Diversity with Universality*

Sonia Ramotowska[1], Paul Marty[2], Jacopo Romoli[1], Yasutada Sudo[3], and Richard Breheny[3]

[1] Heinrich-Heine-Universität Düsseldorf
[2] L-Università ta' Malta
[3] University College London

Abstract

This paper investigates the 'Diversity' inferences (D-INFERENCES) arising from disjunction embedded in the scope of a universal quantifier, e.g., *Every X is A or B* suggests *Some Xs are A* and *Some Xs are B*. It has previously been claimed (i) that D-INFERENCES are independent from 'Negative Universal' inferences (NU-INFERENCES), which are the negations of *Every X is A* and *Every X is B*, but (ii) that for disjunction in the scope of a universal modal the D-INFERENCES cannot be observed independently of the NU-INFERENCES ([3, 7, 9, 13]). Experiment 1 tested the availability of D-INFERENCES in the absence of NU-INFERENCES for the determiner *every* and the epistemic modal *must*. Experiment 2 followed up on Experiment 1 by testing the same two quantifiers, only this time the modal *must* expressed deontic necessity. The results show that, for both types of quantifiers, D-INFERENCES could be derived independently of NU-INFERENCES. While the results for *every* essentially replicate those reported in [7], the results for *must* are new and go against the aforementioned claim (ii). In addition, the response time results from both experiments show that D-INFERENCES are associated with response delay effects in the opposite direction to those observed for regular scalar implicatures in similar tasks ([4, 5]). We argue that these findings about the time course of D-INFERENCES raise a new challenge for an implicature-based approach to these inferences.

1 Introduction

A sentence like (1), where disjunction occurs in the scope of a nominal universal quantifier, gives rise to the inference in (1-a), schematised in (1-b). Inferences like (1-a) are generally referred to as 'Diversity' or 'Distributive' inferences. We will use the more compact label 'D-INFERENCES'.

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

  a. $\rightsquigarrow$ *Some visible box contains a blue ball and some contains a yellow ball*
  b. $\exists x \mathsf{A}x \wedge \exists x \mathsf{B}x$

While the existence of D-INFERENCES is uncontroversial, the status and source of these inferences is still very much debated. Traditionally, they are considered scalar implicatures, derived through negating the alternatives to (1) in (2-a) and (2-b) ([8, 12]).

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

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

This account of D-INFERENCES makes a straightforward prediction: D-INFERENCES should always arise in combination with the additional 'Negative Universal' inferences (henceforth, NU-INFERENCES) in (3-a), schematised in (3-b).

(3) a. $\rightsquigarrow$ *Not all visible box contain a blue ball and not all of them contain a yellow ball*
b. $\neg\forall x \mathsf{A}x \wedge \neg\forall x \mathsf{B}x$

There is experimental evidence that challenges this prediction ([7]): Sentences like (1) were robustly judged as true, when their D-INFERENCES were true while their NU-INFERENCES were false (e.g., in situations where all of the visible boxes contain a yellow ball and some of them contain also a blue one), and rejected when their D-INFERENCES and NU-INFERENCES were both false (e.g., in situations where all of the visible boxes contain a yellow ball and none of them also contains a blue one). These results suggest that D-INFERENCES can arise independently of NU-INFERENCES. In fact, no evidence for the presence of NU-INFERENCES was found.

In addition to this finding, it has also been considered, based on introspective judgments, that for disjunction in the scope of a modal quantifier the D-INFERENCE cannot be observed independently of the NU-INFERENCE ([3, 7, 9, 13]). For example, the D-INFERENCE associated with a sentence like (4) is considered to always arise with its NU-INFERENCE, (4-b).

(4) The mystery box must contain a blue ball or a yellow ball. $\Box(\mathsf{A} \vee \mathsf{B})$
$\rightsquigarrow$ *The mystery box might contain a blue ball and it might contain a yellow ball*
a. $\Diamond\mathsf{A} \wedge \Diamond\mathsf{B}$
b. $\neg\Box\mathsf{A} \wedge \neg\Box\mathsf{B}$

Competing accounts have been proposed for the D-INFERENCES of disjunction under universal nominal quantifiers ([1, 2, 7, 10, 14]), but crucially, some of them predict disjunction under universal modals to give rise to D-INFERENCES independently of NU-INFERENCES ([14, 10, 1]), while others do not ([2, 3]).

