Neglect-Zero Effects in the Interpretation of Quantifiers and Disjunction[1]

Oliver Bott — *University of Bielefeld*
Fabian Schlotterbeck — *University of Tübingen*
Tomasz Klochowicz — *ILLC, University of Amsterdam*
Sonia Ramotowska — *ILLC, University of Amsterdam*
Maria Aloni — *ILLC, University of Amsterdam*

Abstract. This paper presents a cross-experimental comparison investigating the interpretation and processing of four semantic/pragmatic phenomena, for which natural language interpretation deviates from literal meaning. In particular, we compared so-called neglect-zero phenomena in the case of quantifier interpretation – both for empty restrictors and empty scope sets – as well as distributivity inferences of disjunction embedded under universal quantification to the interpretation of scalar implicature in the case of *some*. The experiments employed a timed question-answering task which allowed us to test the just-mentioned phenomena embedded in polar questions. The data lend support to a broad differentiation of the phenomena into 1) presupposition violation in the case of empty restrictors, 2) neglect-zero effects for empty quantifier scope and distributivity inferences, and 3) negation of scalar alternatives for *some*.

Keywords: Neglect-zero, empty restrictor, empty scope, distributivity inferences, scalar inferences.

1. Introduction

The examples in (1) illustrate systematic patterns of interpretation observed in natural language that deviate from classical logic or the literal meaning of the expressions involved. Among them we find phenomena such as scalar implicature, or more generally upper-bounded construals (UB) (Grice, 1975; Levinson, 2000; Geurts, 2010; Chierchia et al., 2011)[2], distributive (DIST) inferences (e.g., Fox, 2007; Crnič et al., 2015), but also certain empty-set effects (ES) systematically showing up in the interpretation of mostly downward-entailing (DE) quantifiers (Bott et al., 2019) or with respect to quantificational domain restriction (Strawson, 1952; de Jong and Verkuyl, 1985; Lappin and Reinhart, 1988; Diesing, 1992; Geurts, 2008).

(1) a. Some of the squares are black ⇒ not all of the squares are black [UB]
b. Every square is black or white ⇒ there are black and also white squares [DIST]
c. Less than three squares are white ⇒ there are some white squares [ES, NE scope]
d. Less than three squares are white ⇒ there are some squares [ES, NE restrictor]

The present paper presents a comparison of the phenomena in (1) in a cross-experimental setup to investigate whether they are processed in a uniform way. If not, we are furthermore interested in whether observed interpretation and processing differences are in accordance with

[1]We would like to thank the audience at the NIHIL workshop 2024 in Amsterdam as well as the anonymous reviewers and audience at the *too many tools* workshop at the 29th *Sinn und Bedeutung* conference for their valuable comments. This research has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – CRC-1646, project number 512393437, project B01 and the Netherlands Organisation for Scientific Research (NWO) OC project *Nothing is Logical* (NihiL).

[2]The question whether the upper-bound inference is pragmatic or grammatical in nature that has been prominently discussed in experimental pragmatics (Breheny, 2019) is largely orthogonal to the current research enterprise, so we remain agnostic about the nature of scalar inferences. However, the data presented below showing a (weak) pragmatic effect of *some* embedded in polar questions may be relevant to this debate.

©2025, Oliver Bott, Fabian Schlotterbeck, Tomasz Klochowicz, Sonia Ramotowska and Maria Aloni.
Neglect-Zero Effects in the Interpretation of Quantifiers and Disjunction. In: Federica Longo and Daniele Panizza (eds.) *Proceedings of Sinn und Bedeutung 29*, pp. 214-232. Noto: University of Messina, C.U.M.O.

certain mechanisms and tools offered by the specific semantic and pragmatic theories on these phenomena.

