Semantics & Pragmatics Volume 19, Article 12, 2026
https://doi.org/10.3765/sp.19.12

This is an EARLY ACCESS version of

Marty, Paul, Sonia Ramotowska, Richard Breheny, Jacopo Romoli & Yasutada Sudo. 2026. On the source of distributive inferences. *Semantics and Pragmatics* 19(12). https://doi.org/10.3765/sp.19.12.

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

EARLY ACCESS

On the source of distributive inferences*

Paul Marty
Universidade de Lisboa (CLUL)

Sonia Ramotowska
Institut Jean Nicod, Département d'Études Cognitives (ENS-PSL, EHESS, CNRS)

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 under universal quantifiers. It has been claimed that DISTRIBUTIVE inferences can be derived independently from NEGATED UNIVERSAL inferences with *nominal*, but not with *modal* quantifiers (Sayre-McCord 1986, Crnič et al. 2015, Fusco 2015, Booth 2022). 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 by testing the same two quantifiers, but this time with *must* expressing deontic necessity. The results from both experiments support the view that DISTRIBUTIVE inferences can arise independently of NEGATED UNIVERSAL inferences with both types of operator. While the findings for *every* essentially reproduce those of Crnič et al. 2015, the findings for *must* are new and go against the aforementioned claim. As such, we take these findings to pose a challenge for accounts of DISTRIBUTIVE inferences that cannot derive them independently of NEGATED UNIVERSAL inferences across universal quantifiers. We outline two recent accounts of DISTRIBUTIVE inferences

* For helpful comments and discussion, 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, the audiences at the 23rd Amsterdam Colloquium and the workshop 'Free Choice Inferences: theoretical and experimental approaches' (May 2024), the anonymous reviewers and the editor. PM acknowledges support from the Fundação para a Ciência e a Tecnologia through the grants 2023.06820.CEECIND/CP2891/CT0008 (individual funding) and UID/214/2025 (institutional funding awarded to the Centro de Linguística da Universidade de Lisboa). SR was supported by the NWO project 'Nothing is Logical' (Nihil, 406.21.CTW.023) and by the French National Research Agency through the projects 'Communicative efficiency, cognitive constraints and lexical meaning' (ANR-23-CE28-0016) and 'FrontCog' (ANR-17-EURE-0017). JR was supported by the Deutsche Forschungsgemeinschaft (RO6384/1-1). The early stages of this project were supported by the Leverhulme Trust (RPG-2018-425).

that can meet this challenge.

Keywords: distributive inferences, disjunction, universal quantifiers, modals, scalar implicatures, neglect-zero, pragmatic reasoning

1 Introduction

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). These 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. Since Sauerland (2004), these inferences have been analysed as a by-product of the Scalar Implicatures (SIs) derived through negating the universally quantified alternatives to (1) in (2), both of which are structurally simpler, stronger and innocently excludable (see also Fox 2007). 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. $\forall x\mathsf{B}x$

Thus, on this account, the D-inferences associated with a sentence like (1) derivationally depend on the 'Negated Universal' inferences (henceforth, NU-inferences) in (3a) and (3b): for the former to obtain, the latter must be derived. This, in turn, predicts that D-inferences cannot arise in the absence of NU-inferences. Hereafter, we will refer to this account as the NU-based approach to D-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 report experimental data that challenges this account. In their study, Crnič et al. found that *every*-sentences like (1) are readily accepted when their D-inferences are true while one of their NU-inferences is false, e.g., in situations where all boxes contain a yellow ball and some of them also contain a blue one.[1] By

1 The stimuli used in Crnič et al. 2015 involve boxes and letters, rather than boxes and balls. The examples in (1) and (4) are used for consistency with the stimuli in our own experiments.

comparison, they found that these same sentences are significantly less accepted when, in addition to one of their NU-inferences, one of their D-inferences is also false, e.g., in situations where all boxes contain a yellow ball but none of them also contain a blue one. These findings partly disconfirm the prediction of the NU-based approach to D-inferences in showing that, when disjunction is embedded under a *nominal* quantifier, D-inferences can arise independently of NU-inferences.

Crnič et al. submit, however, that the relevant prediction is borne out when disjunction is embedded under a *modal* quantifier. 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 (4), where disjunction appears in the scope of *must*. The corresponding D-inferences are given in (4a) and (4b). 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.

(4) The mystery box must contain a yellow ball or a blue ball. $\Box(A \vee B)$

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

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

Based on their own judgments, Crnič et al. claim that, with universal modals, D-inferences cannot arise in the absence of NU-inferences (see Sayre-McCord 1986, Fusco 2015, Booth 2022 for related claims).[2] That is, they claim that the D-inferences above can only arise with the NU-inferences in (5a) and (5b), which arise through negating the simpler universal alternatives to (4).

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

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

If this claim is correct, D-inferences should then go hand-in-hand with NU-inferences for disjunction under universal modals, as predicted by the NU-based approach. Based on the findings of Crnič et al. 2015, one would therefore expect *every*-sentences and *must*-sentences to be judged differently in critical cases where their D-inferences are true but one of their NU-inferences is false: put simply, the former should be accepted, the latter rejected. These predictions were tested in two sentence–picture acceptability experiments reported below.

Before proceeding, we clarify that these predictions do not depend on the Kratzerian view of *must* as a universal quantifier, which we adopt here for simplicity. Similar predictions arise under probabilistic accounts of modals (e.g., Swanson 2006, Yalcin 2007, Lassiter 2016). For example, on Lassiter (2016)'s account, 'must(p)' is

2 Crnič et al. discuss a different but analogous example to (4); see Section 4.

true only if the probability of p meets or exceeds a high threshold θ, i.e., only if $Pr(p) \geq \theta$. On its own, this account does not capture the D-inferences of interest: as long as $Pr(p) \geq \theta$, 'must($p \vee q$)' is predicted to be true even if $Pr(q) = 0$, that is, even if q is impossible. However, this account can be combined with an NU-based account to D-inferences. This is, in fact, what Santorio & Romoli (2017) propose: the meaning of 'must($p \vee q$)' can be enriched via SIs by negating the more informative alternatives 'must(p)' and 'must(q)', yielding the probabilistic counterparts of the NU-inferences above, namely $Pr(p) < \theta$ and $Pr(q) < \theta$. Since the probability of a disjunction is the sum of the probabilities of its disjuncts minus the probability of their conjunction, $Pr(p \vee q) > \theta$ together with $Pr(p) < \theta$ and $Pr(q) < \theta$ entails $Pr(p) > 0$ and $Pr(q) > 0$, which are the D-inferences: p is possible and q is possible. In sum, probabilistic accounts of modals must be supplemented with a theory of D-inferences; if supplemented with the NU-based account, the predictions are the same as those we obtain on a Kratzerian analysis of modals.[3]

2 Experiments

We investigated D-inferences arising from disjunction under universal determiners and universal modals in two acceptability experiments. Experiment 1 tested *every*-sentences like (1) and *must*-sentences like (4), where *must* expressed epistemic necessity. Experiment 2 followed up using a similar design, but this time with *must* expressing deontic necessity. Our aim was to test whether *must*-sentences differ from *every*-sentences in their capacity to yield D-inferences independently of NU-inferences, as claimed by Crnič et al. and others.

