether the utterance was a good description of the box(es) or of the character's options.

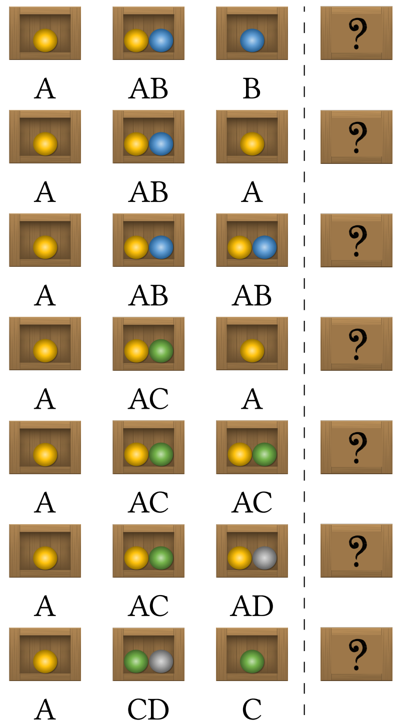
The rest of the design of Experiment 1 and 2 was identical in all respects. In both experiments, the contents of the open boxes were manipulated to create different picture types corresponding to the experimental conditions of the study. The test sentences were paired with four different picture types, which are described in Table 2: the colors of A-balls and B-balls depicted in the open boxes always matched the [A] and [B] color terms used in the sentences (e.g., *yellow* and *blue*) whereas

| | | |
|---|---|---|
| **Experiment 1 – *every* vs. epistemic *must*** | | |
| NOMINAL | | |
| | Test | Every visible box contains a [A] ball or a [B] ball. |
| | C1 | Every visible box contains a [A] ball. |
| | C2 | Every visible box contains a [A] ball and a [B] ball. |
| | C3 | No visible box contains a [A] ball. |
| | C4 | No visible box contains a [A] ball or a [B] ball. |
| MODAL | | |
| | Test | The mystery box must contain a [A] ball or a [B] ball. |
| | C1 | The mystery box must contain a [A] ball. |
| | C2 | The mystery box must contain a [A] ball and a [B] ball. |
| | C3 | The mystery box cannot contain a [A] ball. |
| | C4 | The mystery box cannot contain a [A] ball or a [B] ball. |
| **Experiment 2 – *every* vs. deontic *must*** | | |
| NOMINAL | | |
| | Test | Every box contains a [A] ball or a [B] ball. |
| | C1 | Every box contains a [A] ball. |
| | C2 | Every box contains a [A] ball and a [B] ball. |
| | C3 | No box contains a [A] ball. |
| | C4 | No box contains a [A] ball or a [B] ball. |
| MODAL | | |
| | Test | [Name] must pick a box with a [A] ball or a [B] ball. |
| | C1 | [Name] must pick a box with a [A] ball. |
| | C2 | [Name] must pick a box with a [A] ball and a [B] ball. |
| | C3 | [Name] cannot pick a box with a [A] ball. |
| | C4 | [Name] cannot pick a box with a [A] ball or a [B] ball. |

**Table 1** Schematic description of the sentences tested in Experiment 1 and 2, where [A] and [B] are placeholders for different colour adjectives and [name] a placeholder for a character's name; for a more concrete illustration, you may read [A] as *blue*, [B] as *yellow* and [name] as *Mia*.

the colors of the C-balls and D-balls were randomly chosen from our list of color terms by excluding the color(s) of the matching balls (e.g., *green* and *gray*). The position of the open boxes was randomly assigned; the mystery box displayed on the items of Experiment 1 always appeared in the rightmost position.

Target pictures were designed to make the NU-inferences of the test sentences always false, but their D-inferences either true or false. On the TARGET-1 pictures,



| Condition | | Example picture | | |
|---|---|---|---|---|
| TRUE | | A | AB | B |
| TARGET-1 | i. | A | AB | A |
| | ii. | A | AB | AB |
| TARGET-2 | i. | A | AC | A |
| | ii. | A | AC | AC |
| | iii. | A | AC | AD |
| FALSE | | A | CD | C |

**Table 2** Schematic description and illustration of the picture types paired with the test sentences in Exp.1 and Exp.2. Picture types are illustrated here using the following color assignment: A=yellow, B=blue, C=green and D=gray. *Note: the mystery box was only displayed on the items of Exp.1.*

each of the three open boxes contained an A-ball and at least one of them also contained a B-ball, making the NU-inferences of the test sentences false, but their D-inferences true. TARGET-2 pictures were obtained from the TARGET-1 pictures by replacing the B-ball(s) with balls of a non-matching color, thus making both inference types false. Different variants of the TARGET-1 and TARGET-2 pictures were constructed by varying the number of matching B-balls for the former and by varying both the number and color of non-matching balls for the latter. For the purposes of experimental design, variants of the TARGET-1 and TARGET-2 pictures were treated as sub-conditions of the TARGET-1 and TARGET-2 conditions.[4]

4 No contrast in responses was found between the variants of the TARGET-1 pictures, nor between those of the TARGET-2 pictures, allowing us to aggregate the responses to the TARGET-1 and the

FALSE and TRUE pictures were control pictures, each of which served a different experimental purpose. FALSE pictures were designed to provide a clear baseline for rejection for the test sentences. On these pictures, one of the open boxes contained an A-ball while the other two contained balls of a non-matching color, making the test sentences unambiguously false. TRUE pictures were designed to provide some relevant baseline for acceptance.[5] On these pictures, only some of the open boxes contained an A-ball and only some of them contained a B-ball, making the test sentence true irrespective of the inferences of interest.