1.1. Theoretical Background

Theories based on the negation of stronger alternatives provide a uniform solution to account for all these phenomena – and thus solve a potential too-many-tools problem. Besides the textbook example of the *not all* inference of *some* via the negation of the stronger alternative *all* (Grice, 1975; Geurts, 2010), free choice and distributivity inferences have also received alternative-based accounts via exhaustification of the individual disjuncts (Fox, 2007; Chierchia et al., 2011; Bar-Lev and Fox, 2020; Crnič et al., 2015)[3]. For the distributivity inference in (1b), Crnič et al. (2015: p. 278) proposed that the inferences *some square is black* and *some square is white* can be derived over the exhaustified alternatives of the two disjuncts separately from each other (see Bar-Lev and Fox (2023) for a different implicature account): *every square is only black*, *every square is only white*. Similarly, negation-of-alternatives accounts could also explain empty-set effects in DE quantifiers, i.e. difficulty and deviant responses when verifying situations where the set corresponding to their scope argument is the empty set. Bott et al. (2019) observed that participants reject such situations for quantifiers that have the empty set among their witness sets such as *fewer than n* or *at most n* in more than a quarter of all cases. An alternative-based account could be given since the negative indefinite *no* would be a better device unambiguously signalling the ES-situation. Hence, the pragmatic inference *it's not the case that no squares are white* may result (non-empty scope, NE scope).[4] There has also been broad discussion on determiners' interpretation in situations where they have empty restrictions as in (1d). The classic view is that (strong) determiners carry a presupposition about their domain to be non-empty, i.e. a non-empty (NE) restrictor inference (Strawson, 1952; de Jong and Verkuyl, 1985; Diesing, 1992; Geurts, 2008). However, Reinhart (2004) and Abusch and Rooth (2004) pointed out the possibility to derive the oddness of quantifiers' uses with an empty restriction by means of conversational implicatures. For sentences with strong quantifiers such as *every square is black*, the sentence is trivially true in case there are no squares, and there is an asymmetrical entailment from *there are no squares* to *every square is black*, so Abusch and Rooth (2004) suggested this to be a triggering configuration for a scalar implicature. Of course, the mentioned approaches focus on the particular phenomena and we therefore should emphasize that we do not find the claim in the literature that all four phenomena are uniform in nature. However, aiming at theoretical parsimony it is very appealing to embrace a *Negate Alternatives* approach covering all phenomena. One interesting prediction of a *Negate Alternatives* approach is that embedding under operators such as down-ward entailing operators but also embedding in questions should affect the rate of inferences as has been shown in research on embedded UB construals (e.g., Geurts and Pouscoulous, 2009; Potts et al., 2016; Franke et al., 2016).

Other approaches are more delimited in scope and maintain that the different phenomena are derived by different semantic or pragmatic principles and mechanisms. The non-logical responses in reaction to NE-SCOPE have been attributed to processing difficulty in connection with quantifiers that have the empty set among their witness sets (Bott et al., 2019; Balbach et al., 2022). Bott et al. (2019) proposed a theory of quantifier processing that describes how a

[3]See Franke (2011) for a game-theoretic formalization.

[4]This type of account is discussed and rejected based on (indirect) experimental evidence in Bott et al. (2019).

verification algorithm applicable to any model is constructed during online interpretation and how that algorithm is executed given a specific model to determine a truth value. In their theory they modeled natural language quantifiers as unary functions from non-empty restrictor sets to their witness sets (hence we will refer to their account as W(itness)-quantifier account). They showed a complexity difference between quantifiers that contain the empty set among their witnesses and quantifiers that do not. While the latter can be firmly evaluated with a simple verification algorithm ('*simple expansion*') that operates on positive information only (i.e. information about entities that are in the extension of a predicate) and uses only a single rule, the situation is different for empty-set quantifiers (ESQs). The latter require the encoding of both positive as well as of negative instances of a predicate in addition to a more complex set of a total of four rules. On their processing account comprehenders sometimes fail at applying the more complex verification algorithm explaining why they make more errors for ESQs than Non-ESQs and, in particular, in empty-set situations. Their theory allows to account for interactions of NE-SCOPE effects with other sources of quantificational difficulty, such as quantifier type (Geurts et al., 2010). But also general cognitive constraints such as working memory constraints should matter and quantifier processing is tightly connected to number cognition. In fact, Nieder (2016) makes a similar proposal regarding the processing and representation of the number concept of *zero* in number cognition: The concept involves an inference from the absence of entities to a positive evaluation, reminiscent to the particular rule proposed in Bott et al. (2019) required to evaluate ESQs in empty-set situations. Crucially for the present study, Bott et al.'s W-Quantifier theory predicts that logically incorrect NE-SCOPE interpretations should not involve an extra inferencing step (such as, e.g., negation of alternatives) but should even correspond to a simpler interpretation than logically correct interpretations. Bott et al. (2019) assume a presuppositional analysis of the NE-RESTRICTOR interpretation. One reason for this assumption is that otherwise the universal quantifier could also be considered an empty-set quantifier, which is not intended. Last but not least, the predicted occurrence of NE-SCOPE effects should be unaffected by sentence mood and should equally show up in the evaluation of assertions as well as in the answers to questions involving ESQs. The W-Quantifier account does not make any claims about phenomena such as pragmatic inferences in disjunction or scalars, but is fully compatible with a combination with other frameworks.

Aloni and colleagues proposed *neglect zero* as a more general interpretation principle that can be applied to a variety of phenomena such as free-choice inferences, DIST, homogeneity inferences with plurals and also quantification (Aloni, 2022; Aloni and van Ormondt, 2023; Sbardolini, 2023; Degano et al., 2025). They model these inferences in a bilateral state-based modal logic (BSML). To formalize the neglect zero tendency, they define a pragmatic enrichment function using the non-emptiness atom from team semantics. In this way, they build a formal framework to explain why people eliminate from their representations of sentences like (1b) those interpretations, which are true by the virtue of the empty set. As *simple expansion* in the W-quantifier theory, *Neglect zero* is motivated with recourse to general cognition: "When interpreting a sentence [we] create structures representing reality, pictures of the world (Johnson-Laird, 1983) and in doing so, [we] systematically neglect structures which verify the sentence by virtue of some empty configuration. This tendency, which I will call neglect zero, connects to the general preference for concrete (non-empty) over abstract (empty) configurations [...]." (Aloni, 2022: p. 4) Importantly, *neglect zero* is a tendency, i.e. a cognitive bias, to avoid certain models from which some non-classical inferences follow. These default