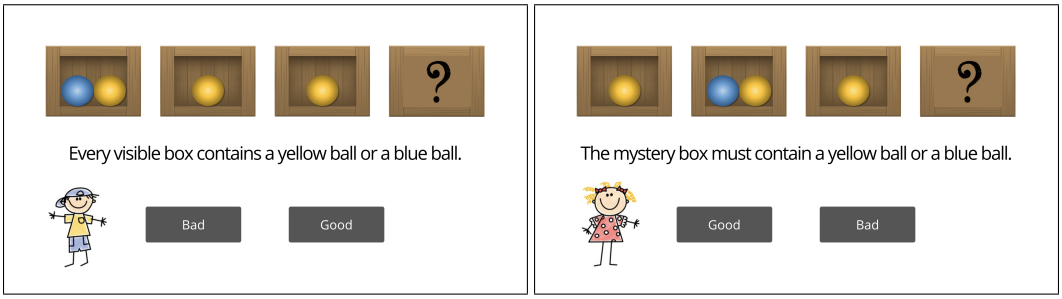
The target conditions in both experiments mirrored the critical conditions in Crnič et al.'s study. In the TARGET-1 conditions, test sentences were paired with pictures that made one of their NU-inferences false but their D-inferences true; in the TARGET-2 conditions, they were paired with pictures that made one of their NU-inferences and one of their D-inferences false. Based on Crnič et al.'s findings, we expected the *every*-sentences to be accepted less often in the TARGET-2 than in the TARGET-1 conditions. The key question was whether a comparable contrast would be observed for *must*-sentences.

2.1 Participants

For each experiment, 100 native speakers of English were recruited online through Prolific using the same pre-screen criteria (First and primary language: English, Nationality: UK/US, Country of birth: UK/US, Colorblindness: No, Approval rate: $\geq 99\%$). The recruitment was set up so that each participant could only take part

3 We are grateful to one of the reviewers for bringing this point to our attention.

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



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

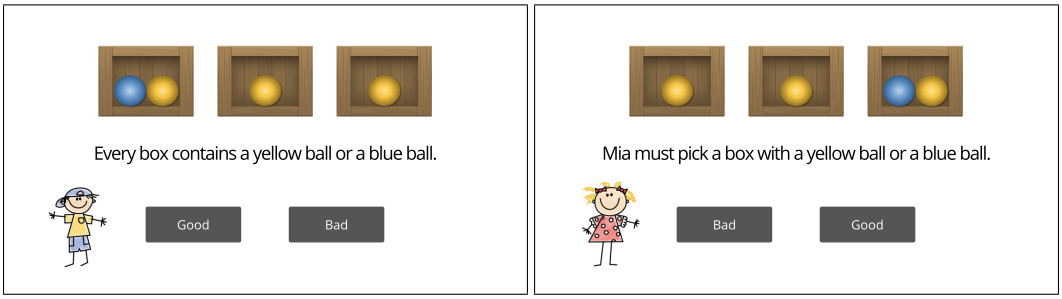


Figure 1 Example items illustrating the items' layout in Experiment 1 and 2. These examples correspond to A-AB-A instances of the TARGET-1 conditions for the nominal (left) and modal (right) cases.

in one of the two experiments. Participants were paid £2. 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.

2.2 Materials and Design

Both experiments adapted the materials and method from Marty et al. (2024b: Experiments 4–6) (see also Degano et al. 2025). Each item consisted of a sentence displayed beneath a row of horizontally aligned boxes and above the picture of one of two characters (see Figure 1 for examples). Sentences were constructed using the frames in Table 1. The color adjectives (the [A] and [B] terms in the table) were randomly selected 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 match the intended interpretation of *must* in each experiment.

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. [A] and [B] are placeholders for different color adjectives and [name] a placeholder for a character's name. For illustration, you may read [A] as *blue*, [B] as *yellow* and [name] as *Mia*.

Experiment 1 built on Marty et al.'s (2024b) implementation of Noveck's (2001) mystery box paradigm to investigate epistemic modality. The test sentences were *every*-sentences like (1) and *must*-sentences like (4), where *must* expressed epistemic necessity. Each item featured a display of four boxes: three open boxes containing one or two balls each, and one covered box marked with the symbol '?'. Participants were told that the characters could see the contents of the first three

boxes—the *visible* boxes—but not the covered one—the *mystery* box. They were also told that the characters had been taught the rule that the mystery box always matched the contents of one of the visible boxes. Their task was to judge whether the character's utterance was an appropriate description of the relevant box(es), given the information available to them and the rule they had been taught.

Experiment 2 extended the investigation to deontic modality, using the same basic materials and design as Experiment 1 but with an adapted cover story and modified modal sentences. The test sentences again included *every*-sentences, as in Experiment 1, and novel *must*-sentences where *must* now expressed deontic necessity. Each item featured a display of three open boxes (no mystery box). Participants were told that the characters were playing games and that, depending on the game, one of them either had to describe the contents of the boxes (NOMINAL cases) or had to pick one of them (MODAL cases). Their task was to judge whether the utterance was an appropriate description of either the box contents or the character's options, depending on the game scenario.

The rest of the design was identical across experiments. The contents of the open boxes were varied to produce different picture types corresponding to the experimental conditions. Test sentences were paired with four picture types, as shown in Table 2. The color of the A-balls and B-balls in the open boxes always matched the [A] and [B] color terms used in the sentences (e.g., *yellow* and *blue*); 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 randomized; the mystery box displayed on the items of Experiment 1 always appeared in the rightmost position.