To adjudicate between these opposite predictions, we carried out two sentence-picture verification experiments. The results replicate the earlier experimental finding that D-INFERENCES can be observed without NU-INFERENCES for disjunction under nominal universal quantifiers, but they also show that for disjunction under universal modals, D-INFERENCES can be observed independently, contrary to what has been assumed in the literature. In addition, we observe that D-INFERENCES trigger effects on response times that are the opposite of what is typically observed for scalar inferences, which we take to be a challenge for accounts of D-INFERENCES as a special case of scalar inferences.

2 Experiments

2.1 Participants

For each experiment, 100 native speakers of English were recruited online through Prolific.ac ([11]) using the same pre-screening criteria (first language: English, nationality: UK/US, country of birth: UK/US, prior approval rate: above 90%). Participants were paid £2.20, and average completion time was about 13 minutes. 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

Examples of sentence-picture items used in Experiment 1 and Experiment 2 are given in Table 1. In each experiment, sentences were presented as being uttered by one of two characters, Sam or Mia, and the participants' task was to decide whether or not the character's utterance was a good description of the picture accompanying that utterance.

EXAMPLE SENTENCE	EXAMPLE PICTURE				CONDITION
Experiment 1					
NOMINAL	A	AB	B	?	TRUE
Every visible box contains a yellow ball or a blue ball.	A	AB	A	?	TARGET-1
MODAL EPISTEMIC	A	AC	A	?	TARGET-2
The mystery box must contain a blue ball or a yellow ball.	A	CD	C	?	FALSE
Experiment 2					
NOMINAL	A	AB	B		TRUE
Every box contains a yellow ball or a blue ball.	A	AB	A		TARGET-1
MODAL DEONTIC	A	AC	A		TARGET-2
Sam must pick a box with a blue ball or a yellow ball.	A	CD	C		FALSE

Table 1: Example items illustrating the experimental conditions for the target sentences in Experiments 1 and 2. There were 6 instances of each condition per quantifier type.

In Experiment 1, target sentences were NOMINAL sentences like (1), involving the determiner *every*, and MODAL sentences likes (4), in which *must* was used as an epistemic modal. Pictures involved three open ('visible') boxes, each of which contained one or two colored balls, and one closed ('mystery') box, with a question mark on it. Participants were instructed that the mystery box always has the same contents as one of the three visible boxes. They were asked to decide whether or not the sentence was a good description of what's inside the relevant box(es). In the control conditions, target sentences were paired with pictures that make them either TRUE or FALSE, regardless of the inferences under investigation. In the TARGET-1 conditions, they were paired with sentences that make their NU-INFERENCES false, but their D-INFERENCES true. In the TARGET-2 conditions, they were paired with sentences that make both these inferences false. Quantifier type (2 levels: NOMINAL, MODAL) and picture type (4 levels: TRUE, TARGET-1, TARGET-2, FALSE) were crossed to obtain a 2×4 within-subjects factorial design.

Experiment 2 was built on Experiment 1 by adapting the materials and instructions from Experiment 1 so as to extend our investigation to deontic modality. Target sentences were NOMINAL sentences with the determiner *every*, similar to those tested in Experiment 1, and novel MODAL sentences in which the modal *must* was used to express deontic necessity. Pictures were similar to those used in Experiment 1, except that they involved only three boxes, all open.

In the instructions, participants were told that the two characters were playing together 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 to test NOMINAL sentences and the second to test MODAL sentences. 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 2 was identical to that of Experiment 1 in all respects.

2.3 Procedure

The procedure was the same in both experiments, provided the differences in instructions we described above. NOMINAL trials and MODAL trials were presented in two separate blocks, the order of which was counterbalanced between participants. Each block started with a short training and included additional control items designed to check that subjects understood the instructions properly and had no problem understanding the quantifiers and connectives used in the target sentences. In each block, trials were presented in a random order. Participants provided their answers by clicking one of two response buttons, labelled 'Good' and 'Bad', respectively. Participants' responses and response times were recorded on each trial.

2.4 Data treatment

8 participants in Exp. 1 and 6 participants in Exp. 2 were excluded prior to analyses because their performance on the control items was below the pre-established threshold of 80% accuracy. For both experiments, participants' responses and response times on the target NOMINAL and MODAL trials were analysed using mixed-effects regression models with the maximal random effect structure justified by the design and supported by the data. Figure 1 shows the mean acceptance rate to the target sentences by quantifier type and experimental condition. Figure 2 shows the mean RTs by quantifier type, experimental condition and response type.