	Non-empty restrictor	Non-empty scope	Upper bound	Distributivity
Negate alternatives	✓	✓	✓	✓
W-quantifier	NA	✓	NA	NA
Neglect Zero	✓	✓	NA	✓
Presupposition	✓	NA	NA	NA

Table 1: Summary of the scope of the different accounts tested in our experimental study.

pragmatic inferences can only be overcome in case extra effort is put into shifting towards a classical logic 'reasoning mode'. This should clearly set *neglect zero* phenomena apart from scalar implicature, which have been shown to show up relatively late in acquisition, take extra time in processing and be subject to capacity limitations in processing (e.g., Chemla and Bott, 2014; Tieu et al., 2016; De Neys and Schaeken, 2007).[5]

For the NE-RESTRICTOR inference, Geurts (2008) and many others adopt a presuppositional account (Strawson, 1964). Under this view, this phenomenon is different from all the others discussed above showing features of presupposition violation and robust projection out of entailment-canceling environments. For our experiments below, we thus expect to find oddness responses to all cases in which NE-RESTRICTOR is violated and we expect this reaction also in return to questions indicating projection behaviour of this inference. The scope of phenomena addressed by the theoretical accounts discussed above is summarized in Table 1.

The working hypothesis employed in the present study is a multi-tool approach treating NE-RESTRICTOR as a presupposition and both NE-SCOPE and DIST as reflexes of *neglect-zero*. The *neglect-zero* interpretations have to be set apart from UB interpretations of scalars, which originate from the negation/exhaustification of alternatives.

1.2. Previous Experimental Research

We aim to investigate the NE-SCOPE (for ESQs) and more general NE-RESTRICTOR constraints as well as the DIST and UB inferences in a direct cross-experimental comparison. Previous studies have investigated the four phenomena separately. To these we will now briefly turn before laying out the experimental designs of and predictions for the present study.

ES effects in quantifier interpretation have been reported in Bott et al. (2019) and Balbach et al. (2022). In a total of four experiments, the same experimental setup was employed as in the present study. First, participants received a quantified sentence about geometrical objects in a self-paced fashion. Then, they moved to a picture screen and had to evaluate the sentence in the depicted situation, recording both responses as well as response times. Across all four experiments, NE-SCOPE effects were observed in response distributions as well as in response RTs. Most crucial for the present study was the finding that participants failed to arrive at the logically correct answer in at least 25% of the trials after having read an ESQ and having received an empty scope picture. This was also the condition exhibiting the longest RTs. Furthermore, the studies showed that difficulty increases (higher proportion of errors) in case quantificational complexity is increased. The experiments also tested quantifiers includ-

[5] According to this account, the DIST inference comes about by a systematic intrusion of pragmatic enrichments with non-empty constraints in combination with split disjunction, which requires the existence of two (after pragmatic enrichment: non-empty) states, one for each disjunct (Aloni, 2022; Aloni and van Ormondt, 2023).

ing exceptive constructions (Balbach et al., 2022) or iteration of quantifiers (Bott et al., 2019: Exp. 3). The latter turned out to be so difficult that participants by the majority gave logically illegitimate responses. Also related to quantificational complexity, the experiments revealed that Aristotelian quantifiers (ESQ: *kein* 'no'; non-ESQ: *jed-* 'every') led to smaller NE-SCOPE effects than comparative quantifiers. They furthermore showed NE-SCOPE effects for both monotone decreasing quantifiers such as the German counterparts of *fewer than n* or *at most n* as well as non-monotone quantifiers such *none or three*. Thus, the NE-SCOPE property is in fact a property of ESQs and therefore has to be distinguished from monotonicity.

Although NE-RESTRICTOR has been broadly discussed in the philosophy of language and in semantics and pragmatics, we are aware of only a single published experimental study on empty restrictors (Mankowitz, 2023). In her first experiment, Mankowitz tested sentences such as *Every/No American king lives in New York* and asked participants for their reaction to sentences with empty restrictors with the three options '*yes*', '*no*' or '*can't say*'. She explicitly tested a presuppositional account, for which she predicted a high proportion of '*can't say*' judgments. Contrary to this prediction, '*yes*' and '*no*' answers were observed with some variance between *every* and *no* sentences. This provides prima facie evidence against a presuppositional account. However, to us it is unclear how the '*can't say*' response relates to intuitively felt oddness of these examples. If *can't say* is interpreted epistemically – a point explicitly discussed in Mankowitz (2023) – it should be deliberately ignored because we know that USA is not a monarchy. We will therefore explicitly provide the possibility to mark the oddness of sentences in our experiments below. In addition, since we are interested in the projection behaviour of NE-RESTRICTOR, all considered quantifiers were tested in an entailment-canceling environment and test conditions involving NE-RESTRICTOR in polar questions.