In addition to the test sentences, there were four different types of control sentences: two positive sentences (C1 and C2) and two negative ones (C3 and C4), involving either one color adjective (C1 and C3) or two (C2 and C4). Each of these sentences were paired with pictures that made them either clearly true (GOOD) or clearly false (BAD), as described and illustrated in Table 3. These items were added to identify low-effort responses. In particular, we worried that some participants may perform the task superficially, simply by checking whether the colors mentioned in the sentence match those of the balls depicted on the pictures. We reasoned that, if a participant follows such a strategy, they should perform relatively poorly on the control items involving negative sentences.

Pairing the test and control sentences with the relevant picture types gave rise in each experiment to 4 test and 8 control conditions for each quantifier type. Each control condition was instantiated 3 times and each target condition 6 times, giving rise to 24 control and 24 test trials per quantifier type. Instances of the TARGET-1 and TARGET-2 conditions were evenly distributed across their respective sub-conditions. In each experiment, trials were blocked by quantifier type to facilitate participants' comprehension of the instructions and reduce the risk that participants' responses to one type of test trials be affected by the presentation of the other.

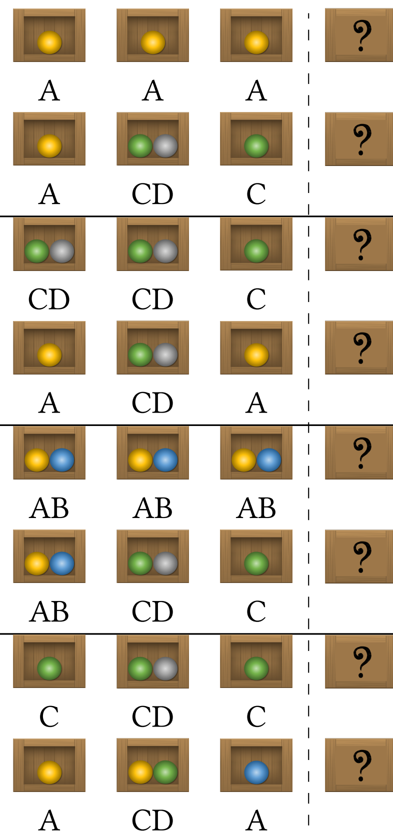
### 2.3 Procedure

The procedure was the same in both experiments, save the differences in instructions we described above. In each experiment, NOMINAL trials and MODAL trials were presented in two separate blocks, the order of which was counterbalanced between participants. Each block started with a set of instructions emphasizing some key aspects of the cover story. Next, participants completed a short practice

---

TARGET-2 trials across sub-conditions without loss of information (see Section 2.6).

5 TRUE pictures were designed to be as close as possible to the TARGET-1 pictures. On these pictures, one of the open boxes contained both an A-ball and a B-ball, just like on the TARGET-1 pictures. In principle, these pictures make the test sentences true unless an exclusivity inference is derived at embedded level. As reported in Section 2.7.3, there is no evidence in our data that this type of inference was ever derived by our participants.



| Sentence | Condition | Example picture |
| --- | --- | --- |
| C1 | GOOD | A A A |
| | BAD | A CD C |
| C2 | GOOD | CD CD C |
| | BAD | A CD A |
| C3 | GOOD | AB AB AB |
| | BAD | AB CD C |
| C4 | GOOD | C CD C |
| | BAD | A CD A |

**Table 3** Schematic description and illustration of the picture types paired with the control sentences in both experiments. Picture types are illustrated using the same color assignment as before. *Note: the mystery box was only displayed on the items of Exp.1.*

to make sure that they understood the instructions properly. The practice included one instance of each control condition associated with the quantifier type tested in the block, hence 8 trials. During this phase, participants received feedback on the accuracy of their responses. They were prevented from moving to the test phase until they correctly answered all practice trials. After the practice, each block continued with 48 experimental trials presented in random order. Participants reported their responses by clicking one of two response buttons labelled 'Good' and 'Bad', respectively. The position of the labels was counterbalanced across participants. Responses and RTs were recorded on each trial.

## 2.4 Data availability

Materials along with the code files for result analysis and raw data can be found on the OSF Platform at https://osf.io/wds4n/.

## 2.5 Software

Data treatment and analysis were carried out in the R statistical environment (R Core Team 2023) using the Hmisc (Harrell 2023), Rmisc (Hope 2022), lme4 (Bates et al. 2015), car (Fox & Weisberg 2019) outliers (Komsta 2022) and emmeans (Lenth 2023) packages for the R statistics program.

## 2.6 Data preparation

8 participants in Experiment 1 and 6 participants in Experiment 2 were excluded because their performance on the control items was below the pre-established threshold of 80% accuracy. The performance of the remaining subjects was uniformly high both in the BAD and the GOOD conditions.

Responses to the TARGET-1 and TARGET-2 trials in both experiments were inspected to check for potential discrepancies among the variants of the TARGET-1 and TARGET-2 pictures (see Table 2). For each target condition, we fitted a generalised linear mixed-effect (GLMER) model with a logit link function, predicting responses from the fixed effect of picture sub-type (dummy coded). All models included by-participant random variance for the intercept, the slope, and their correlation, and by-item random variance for the intercept. Each model was compared to a null model missing the fixed effect. None of the models was significantly different from the null model, meaning that picture sub-type had no reliable effect on responses in the target conditions. Responses to the TARGET-1 and TARGET-2 trials were aggregated across sub-conditions for the main analyses.