2.5 Responses

Overall, the response patterns for MODAL sentences were similar to those for NOMINAL sentences. In both experiments, both sentence types gave rise to intermediate acceptability ratings in their TARGET-2 conditions (comparisons with TRUE and FALSE: all $|\beta|$s$>$ 4.34, all ps$<$ .001), with a bimodal distribution of by-subject mean ratings (unimodality tests: all Ds$>$ 0.10, ps$<$ .001) while they were both uniformly accepted in their TARGET-1 conditions (comparisons with TRUE: all $|\beta|$s$<$ 0.42, *ns*; comparisons with FALSE: all $|\beta|$s$>$ 12.5, ps$<$ .001; unimodality: all Ds$<$ 0.05, *ns*). Responses to NOMINAL and MODAL sentences in the TARGET-1 and TARGET-2 conditions were further compared by testing the interaction between quantifier and picture types. A significant interaction was found in Experiment 1 ($\beta = -1,80$, $SE = .49$, $p < .001$), showing that the acceptability contrast between TARGET-1 and TARGET-2 was greater in the MODAL than in the NOMINAL cases; no such an interaction was found in Experiment 2 ($\beta = -0,42$, $SE = .32$, $p = 0.19$). These results replicate the main results from [7] and establish that the key contrasts previously observed for NOMINAL quantifiers reproduce with MODAL quantifiers.

2.6 Response times

In both experiments, NOMINAL and MODAL sentences showed response delay effects in the opposite direction to those commonly observed for regular scalar implicature in similar acceptability judgement tasks. Specifically, in the TARGET-2 conditions, participants took significantly longer

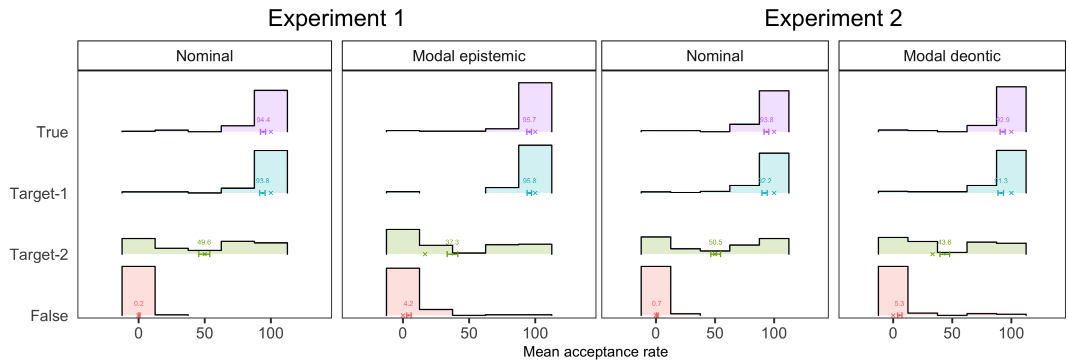

Figure 1: Mean acceptance rate by quantifier type and experimental condition. The distribution of by-participant mean ratings is visualised by a histogram, the grand mean by a thick bar with its value on top and the 95% CI around it, and the median by a cross.

to accept than reject both NOMINAL and MODAL sentences (all βs$>$ 333, all ps$<$ .01). By contrast, in the control conditions, participants were significantly faster to accept than reject these same sentences (all βs< -241, all ps$<$ .001) while they were equally fast at rejecting them in their FALSE and TARGET-2 conditions in both experiments (all $|\beta|$s$<$ 38, *ns*) and equally fast at accepting them in their TRUE and TARGET-1 conditions in Experiment 1 (all $|\beta|$s$<$ 14, *ns*).

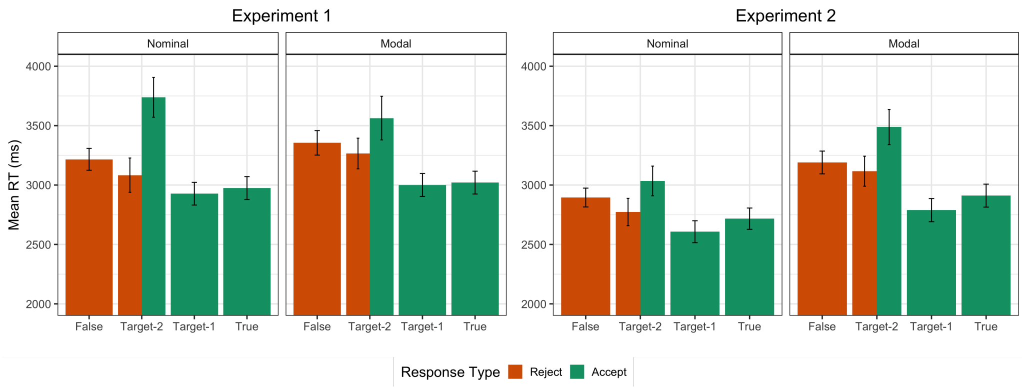

Figure 2: Mean RTs (in ms) by sentence type, experimental condition and response type. Error bars represent 95% CIs. RTs were analysed by considering correct responses in the control TRUE ('accept') and FALSE ('reject') conditions, 'accept' responses in the TARGET-1 conditions, and both 'accept' and 'reject' responses in the TARGET-2 conditions.