Besides a growing number of experimental studies on free choice (e.g., Chemla and Bott, 2014), there have also been experimental investigations of DIST and related inferences (Crnič et al., 2015; Ramotowska et al., 2022; Marty et al., 2024; Degano et al., 2025). Studies eliciting the strength of this inference have shown that disjunctions embedded under universals yields about 20% (Crnič et al., 2015) to 40% DIST inferences (Ramotowska et al., 2022) across different tasks. The inference shows up 60-80% under modals and turns out to be less robust or absent under negation Marty et al. (2024). Ramotowska et al. (2022) furthermore measured RTs in response to situations violating DIST for disjunctions embedded under universals for which they observed a bimodal distribution allowing to separately analyze and compare acceptance and rejection RTs. The analysis revealed what they termed a reversed delay effect on top of a general penalty for DIST violations: acceptance of sentences violating DIST was significantly slower than their rejection – opposite to the delay effect observed for implicature. This is consistent with accounts treating the DIST interpretation as the default interpretation and interpretations without the inference as derivative of these defaults after their cancelation in line with the *Neglect Zero* approach.

The reverse delay effect stands in stark contrast to RT effects related to processing effects observed for scalar implicatures (see Breheny (2019) for a review). A recurring theme in this research is the striking processing difference observed for logical and pragmatic responses in situations in which the literal interpretation is true, but the UB inference is false (see, a.o., Noveck and Posada, 2003; Bott and Noveck, 2004; Bott et al., 2012). The basic finding that has been replicated in later work is that pragmatic responses take longer than literal responses,

i.e. the pragmatic delay effect (for a review and discussion see Khorsheed et al., 2022). This has been shown within the same participants in experiments explicitly instructing participants to react in a certain way but also in spontaneous reactions of the two types. Even though there is still ongoing debate about the timing of scalar implicature interpretation (Grodner et al., 2010; Breheny et al., 2012; Degen and Tanenhaus, 2015), Huang and Snedeker (2009) and more recently Huang and Snedeker (2018) have argued for a two-step processing model of implicated meaning relative to the literal interpretation. This is consistent with results from dual-task studies (De Neys and Schaeken, 2007) and mouse tracking (Tomlinson Jr et al., 2013) also supporting the view that UB interpretations are more effortful and take more time than literal interpretations. Also relevant for the present investigation, studies on scalar diversity have revealed rather high implicature rates *some* (Van Tiel et al., 2016) and the partitive construction *some of the* has been shown to independently increase implicature rates (Degen and Tanenhaus, 2015). This will be the chosen scalar expression in Exp. 3 below.

Since we are investigating the robustness of the four phenomena in polar question environments, we would like to briefly mention experimental work on embedded implicatures. Even though there is still ongoing discussion on this, various studies have shown embedded implicatures to exist (inter alia, Potts et al., 2016; Franke et al., 2016). Crucially, even though embedded implicatures are possible, implicature rates dramatically drop if lexical triggers of UB interpretations are embedded in monotone decreasing or non-monotone contexts. To our knowledge, there are no published studies on local UB readings of UB-triggering expressions in questions (but see the proposal in Bassi et al., 2021).

1.3. Predictions by the Multiple-Tool Approach advocated here

Our working hypothesis is a multiple-tool approach to the phenomena addressed in the experiments below. Having reviewed findings from the literature, we are now in a position to outline the predictions of the current study.

1. A presuppositional analysis of the NE-RESTRICTOR inference predicts oddness for all quantified polar questions *Are Q A B?* in situations in which *A* denotes the empty set. Hence, we expect '*odd question*' responses in these conditions and '*yes*' or '*no*' responses in all other conditions. Our experimental setup includes both strong (*every*) and weak quantifiers (*fewer than n*, *more than n* and *no*), which has played an important role in discussions on the respective presuppositions (see McNally, 2019), and we expect uniform answer behaviour with very robust inferences.

2. For NE-SCOPE we adopt a W-quantifier (as one specific type of *neglect zero*) analysis. We thus expect to observe clear NE-SCOPE effects for ESQs in situations where the scope set is empty. This effect should be modulated by quantificational complexity with higher proportions of non-logical responses and larger RT effects for the comparative ESQ *fewer than three* than for the Aristotelean ESQ *no*. Crucially, we predict NE-SCOPE responses to be as fast or perhaps even faster than logically legit responses. Since *neglect zero* equally applies to assertions as well as it does to questions, we expect NE-SCOPE effects of the same size as in earlier research testing verification of statements.

3. Similarly for DIST: *Neglect zero* lets us expect DIST effects in situations where all objects share the same feature. Since this pragmatic enrichment is a default inference, a

'*yes*' response in a situation violating DIST should be more taxing than responding '*no*'. The effect size of DIST effects should be in the range of earlier studies testing assertions.