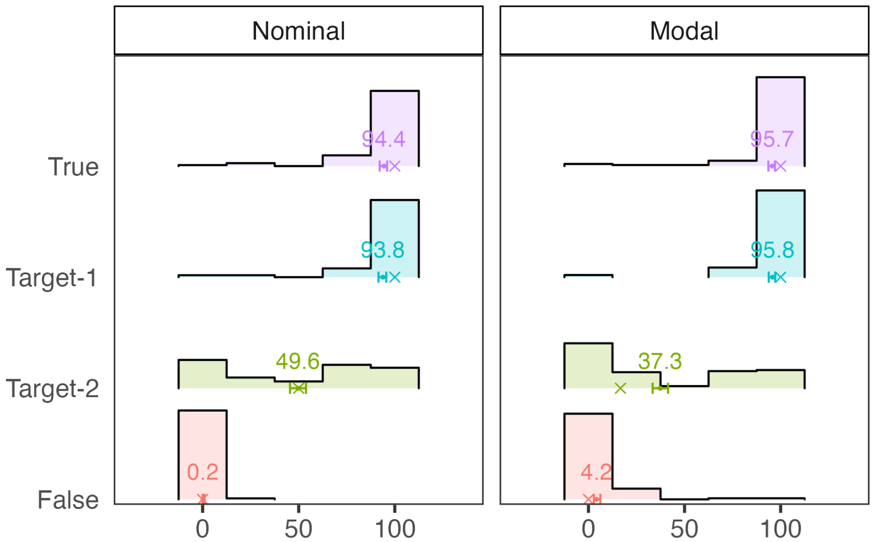
Data treatment and analyses for responses and RTs are described in the relevant sections below. We refer the reader to the code files for the full outputs of the statistical models that we summarize in the following.

## 2.7 Responses

### 2.7.1 Treatment

No further data treatment was applied for the analysis of responses.

**Experiment 1 – *every* vs. epistemic *must***




**Experiment 2 – *every* vs. deontic *must***

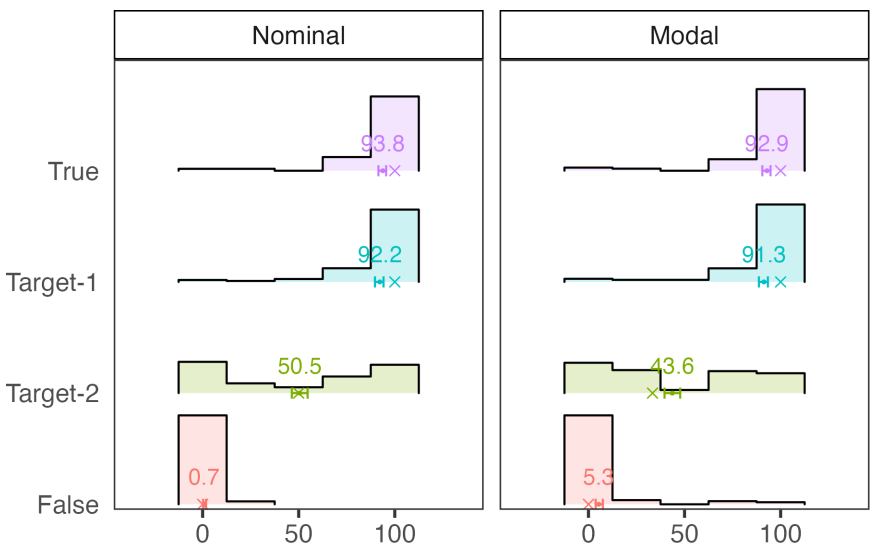



**Figure 2** Mean acceptance rate by quantifier type and condition in Exp.1 (top) and Exp.2 (bottom). The distribution of by-participant mean rates is visualised by a histogram, the grand mean by a thick bar with its value on top and the 95% CI around it, and the median by a cross.

### 2.7.2 Analyses

Responses to the test trials were analysed (i) by carrying out pairwise comparisons between each target condition and all other (sentence-related) conditions and (ii) by testing the interaction effect between quantifier type and condition on the responses to the TARGET-1 and TARGET-2 trials. For (i), we fitted GLMER models (logit link function), predicting responses from the fixed effect of condition (dummy coded). Each model included random intercepts for participants, items and block orders, which was the maximal random effect structure allowing all the models to converge. For (ii), we fitted participants' responses to the TARGET-1 and TARGET-2 trials into a GLMER model (logit link function), predicting responses from quantifier type (dummy coded), condition (dummy coded) and the interaction between the two. The maximal converging model included random effects for the intercept and the slope with their correlation, grouped by each of participant, item, and block.

The $\chi^2$ and p-values that we report were obtained by performing likelihood ratio tests in which the deviance of the models containing the main or interaction effect of interest was compared to another model without the relevant effect, but with the same random effect structure. For the pairwise comparisons, the Bonferroni correction method for multiple testing was used for interpreting the p-values. Concretely, because 10 comparisons were carried out on the data from each experiment, only p-values below 0.005 were treated as significant.

### 2.7.3 Results

Figure 2 shows the mean acceptance rate (i.e., proportion of 'Good' responses) by quantifier type and condition in both experiments. Overall, the response patterns for MODAL sentences were similar to those for NOMINAL sentences across-the-board. In both experiments, the acceptance rates for these sentences were uniformly high in the TRUE and TARGET-1 conditions (all Ms>92%), with no significant difference between the two (all $\chi^2_1$s$<$ 2.06, *ns*). By contrast, the acceptance rates for the TARGET-2 conditions were somewhat intermediate (37%<Ms<51%), between the high(est) rates observed in the TRUE conditions (all $\chi^2_1$s$>$ 150, all $p$s$<$ .001) and the low(est) ones observed in the FALSE conditions (all $\chi^2_1$s$>$ 65, all $p$s$<$ .001), with the TARGET-2 conditions yielding significantly lower rates than the TARGET-1 conditions (all $\chi^2_1$s$>$ 129, all $p$s$<$ .001). These results indicate that the participants in our experiments did not derive the NU-inferences associated with the test sentences whereas they derived the corresponding D-inferences to a large extent.