Table 3 summarizes the truth-values of the literal meaning, NU-inferences, and D-inferences of the test sentence across picture conditions. Target pictures were designed to falsify one of the test sentences' NU-inferences, while ensuring that one or both of their D-inferences remained true. On the TARGET-1 pictures, each of the three open boxes contained an A-ball, and at least one also contained a B-ball–thereby making one NU-inference false but both D-inferences true. TARGET-2 pictures were created from the TARGET-1 pictures by replacing the B-ball(s) with A-ball(s), thus making one NU-inference and one D-inference false. Variants of the TARGET-1 and TARGET-2 pictures were included to introduce visual diversity, manipulating the number of B-balls for the former and A-balls for the latter. For experimental purposes, these were treated as sub-conditions, but since no response differences emerged across variants, responses were aggregated across sub-conditions for analysis (see Section 2.6).

FALSE and TRUE pictures were control pictures and served distinct purposes. FALSE pictures were designed to establish a clear baseline for rejection: one of the open boxes contained an A-ball while the other two contained balls of a non-

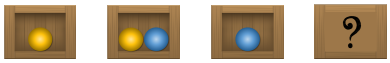
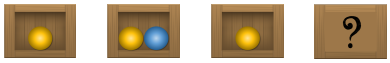
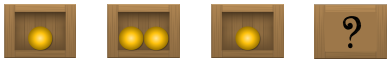
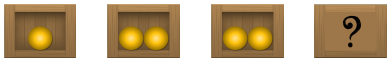
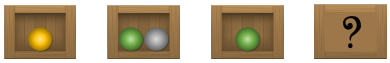
Condition		Example picture			
TRUE					
		A	AB	B	
TARGET-1	i.				
		A	AB	A	
	ii.				
		A	AB	AB	
TARGET-2	i.				
		A	AA	A	
	ii.				
		A	AA	AA	
FALSE					
		A	CD	C	

Table 2 Schematic description of the picture types paired with the test sentences in Experiment 1 and 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.*

matching color, making the test sentences literally false. TRUE pictures were designed to provide some relevant baseline for acceptance.[4] 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 sentences true irrespective of the inferences of interest.

In addition to the test sentences, each experiment included four types of control sentences: two positive (C1 and C2) and two negative (C3 and C4), involving either one color adjective (C1 and C3) or two (C2 and C4). Each control sentence was paired with a picture that made it either clearly true (GOOD) or clearly false (BAD), as described in Table 4. These controls were included to help identify low-effort responses. We were particularly concerned that some participants might engage

4 TRUE pictures were designed to closely resemble TARGET-1 pictures. Specifically, on these pictures, one open box contained both an A-ball and a B-ball, as on the TARGET-1 pictures. These pictures make the test sentences true, unless an exclusivity inference is derived at embedded level. As reported in Section 2.6, there is no evidence that participants ever derived such an inference.

			TRUE	TARGET-1	TARGET-2	FALSE
NOMINAL	Literal:	$\forall x(Ax \vee Bx)$	T	T	T	**F**
	NU:	$\neg\forall xAx$	T	**F**	**F**	
		$\neg\forall xBx$	T	T	T	
	D:	$\exists xAx$	T	T	T	
		$\exists xBx$	T	T	**F**	
MODAL	Literal:	$\Box(A \vee B)$	T	T	T	**F**
	NU:	$\neg\Box A$	T	**F**	**F**	
		$\neg\Box B$	T	T	T	
	D:	$\Diamond A$	T	T	T	
		$\Diamond B$	T	T	**F**	

Table 3 Truth-values of the literal meaning, NU-inferences, and D-inferences associated with the NOMINAL and MODAL test sentences across experimental conditions in the two experiments (T = True, F = False).

with the task superficially—for instance, by simply checking whether the colors mentioned in the sentence appeared in the picture. If so, they would likely perform poorly on control items involving negative sentences.

Pairing the test and control sentences with the relevant picture types yielded 4 test and 8 control conditions per quantifier type. Control conditions were instantiated 3 times each, and test conditions 6 times, yielding 24 control and 24 test trials per quantifier type. Instances of the TARGET-1 and TARGET-2 conditions were evenly distributed across their sub-conditions. In both experiments, trials were blocked by quantifier type to support participants' comprehension of the instructions and minimize the risk that responses to one type of test trial would be influenced by exposure to the other.

2.3 Procedure

The procedure was identical for both experiments, save the differences in instructions described above. In each experiment, NOMINAL and MODAL trials were presented in two separate blocks, with block order counterbalanced across participants. Each block began with instructions highlighting key aspects of the cover story,

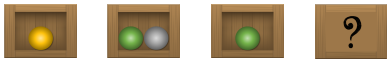
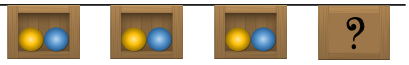
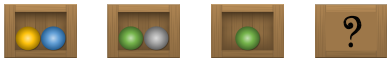
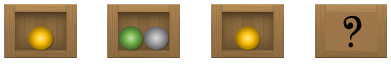
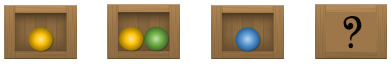
Sentence	Condition	Example picture
C1	GOOD	 A A A
	BAD	 A CD C
C2	GOOD	 AB AB AB
	BAD	 AB CD C
C3	GOOD	 CD CD C
	BAD	 A CD A
C4	GOOD	 C CD C
	BAD	 A CD A

Table 4 Schematic description of the picture types paired with the control sentences in both experiments. Picture types are illustrated using the same color assignment as in Table 2. *Note: the mystery box was only displayed on the items of Exp.1.*

followed by a short practice session to ensure participants understood the task. The practice included one instance of each control condition for the relevant quantifier type (8 trials total), with feedback provided after each response. Participants could only proceed to the test phase after answering all practice trials correctly. The test phase consisted of 48 randomized experimental trials. Participants responded by clicking one of two buttons labeled 'Good' and 'Bad', respectively. The label positions were counterbalanced across participants.

2.4 Predictions

On the NU-based account, D-inferences are contingent on NU-inferences: they arise as logical entailments only when the test sentences are interpreted with their corresponding NU-inferences. In our design, the TARGET-1 pictures falsify one NU-inference associated with the test sentences, whereas the TARGET-2 pictures falsify both one NU-inference and one D-inference (see Table 3 for details). Thus, if participants derive either type of inference, this should be detectable in their responses to the TARGET-2 items. Specifically, following Crnič et al. (2015), we assume that if participants attribute to the speaker an enriched meaning including either of these inferences, they should reject the test sentences as adequate descriptions of the TARGET-2 pictures. On these assumptions, if the NU-based account is correct, participants' responses should not differ between the TARGET-1 and TARGET-2 conditions, since both equally falsify one of the relevant NU-inferences.