4. A different pattern should be observed for UB inferences (for a treatment of scalar implicature allowing for projecting out of questions, cf. Bassi et al., 2021). Informed by existing work on embedded implicatures, which show up less frequently than unembedded ones, we predict relatively few UB interpretations overall and more logical responses in situations only consistent with the literal meaning of *Are some of the As B?*. In addition, the adopted *negation of alternatives* analysis predicts the opposite RT effect of answer polarity than for the *neglect zero* predictions above: Pragmatic '*no*' responses should show a clear delay relative to a logical '*yes*' response. It will also be interesting to see whether violations of scalar implicatures will lead to oddness responses (see Abusch and Rooth 2004, for discussion).

If, alternatively, all phenomena share the same underlying mechanism as a common source we expect similar processing behaviour and a uniform change when moving from declaratives to questions.

The predicted effects are expected to interact with general processing effects well known from the psycholinguistic literature. It is, e.g., a general finding that '*no*' answers tend to take longer than '*yes*' answers (Bott et al., 2019). As already reported in Just and Carpenter (1971), another classic finding is an interaction between *monotonicity* and *answer polarity* . While sentences with monotone increasing quantifiers are easier to verify than to falsify, the opposite pattern is observed for sentences with decreasing quantifiers. Below, we will use appropriate control conditions to estimate these effects on RT and estimate the effects of NE-SCOPE, NE-SCOPE, DIST and UB on top of these general effects in mixed-effects regression models.

2. Experimental Study

The present study included three sub-experiments for a cross-experimental examination of NE-RESTRICTOR and NE-SCOPE effects in ESQs vs. Non-ESQs (Exp. 1), DIST effects in disjunctions in the scope of universals (again in comparison with NE-RESTRICTOR cases) (Exp. 2) and UB effects in pragmatic and literal interpretations of the scalar expression *some of the* (Exp. 3). We introduce the general study design of the three subexperiments by way of example items.

The German sentence stimuli of Exp. 1 employed a 2 (TYPE OF QUANTIFIER: ARISTOTELEAN vs. COMPARATIVE) × 2 (MONOTONICITY: DOWNWARD ENTAILING (DE/ESQ) vs. UPWARD ENTAILING (UE/Non-ESQ)). A concrete example is given in (2):

(2) a. Ist jedes Quadrat blau? (Is every square blue?)
b. Ist kein Quadrat blau? (Is no square blue?)
c. Sind mehr als drei Quadrate blau? (Are more than three squares blue?)
d. Sind weniger als drei Quadrate blau? (Are fewer than three squares blue?)

The four sentence conditions were paired with the four picture types shown in Table 2. The first type violates NE-RESTRICTOR because there are no squares. The second type embodies an empty scope. It thus targets NE-SCOPE for ESQ quantifiers and a '*no*' response for Non-ESQ quantifiers.[6] In addition, two controls were included, one targeting a '*yes*' and

[6]For simplicity, we treat *no* as ESQ here, following the treatment in Bott et al. (2019) and contrary to that of Aloni and van Ormondt (2023). The results below show that *no* behaves in fact differently from *less than n*, though.

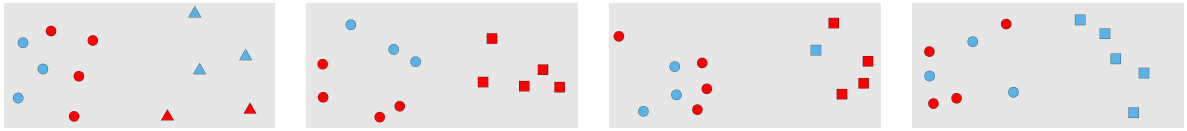


Table 2: Sample pictures tested with *are Q squares blue?* in Exp. 1. Conditions from left to right: empty restrictor, empty scope, control A, control B.

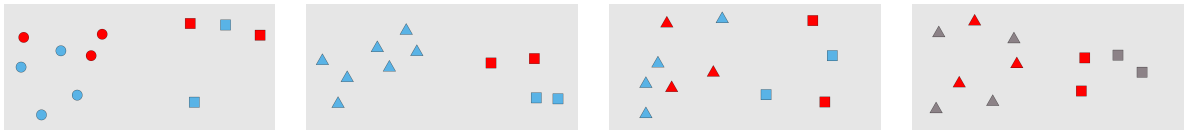


Table 3: Sample pictures tested with *is every triangle either blue or red?* in Exp. 2. Conditions from left to right: empty restrictor, DIST target, control '*yes*', control '*no*'.

one targeting a '*no*' response.[7] Taken together, the experiment employed a 4(PICTURE) × 2(QUANTIFIER TYPE) ×2(MONOTONICITY) within participants and items design.

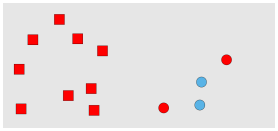
Subexperiment 2 investigating DIST inferences tested disjunctions in the scope of universals counterbalancing the order of the two disjuncts yielding two sentence conditions:

(3) a. Ist jedes Dreieck entweder rot oder blau? (Is every triangle either red or blue?)
b. Ist jedes Dreieck entweder blau oder rot? (Is every triangle either blue or red?)