Responses to NOMINAL and MODAL sentences in the TARGET-1 and TARGET-2 conditions were further compared by testing the interaction between quantifier type and condition. A significant interaction was found in Exp.1 ($\chi^2_1 = 9.64$,

$p < .005$), showing that the acceptability contrast between the TARGET-1 and TARGET-2 conditions was greater for the MODAL than for the NOMINAL sentences; this interaction was not significant in Exp.2 ($\chi^2_1 = 1.69$, $p = .19$). These results replicate the main results from Crnič et al. 2015 and establish that the key contrasts previously observed for the NOMINAL quantifier *every* reproduce with the MODAL quantifier *must*, whether this modal receives an epistemic or a deontic reading.

## 2.8 Response times

### 2.8.1 Treatment

Further data treatment was applied prior to analysing RTs. First, we removed all trials associated with incorrect responses in the TRUE and FALSE conditions and with 'Bad' responses in the TARGET-1 conditions as the response analysis showed no evidence that participants accessed the NU-reading associated with that response type in the TARGET-1 conditions. For the TARGET-2 conditions, both 'Good' (accept) and 'Bad' (reject) responses were kept, in line with the response analysis. About 3% of the trials in Exp.1 and 4.5% of the trials in Exp.2 were removed as a consequence.

Next, we looked at the distribution of RTs in Exp.1 and Exp.2 to check for extreme data points. The distributions were positively skewed (skewness$>$ 8), with a small amount of very high RTs in both datasets. To exclude these extreme values, we used the interquartile range (IQR) criterion and removed all observations above $q_{0.75} + 1.5\times$ IQR and below $q_{0.25} - 1.5\times$IQR. In effect, this procedure removed all observations above 6654 ms in Exp.1 and above 6121 ms in Exp.2. In total, 296 out of 4280 trials in Exp.1 and 248 out of 4310 trials in Exp.2 were removed that way (about 7% and 5.7% of the datasets, respectively). Grubb's tests were performed on the resulting datasets: neither the highest value, nor the lowest value in the sets was found to be an outlier; the distributions were still slightly skewed to the right (skewness$<$ 0.75).[6] RTs were log-transformed prior to statistical analyses.

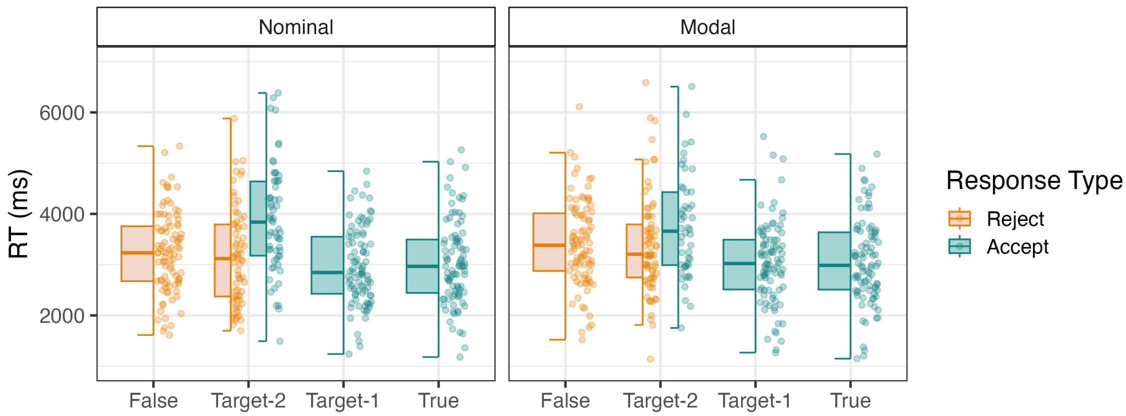
### 2.8.2 Analyses

RTs were analyzed using GLMER models with an inverse Gaussian link function and with the maximal random effect structure justified by the design and supported by the data. For each test sentence in each experiment, we tested (i) the interaction effect between response type (accept vs. reject) and trial type (TARGET-2 vs. control), (ii) the effect of response type on the TARGET-2 and control trials and (iii) the effect

6 Other methods to identify and exclude outliers were considered: a reasonable upper cutoff (e.g., 10000 ms), another quantile-based criterion (e.g., the outer 5% of the distribution), mean$\pm$2SD. These alternative methods were found to be less suitable for the present data in failing to reomve some very extreme values in the upper part of the distribution.

of condition on the accept and reject responses. Every model included random intercepts for participants, items and block orders, and a random slope for all fixed effects in the model with their correlations grouped by participant. All fixed effects were dummy coded. The $\chi^2$ and p-values were obtained through model comparisons following the same procedure as the one used in the response analysis.

**Experiment 1 – *every* vs. epistemic *must***




**Experiment 2 – *every* vs. deontic *must***

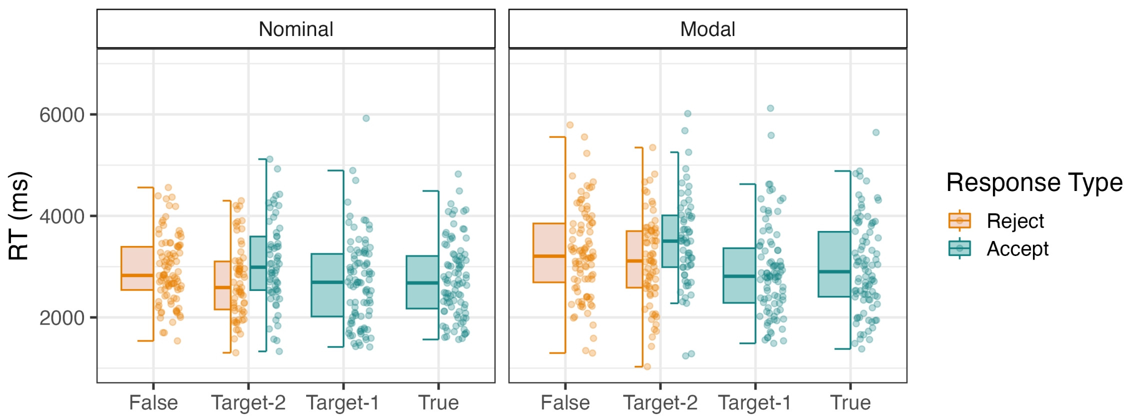



**Figure 3** Distribution of the by-participant mean RTs (in ms) by quantifier type, condition and response type in Exp.1 (top) and Exp.2 (bottom). RTs were analysed by considering correct 'accept' and 'reject' responses in the control conditions, 'accept' responses in the TARGET-1 conditions, and both 'accept' and 'reject' responses in the TARGET-2 conditions.