Alternatively, finding that participants accept the test sentences more often in the TARGET-1 than in the TARGET-2 conditions would suggest that the presence of NU-inferences cannot be the sole basis for rejection in the TARGET-2 conditions, thereby lending support to the view that D-inferences can arise independently of NU-inferences. Based on the findings of Crnič et al. (2015), we expected this latter prediction to be borne out for the NOMINAL cases. Our goal, then, was to determine whether the same pattern holds for the MODAL cases, or whether these instead align with the NU-based predictions. In the latter scenario, we should find little to no difference in acceptance between TARGET-1 and TARGET-2; in the former, we should observe comparable contrasts across the two sentence types.

In formulating the above predictions, we remain mindful of recent critiques concerning the use of truth-value judgment tasks as diagnostics of interpretation. Specifically, several authors have cautioned that participants' responses in such tasks may reflect not only evaluations of literal or pragmatically enriched meanings, but also judgments about the overall quality of the speaker's utterance (Degen & Goodman 2014, Jasbi et al. 2019, Waldon & Degen 2020, Scontras & Pearl 2021). The same considerations apply to the present experiments. In particular, participants might judge the test sentences less favorably in the target conditions because they regard another sentence—e.g., 'Every (visible) box contains an [A] ball'—as a better description of the relevant pictures. Such alternative descriptions may be preferred either because they avoid potentially misleading inferences associated with the disjunctive sentences, or simply because they are less costly to produce.[5]

5 In an earlier version of this paper, we reported on similar experiments that only differed in the composition of the TARGET-2 pictures (see the project's OSF repository at https://osf.io/wds4n). Rather than A-AA-A(A) configurations, they featured A-AC-A(C) configurations, where C was a non-matching ball. A reviewer noted that these configurations may have led participants to reject

We note, however, that simpler descriptions such as 'Every (visible) box contains a [A] ball' are equally true of both the TARGET-1 and TARGET-2 pictures and, on the literal interpretation of the test sentences, truth-conditionally equivalent to the disjunctive descriptions with both picture types. Consequently, while the availability of such alternatives may render the test sentences comparatively prolix and thereby negatively affect participants' judgments for these sentences, this effect should arise similarly across conditions, provided that the relevant alternatives are comparably accessible in the TARGET-1 and TARGET-2 trials.[6] As explained above, the only systematic difference between the TARGET-1 and TARGET-2 pictures is whether the second disjunct (a B-ball) is visually instantiated. Under the hypothesis pursued here, this distinction is expected to matter insofar as participants are sensitive to whether each disjunct has an existential witness in the pictures accompanying the test sentences. Such sensitivity is predicted if the test sentences give rise to D-inferences whose truth-values differ across the two picture types. Accordingly, a reliable contrast between these conditions would support the conclusion that D-inferences can arise independently of NU-inferences.

2.5 Data availability and software

The experimental materials, instructions, data files, and analysis code are openly available on the OSF platform at https://osf.io/wds4n. Data preparation, visualization, and analysis were conducted in the R statistical environment (R Core Team 2023), using the packages `brms` (Bürkner 2017), `ggplot2` (Wickham 2016), `Hmisc` (Harrell 2023), `lme4` (Bates et al. 2015) and `car` (Fox & Weisberg 2019).

2.6 Data preparation

16 participants in Experiment 1 and 11 in Experiment 2 were excluded based on a pre-established criterion of scoring below 80% accuracy on control items. Responses to the TARGET-1 and TARGET-2 trials in both experiments were then examined for potential differences across picture variants (see Table 2). For each target condition, we fitted a generalized linear mixed-effects model (GLMER) with a logit link, predicting responses from the fixed effect of picture sub-type (dummy coded). All models included by-participant random intercepts, slopes, and their correlation, as

the test sentences because a salient alternative—e.g., 'Every (visible) box contains an [A] ball or a [C] ball'—offered a better description of these pictures. More generally, the presence of C-balls may have encouraged participants to consider ad-hoc alternatives to '[A]-ball or [B]-ball', resulting in a strengthened reading of the disjunction—e.g., *every (visible) box contains an [A] ball or a [B] ball, but no [C] balls*—which is false in these contexts. To eliminate this confound, the present experiments exclusively used A-AA-A(A) configurations.

6 We return to this assumption and discuss it critically in Section 3.

well as by-item random intercepts. Each model was compared to a corresponding null model lacking the fixed effect. None of the comparisons revealed a significant difference, indicating that picture sub-type did not reliably influence responses. Accordingly, responses within each of the TARGET-1 and TARGET-2 conditions were aggregated across picture sub-types for the main analyses.

2.7 Results

Figure 2 shows the mean acceptance rate (i.e., the proportion of 'Good' responses) by sentence type and picture type in both experiments. Overall, NOMINAL and MODAL sentences exhibited similar response patterns: in both experiments, acceptance rates were near ceiling in the TRUE and TARGET-1 conditions (all means$\geq$ 90%) with minimal variability, declined in the TARGET-2 conditions (62-80%) and were lowest in the FALSE conditions (all means$\leq$8%).