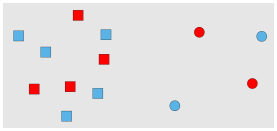
Each question was paired with four pictures, 1) empty restrictors, 2) DIST violations, the target condition, and 3) true and 4) false controls (see Figure 3). Thus, Exp. 2 employed a 4(PICTURE) × 2(DISJUNCTION ORDER) within-participants and items design.

Finally, in the UB subexperiment (Exp. 3) a simple one-factorial design was used comparing two picture conditions for a question involving *some of the* always with the partitive, which should yield high UB rates (Grodner et al., 2010; Degen and Tanenhaus, 2015; Huang and Snedeker, 2018). Here is a sample item:

(4) Sind einige der Quadrate rot? (Are some of the squares red?)



UB violation (literal)



control (literal + UB)

2.1. Methods

2.1.1. Participants

80 native German speakers (age between 18–40 y., no issues seeing colours, gender-balanced samples) were recruited from Prolific.com for payment of £3.75 and 20 were tested in each of the four experiment lists. From these, one participant was excluded during data preprocessing due to extremely long session time and 7 others were excluded due to error rates of more than 20% on control conditions and fillers. Thus, the sample included in the analysis comprised of

[7]Except for *kein* 'no', here both controls had to target a '*no*' response.

72 participants (39 female, 33 male; mean age 29.2 years; range 20–39 years), all scoring at least 80% correct on control conditions and fillers. A typical experimental session for a given participant had a median time of about 18 minutes.

2.1.2. Materials

All sentence materials are provided in the osf archive accompanying this paper (link to osf). For the creation of experimental items five shapes (square, circle, triangle, heart, cross) were combined with 11 colours (red, blue, yellow, green, orange, purple, turquoise, brown, pink, black, grey). The layout of the pictures always showed two sets of randomly distributed objects on the left and on the right, of which one served as a distractor, and the other involved the picture manipulation outlined in the experimental designs. The pictures also varied the number of objects shown in each half of the pictures (range 7 – 11 objects). The picture materials for the sentence materials were automatically generated by a Python script based on a spreadsheet summarizing the shapes, their relative position, their colours and their cardinalities.

Exp. 1 (NE-RESTRICTOR/NE-SCOPE): 20 items were created in 16 conditions each. The stimuli were distributed over four lists in a Latin Square design making sure that each question was only tested once per list with one of the four pictures. In total this yielded 80 trials in Exp. 1 with 5 trials for each participant in each condition.

Exp. 2 (NE-RESTRICTOR/DIST): 10 items were created in 8 conditions each. Lists were created in a Latin Square design such that each picture was only tested once within each list. Thus each item was tested four times in each list. In total, Exp. 2 comprised of 40 trials with 5 trials for each participant in each condition as in Exp. 1.

Exp. 3 (UB): 10 items were constructed in two conditions each. A Latin Square was used to test each item in only one condition per list. Thus, in total 5 trials per condition and participant were tested in Exp. 3, too.

In addition to the 130 experimental trials, the experiment included 60 filler trials testing materials involving clear cases of presupposition failures (uniqueness presupposition of singular definites, cardinality presupposition of *both*, . . .) but also trials in which clearly '*yes*' or '*no*' responses were rejected.[8]

In the results and discussion sections below we will often refer to the English counterparts of the German quantifiers used in the study. Whenever we do so, please keep in mind that this is just a convenient shorthand to actually refer to the German constructions tested.

2.1.3. Procedure

Adopting the methods from Bott et al. (2019) and Balbach et al. (2022), a combined self-paced reading plus picture verification web experiment was implemented in JavaScript. In the web-based experiment participants provided informed consent, entered relevant demographic data, then read written instructions. Before they moved to the experiment with randomized presentation of a total of 190 trials, they did a short practice of ten trials.

[8]The materials also included small pilot study testing homogeneity inferences and respective controls.

In each trial participants first were presented a screen with the question. After reading, they pressed space bar with their left hand to move to the picture screen for which they had to answer the question with the three response options: *komische Frage* ('odd question'), *ja, stimmt* ('yes, true') and *nein, stimmt nicht* ('no, false'). Answers were provided by pressing the left, down and right arrow keys with the index, middle and ring finger of their right hand, respectively. The mapping of keys to responses was randomly assigned for each participant. The question-answering stage had a timeout of 12 s.

In the reading stage, sentence reading time was recorded. In the question-answering stage, both judgments and judgment RTs were recorded.

2.1.4. Data analysis

Reading times and judgment RTs were preprocessed as follows. Whole sentence reading times above 10 seconds were excluded from the analysis. Key presses were coded into the three responses '*odd question*', '*yes*' or '*no*', respectively. Judgment RTs were trimmed by eliminating judgment RTs below 200 ms or more than three standard deviations above a participant's condition mean. The preprocessing and all analysis scripts are contained as RMD files in the accompanying osf archive (link to osf).