### 2.8.3 Results

Figure 3 shows the distribution of the by-participant mean RTs by quantifier type, condition and response type. In both experiments, a significant interaction between response type and trial type was found for the NOMINAL sentences (all $\chi^2_1$s$>5.96$, all $p$s$<.05$) and the MODAL sentences (all $\chi^2_1$s$>6.04$, all $p$s$<.05$). For the NOMINAL sentences, the interpretation of this interaction is the same in both experiments: participants took significantly longer to accept than reject these sentences in the TARGET-2 conditions (all $\chi^2_1$s$>5.18$, all $p$s$<.05$) and significantly longer to accept them in the TARGET-2 than in the TRUE conditions (all $\chi^2_1$s$>6.52$, all $p$s$<.05$); by contrast, participants were equally fast at accepting and rejecting these same sentences in the control conditions (all $\chi^2_1$s$<2.29$, *ns*) and equally fast at rejecting them in the TARGET-2 and FALSE conditions (all $\chi^2_1$s$<0.98$, *ns*).

The results for the MODAL sentences were overall similar, with some minor differences across experiments and across quantifiers. First, participants took longer to accept than reject these sentences in Exp.2 ($\chi^2_1 = 5.18$, $p < .05$), but not in Exp.1 ($\chi^2_1 = 2.09$, $p = 0.19$). Second, in both experiments, participants were faster to accept than reject them in the control conditions (all $\chi^2_1$s$>3.93$, all $p$s$<.05$). The remaining comparisons with the controls yielded the same results as before. That is, participants were slower to accept the MODAL sentences in the TARGET-2 than in the TRUE conditions (all $\chi^2_1$s$>9.97$, all $p$s$<.01$) whereas they were equally fast at rejecting them in the TARGET-2 and FALSE conditions (all $\chi^2_1$s$<0.24$, *ns*).

In sum, 'accept' responses to the NOMINAL and MODAL trials in the TARGET-2 conditions were delayed compared to any other response type in any other condition in both experiments. These results suggest that responding on the basis of a meaning of the test sentences without their D-inferences is linked with processing effort. We discuss the consequences of these findings in the next section.

# 3 Discussion

## 3.1 Main results

This study had two goals: (i) test whether the universal modal *must* differs from the nominal quantifier *every* in its ability to give rise to D-inferences independently of NU-inferences, and (ii) explore the time course of D-inferences.

In relation to (i), we found that MODAL sentences patterned with NOMINAL sentences in all relevant aspects. First, both sentence types were far less accepted in their TARGET-2 than in their TARGET-1 conditions, as expected if both sentence types can give rise to D-inferences independently of NU-inferences. In fact, as previously mentioned, the TARGET-1 conditions systematically yielded near-ceiling acceptance rates, comparable to those found in the TRUE conditions. Hence, there is no evidence in our data that participants ever derived the NU-inferences associated with the test sentences.[7] By contrast, the TARGET-2 conditions systematically yielded intermediate acceptance rates, showing that participants often derived their D-inferences. Second, the comparisons that we carried out between target conditions across quantifier types suggest that MODAL and NOMINAL sentences did not remarkably differ in their propensity to give rise to D-inferences. Finally, the fact that these results were found in both experiments suggests that the findings for MODAL sentences generalize to different flavors of *must*.

While the results for *every* replicate previous finding from Crnič et al. 2015, the results for *must* are novel and directly go against the claim that, for disjunction under universal modals, D-inferences are always observed in tandem with NU-inferences (Booth 2022, Crnič et al. 2015, Fusco 2015, Sayre-McCord 1986). Our results, therefore, present a challenge for accounts of D-inferences which, for modals like *must*, cannot derive the former without also deriving the latter.[8]

In relation to (ii), we found that acceptance of the NOMINAL and MODAL sentences in the TARGET-2 conditions resulted in a slowdown. Specifically, the RT data from both experiments show that 'accept' responses to the test sentences in these conditions were generally delayed compared to any other response type in any other condition. In the context of our experiments, where NU-inferences had no detectable effect on participants' responses, these delay effects can be taken to inform us directly about the processing of D-inferences and, specifically, to show that *not deriving* these inferences is cognitively demanding. In that regard, it is worth noting that the delay effects reported here are in the opposite direction to those commonly found for regular SIs in similar judgment tasks. In the case of scalar sentences, when such effects are present, they are found for 'reject' responses, the response type associated with the strong reading of these sentences, i.e., the one with the SI. In the present case, however, these effects are found for 'accept' responses, the response type associated with the weak reading of the relevant sentences, i.e., the one without D-inferences.

From a psycholinguistic standpoint, these results support the idea that the

7 This result, of course, is compatible with NU-inferences being derived in other cases, as suggested in the literature. Most notably, Crnič et al. 2015 (p.32) observe that a sentence like *You are required to wear sneakers or running shorts* feels misleading in a context in which it is required to wear sneakers in the gym (while running shorts are optional). As the authors note, this observation is expected if this sentence gives rise to NU-inferences; it would be unexpected, however, if only D-inferences were present. We take this observation to suggest that a fully-fledged theory of D-inferences should ultimately be able to explain when DISTRIBUTIVE-only readings are possible and when they are not.