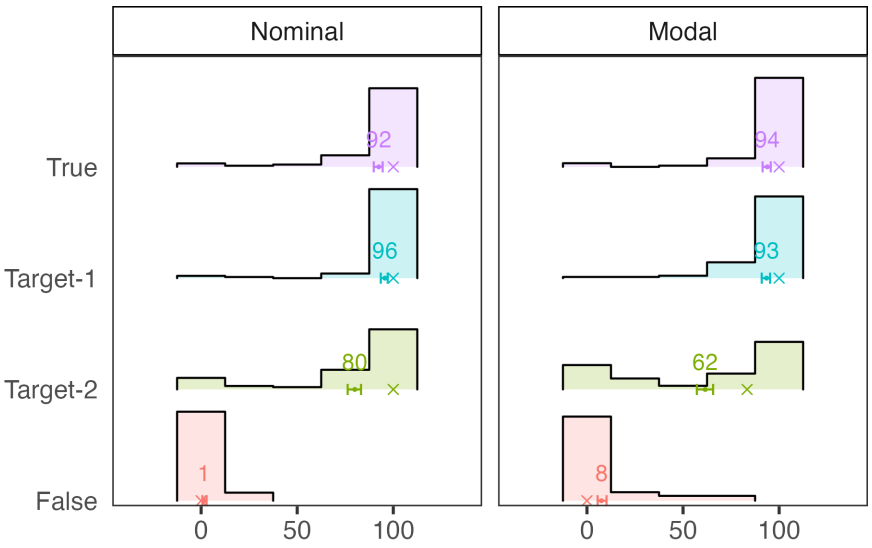
We analyzed the binary response data using Bayesian mixed-effects logistic regression models, implemented in the `brms` package in R (Bürkner 2017), with weakly informative priors.[7] For each experiment, the model predicted the probability of 'Good' response from the fixed effects of Picture (four levels: TRUE, TARGET-1,TARGET-2, FALSE), Sentence (two levels: NOMINAL, MODAL), and their interaction. Condition was dummy-coded with TARGET-2 as the reference level; Sentence was dummy-coded with NOMINAL as the reference level. We used the maximal random effects structure justified by the design (Gelman et al. 2020): by-subject random intercepts and slopes for the fixed effects and their interaction, and by-item random intercepts (each item appeared in only one condition-sentence pairing). Models were run with four chains, each with 4,000 iterations (warmup=2,000), using the NUTS sampler (adapt_delta=0.95). Convergence was confirmed by $\hat{R} = 1.00$ and sufficient effective sample sizes ($ESS > 1000$) for all parameters. Table 5 summarizes the fixed effects estimates and their 95% credible intervals.

The posterior estimates from the models were further used to test three directed hypotheses of primary interest: (i) TARGET-2<TARGET-1, (ii) TARGET-2<TRUE, and (iii) TARGET-1<TRUE, for each sentence type. For the MODAL level of Sentence, we leveraged the model's estimated population-level effects and interactions to assess contrasts without refitting models separately. Table 6 presents the results.

Both experiments yielded strong evidence that acceptance rates were lower

7 Normal(0, 2.5) for the fixed effects, Normal(0, 5) for the intercept, Exponential(1) for random effect standard deviations, and an LKJ(2) prior for correlation matrices, following the recommendations in Gelman et al. (2008). For completeness, we also ran our models using the default priors of the `brms` package (Normal(0, 10) for fixed effects and intercept, half Student-t(3, 0, 10) for random SDs, and LKJ(1) for correlations). These models yielded results supporting the same conclusions. We refer to the `brms` package manual for more information about the default priors, and to the code files for details on the statistical models and outputs.

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



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

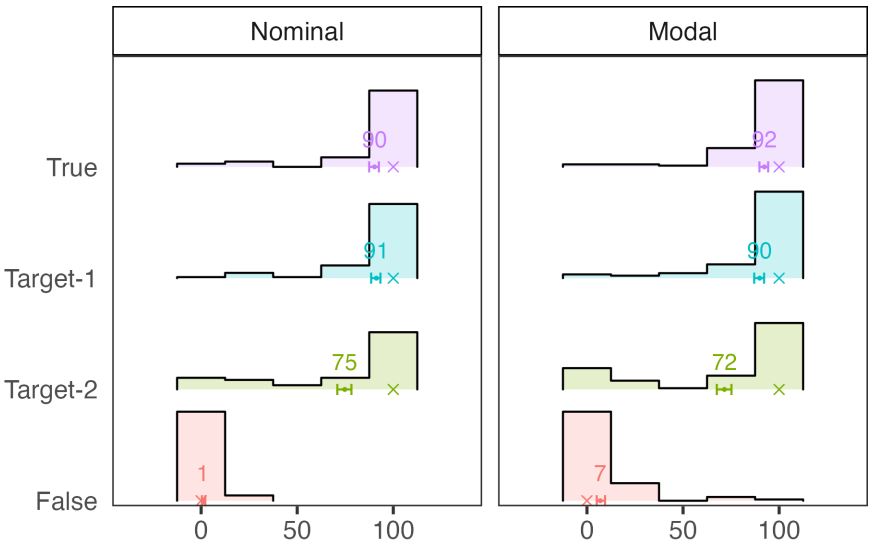


Figure 2 Mean acceptance rates by sentence and picture type in Experiment 1 and 2. Histograms show the distribution of mean acceptance rates across participants. The grand mean is indicated by a point with its rounded value labeled above, the 95% confidence interval is represented by an error bar, and the median is marked with a cross.

	Predictor	Estimate	2.5%	97.5%	Evidence
Exp.1	Intercept	2.79	1.97	3.67	*
	Picture=True	2.83	1.42	4.44	*
	Picture=Target-1	3.52	2.14	5.10	*
	Picture=False	-9.14	-10.78	-7.69	*
	Sentence=Modal	-1.64	-2.57	-0.68	*
	Picture=True : Sentence=Modal	3.01	1.49	4.68	*
	Picture=Target-1 : Sentence=Modal	1.36	-0.33	2.97	
	Picture=False : Sentence=Modal	1.57	-0.95	3.81	
Exp.2	Intercept	2.45	1.63	3.35	*
	Picture=True	2.86	1.57	4.34	*
	Picture=Target-1	2.21	1.18	3.32	*
	Picture=False	-9.01	-10.69	-7.44	*
	Sentence=Modal	-0.07	-0.87	0.79	
	Picture=True : Sentence=Modal	0.30	-1.41	2.00	
	Picture=Target-1 : Sentence=Modal	0.48	-0.86	2.00	
	Picture=False : Sentence=Modal	1.14	-1.13	3.11	

Table 5 Model parameter estimates and their 95% credible intervals for Experiments 1 and 2. Picture and Sentence were dummy-coded with TARGET-2 and NOMINAL as reference levels, respectively. $\hat{R} = 1$ for all estimated parameters. Rows marked with an asterisk indicate credible intervals that do not include zero, providing evidence for an effect.

in the TARGET-2 condition compared to both TARGET-1 and TRUE conditions, for both NOMINAL and MODAL sentences (all posterior probabilities=1, all evidence ratios>1000). There was no conclusive evidence that TARGET-1 yielded lower acceptance rates than TRUE for either sentence type (all posterior probabilities$\leq$0.83, all evidence ratios$\leq$4.87). In Experiment 1, there was evidence that the TARGET-1 vs. TRUE difference was larger for MODAL compared to NOMINAL sentences (see Table 5). However, no evidence was found for an interaction between Condition and Sentence for the TARGET-1 vs. TARGET-2 contrast in Experiment 1, nor for any interaction