RT data were entered untransformed into linear mixed-effects regression (LMER) analyses and the response data were subjected to logit mixed-effects regression (GLMER) analyses with crossed random effects of participants and items. Analyses were conducted in R using the lme4 package (Bates et al., 2015). Linguistic stimuli only varied in Exp. 1, therefore **Reading times** were only analyzed in this experiment, including the fixed effects and random slopes of QUANTIFIER TYPE and MONOTONICITY as well as of SENTENCE LENGTH (number of characters). **Answers to questions** were analyzed in two separate logit mixed-effects models: The first analyzing '*odd question*' responses versus other responses and the second analyzing (logically) '*correct*' vs. '*incorrect*' responses, respectively. The analysis of **judgment RTs** was conducted on legitimate (logically or pragmatically) responses only, i.e. RTs of '*odd question*' responses in the empty restrictor conditions, either 'yes' or 'no' responses in the empty scope, DIST or UB target conditions, and '*yes*' (and '*no*) answers in the true (and false) control conditions.

In Exp. 1 and Exp. 2 the factor PICTURE was simplified by collapsing picture conditions into just three levels EMPTY RESTRICTOR, EMPTY SCOPE and CONTROLS. This factor was Helmert coded with the following two contrasts: 1) EMPTY RESTRICTOR vs. EMPTY SCOPE/CONTROL and 2) EMPTY SCOPE vs. CONTROL. The other two factors were centered and sum-coded. Exp. 3 employed simple treatment coding of the factor PICTURE with two levels (LITERAL ONLY vs. UB). Where needed RESPONSE was entered as an additional factor into analyses on judgment RTs of literal vs. non-literal responses. Last but not least, cross-experimental comparisons were conducted entering EXPERIMENT as yet another factor into the LMER analyses.

Each analysis started with the saturated model including all fixed effects and their interactions as well as crossed random intercepts and maximal random slopes by-participants and by-items.

2.2. Results and Discussion

We will first present and discuss the results of all three subexperiment in turn before we will present a cross-experimental comparison.

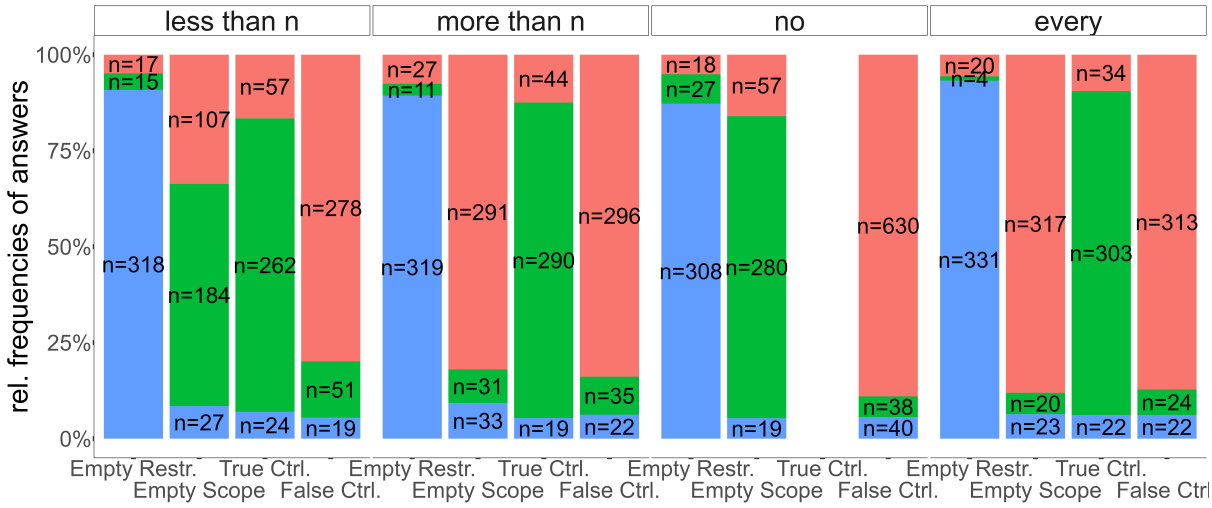


Figure 1: Distributions of answers chosen in response to questions about the picture conditions tested in Exp. 1 as a function of QUANTIFIER. Note: blue corresponds to '*odd question*' responses, green to '*yes, correct*' and red to '*no, incorrect*'; restr. = restrictor, ctrl. = control. For *no*, only false control conditions could be tested.

2.2.1. Experiment 1 – NE-RESTRICTOR and NE-SCOPE effects in Quantifiers

The analysis of reading times revealed main effects of QUANTIFIER TYPE and of MONOTONICITY on top of a significant main effect of SENTENCE LENGTH. Sentences with DE quantifiers took reliably longer to read than those with UE quantifiers ($\hat{\beta} = 86.2; SE = 22.7; t = 3.79; p < .01$) and sentences with comparative quantifiers had significantly longer RTs than sentences with Aristotelean quantifiers ($\hat{\beta} = 199.3; SE = 82.6; t = 2.41; p < .01$). These were the only reliable effects observed in the reading stage. Note that the observed reading time effect of MONOTONICITY points in the same direction as the effects observed in question-answering supporting the standard view that DE quantifiers are harder to process than UE quantifiers (e.g., Just and Carpenter, 1971; Geurts et al., 2010).