8 All such accounts can also derive, for MODAL and NOMINAL sentences, a reading without either inference. On this reading, however, these sentences should be judged as true and thus accepted in both our target conditions. Hence, this reading cannot account for the contrasts observed between the TARGET-1 and TARGET-2 conditions.

reading with D-inferences is, in some psychological sense, more primitive than the one without. In principle, this idea remains compatible with D-inferences being SIs, if we assume that (i) these SIs are computed by default, and (ii) cancelling them comes at an extra processing cost. On this view, the delay effects that we found would reflect the fact that speakers first had to cancel these SIs before they could access the literal reading of the relevant sentences. While these assumptions seem fairly reasonable, the challenge for this proposal remains to explain why D-inferences would differ from other types of SIs in being computed by default.

Taken together, these findings establish two main challenges for any account of D-inferences. First, the response data suggest that we need a theory of D-inferences that can derive them independently of NU-inferences across nominal and modal quantifiers. Second, the RT data challenge the view that D-inferences should be treated as regular SIs, as proposed in the traditional approach.

### 3.2 Two recent approaches

Several accounts have been proposed for deriving the D-inferences of disjunction independently from their NU-inferences with nominal quantifiers (Aloni 2022, Bar-Lev & Fox 2020, 2023, Crnič et al. 2015, Minor 2022, Sudo to appear). In the following, we focus on two such accounts which can extend these good results to modal quantifiers: the implicature approach by Bar-Lev & Fox (2023), based on recursive exhaustification, and the non-implicature approach by Aloni (2022), based on 'neglect-zero'. While both approaches can take up the challenge raised by our response data, we will see that they lead to divergent expectations regarding the time course of D-inferences.

#### 3.2.1 An implicature account

Bar-Lev & Fox (2023) offer a novel implicature account of D-inferences which can derive these inferences without also deriving NU-inferences.[9] The novelty of Bar-Lev & Fox's proposal has to do with the set of alternatives that speakers are assumed to consider when exhaustifying the meaning of a sentence like (1). Under their account, the set of alternatives associated with such sentences includes all those formal alternatives where *every* has been replaced with its existential counterpart *some*, as shown in (10).

9 Crnič et al. (2015) had previously proposed a related implicature account deriving D-inferences independently from NU-inferences. See Bar-Lev & Fox (2023) for criticisms of this previous account.

$$
(10)\quad \left\{\begin{array}{ll}
\text{Every box contains a yellow ball or a blue ball} & \forall x(\mathsf{A}x \vee \mathsf{B}x) \\
\text{Every box contains a yellow ball} & \forall x\mathsf{A}x \\
\text{Every box contains a blue ball} & \forall x\mathsf{B}x \\
\text{Every box contains a yellow ball and a blue ball} & \forall x(\mathsf{A}x \wedge \mathsf{B}x) \\
\text{Some box contains a yellow ball or a blue ball} & \exists x(\mathsf{A}x \vee \mathsf{B}x) \\
\text{Some box contains a yellow ball} & \exists x\mathsf{A}x \\
\text{Some box contains a blue ball} & \exists x\mathsf{B}x \\
\text{Some box contains a yellow ball and a blue ball} & \exists x(\mathsf{A}x \wedge \mathsf{B}x)
\end{array}\right\}
$$

As demonstrated in (11), it is possible to derive D-inferences by recursively exhaustifying the meaning of (1) over the alternatives above. On the first application of EXH, only the conjunctive alternatives $\forall x(\mathsf{A}x \wedge \mathsf{B}x)$ and $\exists x(\mathsf{A}x \wedge \mathsf{B}x)$ can be excluded. On the subsequent application of EXH, D-inferences are generated thanks to the presence of the individual existential alternatives.

$$
\begin{array}{ll}
(11) & \textsc{exh}(\textsc{exh}\ \forall x(Ax \vee Bx)) \\
& \Leftrightarrow \forall x(Ax \vee Bx) \wedge \neg\textsc{exh}(\forall xAx) \wedge \neg\textsc{exh}(\forall xBx) \wedge \neg\exists x(\mathsf{A}x \wedge \mathsf{B}x) \\
& \Leftrightarrow \forall x(Ax \vee Bx) \wedge \neg(\forall xAx \wedge \neg\exists xBx) \wedge \neg(\forall xBx \wedge \neg\exists xAx) \wedge \neg\exists x(\mathsf{A}x \wedge \mathsf{B}x) \\
& \Leftrightarrow \forall x(Ax \vee Bx) \wedge \exists xAx \wedge \exists xBx \wedge \neg\exists x(\mathsf{A}x \wedge \mathsf{B}x)
\end{array}
$$

Note that, given the combination of the prejacent with the D-inferences $\exists xAx \wedge \exists xBx$ and the strong exclusivity inference $\neg\exists x(\mathsf{A}x \wedge \mathsf{B}x)$, the derivation above still entails NU-inferences. Bar-Lev & Fox observe, however, that a reading with D-inferences only can be derived by pruning the conjunctive alternatives in (10), hence removing exclusivity inferences from the picture.[10] In this case, the pruned set of alternatives on which EXH operates is as follows:

$$
(12)\quad \left\{\begin{array}{ll}
\text{Every box contains a yellow ball or a blue ball} & \forall x(\mathsf{A}x \vee \mathsf{B}x) \\
\text{Every box contains a yellow ball} & \forall x\mathsf{A}x \\
\text{Every box contains a blue ball} & \forall x\mathsf{B}x \\
\text{Some box contains a yellow ball or a blue ball} & \exists x(\mathsf{A}x \vee \mathsf{B}x) \\
\text{Some box contains a yellow ball} & \exists x\mathsf{A}x \\
\text{Some box contains a blue ball} & \exists x\mathsf{B}x
\end{array}\right\}
$$

The novel outcome is shown in (13). On the first application of EXH, none of the alternatives above is excluded; on the second application, D-inferences are derived just as before. Crucially, the resulting reading for (1) is compatible this time with the NU-inferences being false.