	Hypothesis	Estimate	2.5%	97.5%	Post.Prob	Evidence
Exp.1						
NOMINAL	TARGET-1 < TARGET-2	3.52	2.34	4.81	1.00	*
	TARGET-2 < TRUE	2.83	1.63	4.15	1.00	*
	TARGET-1 < TRUE	-0.69	-2.25	0.81	0.23	
MODAL	TARGET-1 < TARGET-2	4.88	3.72	6.20	1.00	*
	TARGET-2 < TRUE	5.85	4.42	7.42	1.00	*
	TARGET-1 < TRUE	0.97	-0.66	2.70	0.83	
Exp.2						
NOMINAL	TARGET-1 < TARGET-2	2.21	1.33	3.13	1.00	*
	TARGET-2 < TRUE	2.86	1.77	4.07	1.00	*
	TARGET-1 < TRUE	0.65	-0.62	1.95	0.80	
MODAL	TARGET-1 < TARGET-2	2.69	1.61	3.97	1.00	*
	TARGET-2 < TRUE	3.16	1.88	4.60	1.00	*
	TARGET-1 < TRUE	0.47	-1.07	2.06	0.69	

Table 6 Summary of hypothesis tests comparing conditions within each sentence type. For each hypothesis, the table reports the posterior estimate, 95% credible interval, posterior probability, and evidence. Rows marked with an asterisk indicate credible intervals that do not include zero, providing evidence for an effect.

between Condition and Picture in Experiment 2.

3 Discussion

We found that MODAL sentences pattern closely with NOMINAL sentences. First, both sentence types were accepted less often in the TARGET-2 than in the TARGET-1 conditions. In fact, the TARGET-1 conditions showed near-ceiling acceptance rates, comparable to the TRUE conditions, providing no evidence that participants entertained the NU-inferences typically associated with the test sentences. Second, our comparisons across sentence types revealed no notable difference between

MODAL and NOMINAL sentences in their tendency to trigger D-inferences. Finally, the consistency of these patterns across both experiments suggests that the findings for MODAL sentences generalize across different flavors of *must*.

Prima facie, these findings support the view that both NOMINAL and MODAL sentences can similarly give rise to D-inferences independently of NU-inferences. In particular, while the results for *every* align with those of Crnič et al. 2015, the findings for *must* are novel and challenge the claim that, for disjunction under universal modals, D-inferences always occur alongside NU-inferences (Sayre-McCord 1986, Crnič et al. 2015, Fusco 2015, Booth 2022). Before endorsing this conclusion, however, we consider in turn two concerns regarding the setup of our experiments and discuss their possible ramifications for the interpretation of our results.

The first concern relates to the absence of evidence for NU-inferences and the possible role of the binary judgment paradigm in this outcome. Previous work has suggested that binary tasks using only the extreme ends of a response scale (e.g., 'wrong' vs. 'right') may underestimate participants' pragmatic sensitivity relative to tasks offering participants more graded response options (Katsos & Bishop 2011, Chemla & Spector 2011, Marty et al. 2015, Jasbi et al. 2019). In the present context, it is conceivable that some participants entertained interpretations of the test sentences enriched with NU-inferences, but nevertheless accepted these sentences in the TARGET-1 conditions because they were insufficiently confident that such strengthened interpretations were intended. On this view, a more graded response procedure—for instance, a Likert scale or slider task—might reveal subtle sensitivity to NU-inferences that remained undetected in our binary paradigm.

While this possibility cannot be excluded, it does not by itself explain why participants systematically rejected the test sentences more often in the TARGET-2 than in the TARGET-1 conditions. If D-inferences derivationally depend on NU-inferences, as assumed by NU-based accounts, uncertainty about whether NU-inferences are intended should affect responses in both conditions similarly. A stronger version of the objection could instead appeal to a threshold effect in participants' tolerance for pragmatic deviance. On such a view, participants may have regarded only the TARGET-2 cases as sufficiently misleading to warrant rejection because, unlike in the TARGET-1 cases, both inference types were false there. Although this possibility deserves consideration, we note that there is currently no evidence that binary judgment tasks systematically fail to detect a true contrast that would reliably emerge using a graded response procedure, provided a sufficiently large sample size.[8] In that sense, the consistent absence of a reliable difference between the TRUE and TARGET-1 conditions across both experiments remains difficult

8 We refer to Sprouse & Almeida (2017) and Marty et al. (2020) for a detailed discussion of this point in relation to the sensitivity of acceptability judgment tasks.

to reconcile with the idea that NU-inferences were nevertheless entertained by participants.

The second concern pertains to the role of alternative descriptions in the target trials. As noted earlier, participants may have noticed that the target pictures could be described more economically by non-disjunctive sentences such as 'Every (visible) box contains an [A] ball'. In interpreting our results, we assumed that, although the availability of such alternatives might negatively affect participants' judgments of the test sentences by comparison, this effect should remain relatively constant across target conditions. As a reviewer pointed out to us, however, this assumption is disputable. In particular, the simpler alternatives in question may have been more salient with the TARGET-2 pictures (e.g., A-AA-A(A)) than with the TARGET-1 pictures (e.g., A-AB-A(B)), since the former, unlike the latter, contain only A-balls. In sum, whether these competing alternatives prompted manner considerations or scalar inferencing, the lower acceptance rates observed in the TARGET-2 conditions could in principle reflect differences in the salience of these alternatives induced by the visual properties of the stimuli.

This explanation must nevertheless account for another key aspect of the data, namely the TRUE-like, near-ceiling acceptance rates observed in the TARGET-1 conditions. As we see it, one would have to assume here a fairly pronounced asymmetry between the TARGET-1 and TARGET-2 trials, either in the prominence of the A-AX-A(X) pattern or in participants' attention to this pattern, such that the simpler A-alternatives remained virtually inaccessible in the TARGET-1 trials. This is not impossible, but the required asymmetry appears somewhat stipulative. Furthermore, in our experiments, participants repeatedly encountered control items that required attending to whether all boxes contained balls of the same color (e.g., the C1 and C2 trials). Consequently, a salience-based explanation must also assume that the salience of these simpler descriptions remained highly localized and, specifically, did not extend to the TARGET-1 trials. Such an assumption sits somewhat uneasily with previous findings suggesting that the salience of an alternative tends to persist across trials once locally established (Skordos & Papafragou 2016, Marty et al. 2024a). In sum, while a salience-based account of the data remains possible, it appears more ad hoc and less parsimonious than the hypothesis that participants were sensitive to the truth or falsity of the relevant D-inferences.