3 Discussion

We experimentally investigated the D-INFERENCES and NU-INFERENCES associated with disjunction embedded under the nominal universal quantifier *every* and the universal modal *must* in its epistemic and deontic readings. Our results show that disjunction under both *every* and *must* gives rise to D-INFERENCES independently of NU-INFERENCES. Also, the D-INFERENCES were overall observed to the same degree of robustness. The results for *every* essentially replicate previous finding by [7], but the results for *must* go against the claim in the literature that for disjunction under a universal modal, the D-INFERENCE is always observed in tandem with the NU-INFERENCE ([3, 7, 9, 13]). Our results, therefore, present a challenge to accounts of D-INFERENCES which cannot derive them without NU-INFERENCES for disjunction under universal modals.[1]

In addition, we found response delay effects in the opposite direction to those generally observed for regular scalar implicatures in similar verification tasks ([4, 5] among others). Specifically, we found that it takes more time to accept than reject nominal and modal sentences with disjunction in cases where their D-INFERENCES and NU-INFERENCES are false, compared to cases where only their NU-INFERENCES are. This processing slowdown, established across multiple comparison points, suggests that, in contrast to what is usually observed with regular scalar inferences, it is the suspension rather than the derivation of the D-inferences that is costly. This finding is thus challenging for accounts of D-INFERENCES as a special case of scalar inferences.[2]

Taken together, the present findings suggest that we need a theory of D-INFERENCES that can derive them independently of NU-INFERENCES for both disjunction under universal nominal quantifiers and disjunction under universal modals. The reaction time data also suggest that they should not be treated as a species of scalar inferences, but as a different type of inference, e.g., as proposed by [1].

Lastly, we would like to remark on the discrepancy between our experimental finding and the introspective judgments reported in the literature for disjunction under universal modals. Most notably, [7] observes that (5) feels misleading in a context in which there is a requirement to wear sneakers in the gym (while running shorts are optional).

(5) You are required to wear sneakers or running shorts.

Importantly, the oddness of (5) is unexpected if the sentence readily receives a reading with a D-INFERENCE but without an NU-INFERENCE. We do not mean to deny the validity of the observation regarding (5). Rather, we take it as suggesting that D-INFERENCES cannot always be observed independently of NU-INFERENCES and that a theoretical account of D-INFERENCES ultimately must be able to explain what the relevant restrictions consist in and why these restrictions exist.

References

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

[2] Moshe Bar-Lev and Danny Fox. Free choice, simplification, and Innocent Inclusion. *Natural Language Semantics*, 28:175–223, 2020.

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

[1] All such accounts can also derive a reading without either inference. This reading, however, is compatible with both our target conditions and, thus, it cannot account for the differences we found between the two.

[2] A similar challenge has been raised to implicature accounts of free choice inferences, which also do not give rise to the delay effects found with regular scalar implicatures ([6]). The main response to that challenge locates the source of this difference between free choice inferences and regular implicatures in the type of alternatives involved in their derivation. In particular, the idea is that the difference would be linked to the fact that regular implicatures are based on alternatives constructed by replacing lexical items in the asserted sentence, while free choice inferences only involve alternatives which are part of the asserted sentence. We note that this response does not extended straightforwardly to the case of D-INFERENCES. This is because, at least in some implicature accounts of D-INFERENCES (e.g. [2]), the alternatives involved in their derivation are constructed with lexical substitutions in the same way as in the case of regular implicatures.

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

[5] Lewis Bott and Ira Noveck. Some utterances are underinformative. *Journal of Memory and Language*, 51(3):437–457, 2004.

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

[7] Luka Crnič, Emmanuel Chemla, and Danny Fox. Scalar implicatures of embedded disjunction. *Natural Language Semantics*, 23(4):271–305, 2015.

[8] Danny Fox. Free choice and the theory of scalar implicatures. In Uli Sauerland and Penka Stateva, editors, *Presupposition and Implicature in Compositional Semantics*, pages 71–120. Palgrave, 2007.

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

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

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

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

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

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