10 Strictly speaking, weaker exclusivity inferences of the form $\neg\forall x(Ax \wedge Bx)$ could still be derived by keeping the universal, conjunctive alternative $\forall x(Ax \wedge Bx)$. As these inferences are inconsequential for our point, we set this alternative aside for simplicity.

(13) $\textsc{exh}(\textsc{exh}\ \forall x(Ax \vee Bx))$
$\Leftrightarrow \forall x(Ax \vee Bx) \wedge \neg\textsc{exh}(\forall x Ax) \wedge \neg\textsc{exh}(\forall x Bx)$
$\Leftrightarrow \forall x(Ax \vee Bx) \wedge \neg(\forall x Ax \wedge \neg\exists x Bx) \wedge \neg(\forall x Bx \wedge \neg\exists x Ax)$
$\Leftrightarrow \forall x(Ax \vee Bx) \wedge \exists x Ax \wedge \exists x Bx$

In sum, this account allows D-inferences to be generated independently of NU-inferences. In addition, it can be extended to universal modals in a straightforward way by assuming, in the same way as above, that the set of alternatives in such cases includes those alternatives derived by replacing the quantifier in question (e.g., *must*) with an existential one (e.g., *might*). This account is, therefore, fully compatible with our response data.

Our RT data, however, is more challenging for this account. Specifically, on this account, one would expect the processing profile of D-inferences to be similar to that of regular SIs. In particular, one would expect the DISTRIBUTIVE-only reading of sentences like (1) to be harder to process than their literal reading. The problem is that the delay effects we found go the other way around, showing that accessing the literal reading of these sentences is in fact more effortful than accessing their DISTRIBUTIVE-only reading. As previously discussed, these 'reverse' delay effects are unexpected if D-inferences are to be treated on a par with regular SIs.

The challenge here is reminiscent of the one raised for implicature accounts of free choice (FC) inferences. Experimental studies investigating these inferences have found that they are readily derived by speakers without causing any remarkable slowdown (Chemla & Bott 2014b, Van Tiel & Schaeken 2017; see Marty et al. 2020 for discussion). (Though we note that our results do not just show that responses based on D-inferences are not slower than those based on the literal reading, they show that these response are actually faster). Thus far, the main response to the challenge raised by FC has been to assume that the difference with regular SIs stems from the type of alternatives involved in their derivation. In a nutshell, the derivation of regular SIs involves alternatives constructed by lexical substitution (e.g., replacing *some* with *all*) while the derivation of FC involves alternatives constructed by simplification (e.g., simplifying *A or B* as *A*). Based on these considerations, it has been hypothesized that appealing to lexical substitution when building alternatives is what slows down the processing of a SI, hence the absence of delay effects for FC inferences (Pagliarini et al. 2018, Van Tiel & Schaeken 2017, Tieu et al. 2016, Singh et al. 2016, Chemla & Bott 2014b, Barner et al. 2011).

This line of explanation does not offer a solution to the challenge raised by D-inferences for the approach outlined above. The reason is that the critical alternatives involved in the derivation of these inferences are constructed by lexical substitution (e.g., by replacing *every* with *some*), as in the case of regular SIs. Therefore, we would still expect the processing profile of D-inferences to be similar to that of regular SIs, contrary to what we found.[11]

### 3.2.2 A non-implicature account

Aloni (2022) has recently put forward an account of D-inferences and related inferences in which D-inferences result from a pragmatic tendency referred to as 'neglect-zero'. The main idea behind Aloni's account is that, when interpreting a sentence, language users construct structures that represent reality and, in doing so, they tend to disregard structures that vacuously satisfy the sentence due to an empty configuration, also referred to as zero-models. These structures are built upon the notion of a state – a set of possible worlds – which reflects the speaker's information state and upon which formulas are interpreted. Aloni (2022) formalizes this notion in a bilateral state-based modal logic (henceforth, BSML).

For what is most relevant to us, BSML employs a split notion of disjunction on which disjunction is supported in a state if that state can be split into two substates, each of which supports one of the disjuncts. According to this definition, a disjunction of the form $A \vee B$ is thus supported by a state even if one or both of the disjuncts are vacuously supported by the empty state. Such cases, however, are ruled out by the neglect-zero option of interpretation. Aloni (2022) implements the neglect-zero effect by means of an enrichment function, $(\cdot)^+$, which excludes the empty state as a possible verifier. Focusing here on the case of disjunction, an enriched disjunction of the form $(A \vee B)^+$ is supported in a state when the state can be split into two *non-empty* substates, each supporting one of the disjuncts.

As discussed in Aloni (2022), the neglect-zero option readily accounts for D-inferences when disjunction is embedded under a universal modal since $[\Box(A \vee B)]^+$ entails $\Diamond A$ and $\Diamond B$, regardless of the flavor of the modal involved, as exemplified in (14). Importantly, since NU-inferences are not derived by the neglect-zero option, it is expected that D-inferences can be observed on their own, consistent with our response data. We also note that a first-order extensions of Aloni (2022)'s system, such as Aloni & van Ormondt (2023), can capture the D-inferences arising from disjunction under nominal quantifiers in the exact same way.