Thus, for the time being, we conclude that our results challenge accounts of D-inferences that do not allow them to arise independently of NU-inferences across both nominal and modal quantifiers. Fortunately, several recent accounts derive D-inferences independently from NU-inferences (Bar-Lev & Fox 2020, Aloni 2022, Bar-Lev & Fox 2023, Crnič et al. 2015, Sudo 2023). In the remainder of this paper, we focus on two such accounts that can apply to both nominal and modal quantifiers

and thus explain our data.

The first is the novel implicature account of D-inferences proposed by Bar-Lev & Fox (2023).[9] The novelty of this account lies in the set of alternatives assumed for exhaustification. For a sentence like (1), this set is assumed to include all formal alternatives with the universal quantifier *every* as well as those where *every* is replaced with its existential counterpart, *some*. On this assumption, D-inferences can be derived by applying EXH recursively, as shown in (6). On the first application, the conjunctive alternatives $\forall x(\mathsf{A}x \wedge \mathsf{B}x)$ and $\exists x(\mathsf{A}x \wedge \mathsf{B}x)$ are excluded; on the second, D-inferences arise due to the presence of individual existential alternatives.

(6) $\textsc{exh}(\textsc{exh}\ \forall x(\mathsf{A}x \vee \mathsf{B}x))$
$\Leftrightarrow \forall x(\mathsf{A}x \vee \mathsf{B}x) \wedge \neg\textsc{exh}(\forall x\mathsf{A}x) \wedge \neg\textsc{exh}(\forall x\mathsf{B}x) \wedge \neg\exists x(\mathsf{A}x \wedge \mathsf{B}x)$
$\Leftrightarrow \forall x(\mathsf{A}x \vee \mathsf{B}x) \wedge \neg(\forall x\mathsf{A}x \wedge \neg\exists x\mathsf{B}x) \wedge \neg(\forall x\mathsf{B}x \wedge \neg\exists x\mathsf{A}x) \wedge \neg\exists x(\mathsf{A}x \wedge \mathsf{B}x)$
$\Leftrightarrow \forall x(\mathsf{A}x \vee \mathsf{B}x) \wedge \boxed{\exists x\mathsf{A}x} \wedge \boxed{\exists x\mathsf{B}x} \wedge \neg\exists x(\mathsf{A}x \wedge \mathsf{B}x)$

The combination of the prejacent, the D-inferences $\exists x\mathsf{A}x$ and $\exists x\mathsf{B}x$, and the strong exclusivity inference $\neg\exists x(\mathsf{A}x \wedge \mathsf{B}x)$ entails $\neg\forall x\mathsf{A}x$ and $\neg\forall x\mathsf{B}x$. However, Bar-Lev & Fox show that, on this account, D-inferences can also be generated without NU-like inferences: if the conjunctive alternatives are pruned from the set of alternatives, D-inferences remain while exclusivity inferences disappear, and the resulting reading is now compatible with both $\forall x\mathsf{A}x$ and $\forall x\mathsf{B}x$ being true.[10] Moreover, this account straightforwardly extends to universal modals if we assume, as above, that the relevant set of alternatives also includes forms where the universal modal (e.g., *must*) is replaced with its existential counterpart (e.g., *might*).[11]

The second account is the one offered in Aloni (2022), on which D-inferences arise from an interpretive bias dubbed *neglect-zero*. The key idea is that when interpreting a sentence, language users construct representations of possible situations (or states) that satisfy the sentence, but tend to disregard trivial configurations that would satisfy the sentence merely due to emptiness. Formally, Aloni develops this idea within Bilateral State-based Modal Logic (BSML), where a state is understood as a set of possible worlds reflecting the speaker's information state. In BSML, disjunction is supported in a state if that state can be split into two substates, each

9 Crnič et al. (2015) offer a different SI-based account that, for nominal cases, can also derive D-inferences independently of NU-inferences. For space reasons, we do not review their proposal here; see Bar-Lev & Fox (2023) for a detailed critique.

10 Strictly speaking, only the existential conjunctive alternative $\exists x(\mathsf{A}x \wedge \mathsf{B}x)$ needs to be pruned to prevent NU-like inferences from arising; in particular, the universal conjunctive alternative $\forall x(\mathsf{A}x \wedge \mathsf{B}x)$ can still be retained, yielding then the weaker exclusivity inference $\neg\forall x(\mathsf{A}x \wedge \mathsf{B}x)$.

11 We note that Bar-Lev & Fox (2023), building on the intuition by Crnič et al. (2015) presented below, introduce a stipulation about alternatives of the modal to prevent their account from extending to the modal case. However, given our results, this move appears unwarranted; without it, their account predicts D-inferences independently from NU-inferences across quantifier types.

supporting one of the disjuncts. Under this definition, a disjunction may thus be trivially supported if one or both disjuncts are vacuously supported by the empty substate. This is where the *neglect-zero* bias comes in.

Aloni (2022) formalizes this bias via an enrichment function, $(\cdot)^+$, which excludes the empty substate as a valid verifier. Consequently, an enriched disjunction $(\mathsf{A} \vee \mathsf{B})^+$ is supported in a state only if that state can be split into two *non-empty* substates, each genuinely supporting one of the disjuncts. This enrichment mechanism accounts for D-inferences under universal modals in predicting an enriched disjunction $[\Box(\mathsf{A} \vee \mathsf{B})]^+$ to entail both $\Diamond\mathsf{A}$ and $\Diamond\mathsf{B}$, regardless of whether the modal is epistemic or deontic. Since this mechanism does not generate NU-inferences, the account is also compatible with D-inferences being present without accompanying NU-inferences, consistent with our experimental findings. Finally, first-order extensions of this system, such as Aloni & van Ormondt (2023), generalize Aloni's account of D-inferences to disjunction under nominal quantifiers like *every*, thereby capturing the pattern we observed across both quantifier types.

4 Conclusion

The results of this study extend earlier findings from Crnič et al. (2015) by showing that D-inferences may arise independently of NU-inferences across universal quantifiers. Based on these findings, we argued that an adequate theory of D-inferences must allow for their derivation independently of NU-inferences, regardless of the quantifier involved, and we outlined two promising theoretical approaches that meet this challenge, the recent implicature-based approach by Bar-Lev & Fox (2023) and the neglect-zero approach by Aloni (2022). Importantly, although our data do not provide evidence for NU-inferences with modal sentences, prior work suggests that such inferences do arise in certain contexts. 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 where only sneakers are required (running shorts are optional). This observation aligns with the presence of NU-inferences, and it would be unexpected if only D-inferences were in play. We take this to suggest that a comprehensive theory of D-inferences should ultimately not only account for when distributive-only readings arise but also explain when they do not.