(14) a. $[\Box(A \vee B)]^+ \models \Diamond A \wedge \Diamond B$
b. $[\Box(A \vee B)]^+ \not\models \neg\Box A \wedge \neg\Box B$

Finally, this approach is compatible with our RT data: if D-inferences are derived

11 On the implicature account by Crnič et al. (2015), D-inferences are based on alternatives derived by simplification. When supplemented with the alternative-based hypothesis above, this account predicts D-inferences to be fast to process, consistent with our data. It does not predict, however, the slowdown for 'accept' responses in our data; see Bar-Lev & Fox 2020 for a critical discussion of this account and Marty et al. 2020 for other challenges for the alternative-based hypothesis.

as neglect-zero effects, there is no reason *a priori* to expect their processing profile to be similar to that of SIs. In fact, Aloni (2022) suggests that the neglect-zero interpretation is the default option, which leads to the expectation that responses based on neglect-zero should be faster than those where zero-models are left in and taken into consideration. This proposal correctly predicts the direction of the delay effects that we found.

## 4 Conclusion

The results of this study are twofold. First, the response results extend earlier experimental findings from Crnič et al. (2015) in showing that D-inferences can be observed without the corresponding NU-inferences across universal quantifiers. Second, the RT results show that, in situations where D-inferences are false, responses ignoring these inferences are slower than responses informed by them. Based on these findings, we argued that a theory of D-inferences should satisfy (at least) two desiderata: it should allow D-inferences to be derived independently from NU-inferences, irrespective of the quantifier involved, and explain the *raison d'être* of the the response delay effects that we unveiled.

We discussed two promising directions to take up on these new challenges, the recent implicature approach by Bar-Lev & Fox (2023) and the neglect-zero approach by Aloni (2022). The first approach can account for the independent derivation of D-inferences, both with nominal quantifiers and modals, but falls short of an explanation regarding the response delay effects. The second approach can also account for our response data by deriving D-inferences as neglect-zero effects and it offers some insights as to why literal responses appear to be linked with processing effort: if neglect-zero is a default option of interpretation, interpretations overriding neglect-zero should be harder to process than those that do not.

## References


Aloni, Maria. 2022. Logic and conversation: the case of free choice. *Semantics and Pragmatics* 15. 5.

Aloni, Maria & Peter van Ormondt. 2023. Modified numerals and split disjunction: the first-order case. *Journal of Logic, Language and Information* 1–29.

Bar-Lev, Moshe & Danny Fox. 2020. Free choice, simplification, and Innocent Inclusion. *Natural Language Semantics* 28. 175–223. http://dx.doi.org/10.1007/s11050-020-09162-y.

Bar-Lev, Moshe E & Danny Fox. 2023. On fatal competition and the nature of distributive inferences. Manuscript TAU-MIT. https://lingbuzz.net/lingbuzz/007055.

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. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Booth, Richard Jefferson. 2022. Independent alternatives: Ross's puzzle and free choice. *Philosophical Studies* 179(4). 1241–1273.

Bott, Lewis, Todd M Bailey & Daniel Grodner. 2012. Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language* 66(1). 123–142.

Bott, Lewis & Ira A Noveck. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language* 51(3). 437–457.

Breheny, Richard, Napoleon Katsos & John Williams. 2006. Are generalised scalar implicatures generated by default? an on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100(3). 434–463.

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

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

Chevallier, Coralie, Ira A Noveck, Tatjana Nazir, Lewis Bott, Valentina Lanzetti & Dan Sperber. 2008. Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology* 61(11). 1741–1760.

Cremers, Alexandre & Emmanuel Chemla. 2014. Direct and indirect scalar implicatures share the same processing signature. In *Pragmatics, semantics and the case of scalar implicatures*, 201–227. Springer.

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

Degano, Marco, Paul Marty, Sonia Ramotowska, Maria Aloni, Richard Breheny, Jacopo Romoli & Yasutada Sudo. 2023. The ups and downs of ignorance. Under review. https://ling.auf.net/lingbuzz/007389.

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, John & Sanford Weisberg. 2019. *An R companion to applied regression*. Thousand Oaks CA: Sage 3rd edn. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Fusco, Melissa. 2015. Deontic modality and the semantics of choice. *Philosophers' imprint* 15.

Harrell, Frank. 2023. *Package 'Hmisc'*. https://CRAN.R-project.org/package=Hmisc. R package version 5.1-0.

Hope, Ryan. 2022. *Package 'Rmisc'*. https://CRAN.R-project.org/package=Rmisc. R package version 1.5.1.

Huang, Yi Ting & Jesse Snedeker. 2009. Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology* 58(3). 376–415.

Komsta, Lukasz. 2022. *Outliers: Tests for Outliers*. https://CRAN.R-project.org/package=outliers. R package version 0.15.

Lenth, Russell. 2023. *'emmeans': Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans. R package version 1.8.7.

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

Marty, Paul, Jacopo Romoli, Yasutada Sudo & Richard Breheny. 2023. What makes an inference robust? *Journal of Semantics* To appear.

Minor, Serge. 2022. Implicatures and events: A study of distributive inferences. *under review* .

Noveck, Ira A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78(2). 165–188.

Noveck, Ira A & Andres Posada. 2003. Characterizing the time course of an implicature: An evoked potentials study. *Brain and language* 85(2). 203–210.

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.

Palan, Stefan & Christian Schitter. 2018. Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17. 22–27.

R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

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

Sayre-McCord, Geoffrey. 1986. Deontic logic and the priority of moral theory. *Noûs* 179–197.

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.

Sudo, Yasutada. to appear. Scalar implicatures with discourse referents: A case study on plurality inferences. *Linguistics and Philosophy* .

van Tiel, Bob, Paul Marty, Elizabeth Pankratz & Chao Sun. 2019a. Scalar inferences and cognitive load. In *Proceedings of sinn und bedeutung*, vol. 23 2, 427–442.

van Tiel, Bob, Elizabeth Pankratz & Chao Sun. 2019b. 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.

Tomlinson, John M, Todd M Bailey & Lewis Bott. 2013. Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of memory and language* 69(1). 18–35.

Van Tiel, Bob & Elizabeth Pankratz. 2021. Adjectival polarity and the processing of scalar inferences. *Glossa: a journal of general linguistics* 6(1).

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