References

Aloni, Maria. 2022. Logic and conversation: The case of free choice. *Semantics and Pragmatics* 15(5). 1–60. http://dx.doi.org/10.3765/sp.15.5.

Aloni, Maria & Peter van Ormondt. 2023. Modified numerals and split disjunction:

The first-order case. *Journal of Logic, Language and Information* 32(4). 539–567. http://dx.doi.org/10.1007/s10849-023-09399-w.

Bar-Lev, M.E. & 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, M.E. & Danny Fox. 2023. On fatal competition and the nature of distributive inferences. *Natural Language Semantics* 31(4). 315–348. http://dx.doi.org/10.1007/s11050-023-09210-3.

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. http://dx.doi.org/10.1007/s11098-021-01706-0.

Bürkner, Paul-Christian. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80(1). 1–28. http://dx.doi.org/10.18637/jss.v080.i01.

Chemla, Emmanuel & Benjamin Spector. 2011. Experimental evidence for embedded scalar implicatures. *Journal of Semantics* 28(3). 359–400. http://dx.doi.org/10.1093/jos/ffq023.

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. 2025. The ups and downs of ignorance. *Natural Language Semantics* 33. 1–41. http://dx.doi.org/10.1007/s11050-024-09226-3.

Degen, Judith & Noah Goodman. 2014. Lost your marbles? the puzzle of dependent measures in experimental pragmatics. In *Proceedings of the annual meeting of the Cognitive Science Society*, vol. 36, https://cocolab.stanford.edu/papers/DegenGoodman2014-Cogsci.pdf.

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. London: Palgrave Macmillan UK. http://dx.doi.org/10.1057/9780230210752_4.

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(28). 1–27. http://hdl.handle.net/2027/spo.3521354.0015.028.

Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau & Yu-Sung Su. 2008. A weakly

informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* 2(4). http://dx.doi.org/10.1214/08-aoas191.

Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner & Martin Modrák. 2020. Bayesian workflow. https://arxiv.org/abs/2011.01808.

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

Jasbi, Masoud, Brandon Waldon & Judith Degen. 2019. Linking hypothesis and number of response options modulate inferred scalar implicature rate. *Frontiers in Psychology* Volume 10 - 2019. http://dx.doi.org/10.3389/fpsyg.2019.00189.

Katsos, Napoleon & Dorothy V.M. Bishop. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition* 120(1). 67–81. http://dx.doi.org/https://doi.org/10.1016/j.cognition.2011.02.015.

Lassiter, Daniel. 2016. Must, knowledge, and (in)directness. *Natural Language Semantics* 24. 117–163. http://dx.doi.org/10.1007/s11050-016-9121-8.

Marty, Paul, Emmanuel Chemla & Benjamin Spector. 2015. Phantom readings: The case of modified numerals. *Language, Cognition and Neuroscience* 30(4). 462–477. http://dx.doi.org/10.1080/23273798.2014.931592.

Marty, Paul, Emmanuel Chemla & Jon Sprouse. 2020. The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: a journal of general linguistics* 5(1). 72. http://dx.doi.org/https://doi.org/10.5334/gjgl.980.

Marty, Paul, Jacopo Romoli, Yasutada Sudo & Richard Breheny. 2024a. Implicature priming, salience, and context adaptation. *Cognition* 244. 105667. http://dx.doi.org/https://doi.org/10.1016/j.cognition.2023.105667.

Marty, Paul, Jacopo Romoli, Yasutada Sudo & Richard Breheny. 2024b. What Makes Linguistic Inferences Robust? *Journal of Semantics* 41(1). 1–52. http://dx.doi.org/10.1093/jos/ffad010.

Noveck, Ira A. 2001. When children are more logical than adults: experimental investigations of scalar implicature. *Cognition* 78(2). 165–188. http://dx.doi.org/https://doi.org/10.1016/S0010-0277(00)00114-1.

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

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

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

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

20(2). 179–197. http://dx.doi.org/10.2307/2215390.

Scontras, Gregory & Lisa S Pearl. 2021. When pragmatics matters more for truth-value judgments: An investigation of quantifier scope ambiguity. *Glossa: a journal of general linguistics* 6(1). http://dx.doi.org/10.16995/glossa.5724.

Skordos, Dimitrios & Anna Papafragou. 2016. Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition* 153. 6–18. http://dx.doi.org/https://doi.org/10.1016/j.cognition.2016.04.006.

Sprouse, Jon & Diogo Almeida. 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: a journal of general linguistics* 2(1). http://dx.doi.org/10.5334/gjgl.236.

Sudo, Yasutada. 2023. Scalar implicatures with discourse referents: a case study on plurality inferences. *Linguistics and Philosophy* 46(5). 1161–1217. http://dx.doi.org/10.1007/s10988-023-09381-6.

Swanson, Eric. 2006. *Interactions with context*: MIT dissertation.

Waldon, Brandon & Judith Degen. 2020. Modeling behavior in truth value judgment task experiments. In Allyson Ettinger, Gaja Jarosz & Joe Pater (eds.), *Proceedings of the society for computation in linguistics 2020*, 238–247. New York, New York: Association for Computational Linguistics. https://aclanthology.org/2020.scil-1.29/.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Yalcin, Seth. 2007. Epistemic modals. *Mind* 116(464). 983–1026. http://dx.doi.org/10.1093/mind/fzm983.

Paul Marty
Center of Linguistics of the University of Lisbon
Universidade de Lisboa
Alameda da Universidade
1600-214 Lisboa, Portugal
pmarty@edu.ulisboa.pt

Sonia Ramotowska
Institut Jean Nicod
29 Rue d'Ulm
75005 Paris, France
sonia.ramotowska@ens.psl.eu

Richard Breheny
Division of Psychology and Language Sciences
University College London
Chandler House, 2 Wakefield Street
WC1N 1PF London, UK
r.breheny@ucl.ac.uk

Jacopo Romoli
Institut für Linguistik
Heinrich Heine Universität Düsseldorf
Universitätstr. 1
40225 Düsseldorf, Germany
jacopo.romoli@hhu.de

Yasutada Sudo
Division of Psychology and Language Sciences
University College London
Chandler House, 2 Wakefield Street
WC1N 1PF London, UK
y.sudo@ucl.ac.uk