Figure 1 shows the relative and absolute distributions of answers in Exp. 1 as a function of quantifiers and pictures. The analysis of '*odd question*' revealed that this was the answer generally chosen in answering questions about empty restrictor situations. These conditions gave rise to rather uniform responses of 90.2 % '*odd question*' responses and the respective GLMER analysis revealed a significant effect for the Helmert contrast between empty-restrictor pictures versus all other situation types ($\hat{\beta} = 5.02; SE = .12; z = 41.13; p < .01$). This effect was not modulated by either QUANTIFIER TYPE or MONOTONICITY (all $|t| < 1.67$) showing in fact uniform presupposition failure for all quantifiers in questions about situations involving empty restrictor sets. The analysis of non-logical responses to empty-scope situations revealed a fundamentally different picture. In the analysis of logically correct vs. erroneous responses, only a significant two-way interaction between the Helmert contrast of empty scope models vs. controls and quantifiers' monotonicity was observed ($\hat{\beta} = 0.75; SE = .12; z = 6.02; p < .01$): The difference between empty-scope pictures and controls was larger for monotone decreasing quantifiers than for the control conditions. To confirm that empty-scope effects showed up for both quantifier types, we investigated the two quantifier types separately from each other. Both analyses re-

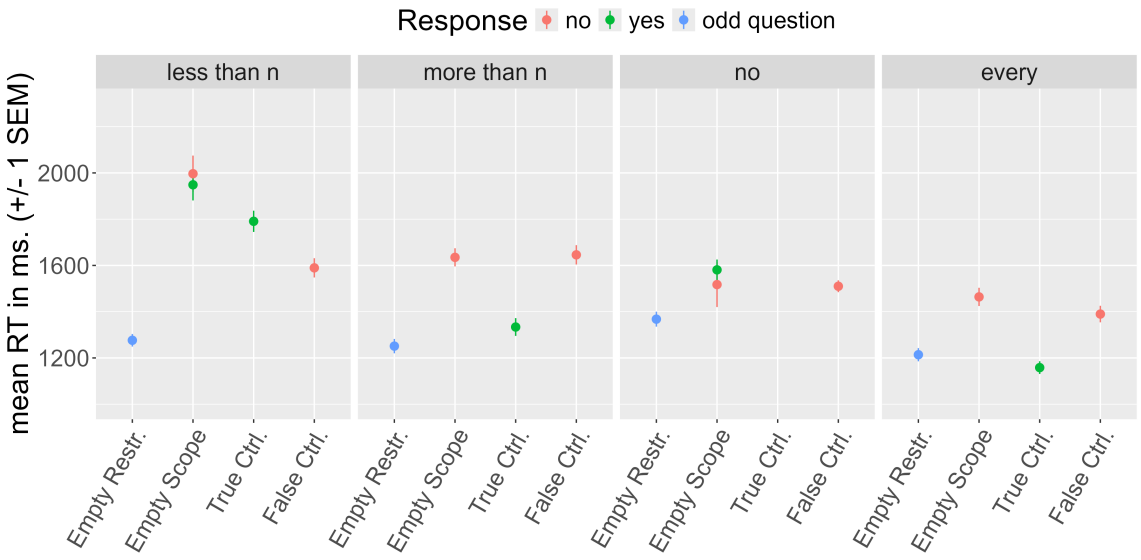


Figure 2: Mean judgment RTs in Exp. 1. Note that only "legitimate" responses are shown. In the case of ESQs both '*yes*' and '*no*' responses to empty-scope situations are included.

vealed significant two-way interactions. Thus, the 36.8% non-logical responses for *less than n* in empty-scope situations (interaction: $\hat{\beta} = 0.62; SE = .17; z = 3.76; p < .01$) as well as the 16.9% non-logical responses for *no* ($\hat{\beta} = 1.00; SE = .19; z = 5.17; p < .01$) relative to their respective controls gave rise to significant empty-scope effects in response behaviour. The size of these effects are as large as reported in the literature or even somewhat larger (Bott et al., 2019; Balbach et al., 2022). Thus, they turn out to occur robustly also in polar questions.

Figure 2 shows the mean response RTs in the question-answering task as a function of quantifiers and situation types shown in the pictures. The evaluation of empty-restrictor situations led to the fastest reactions relative to all other conditions with mean RTs well below 1500 ms for all four quantifiers. This finding is in line with the presuppositional view according to which the justification of the presupposition is a prerequisite for truth evaluation. This pattern was again uniformly observed for all quantifier types. Turning to the empty-scope conditions, a significant three-way interaction was observed in the LMER analysis between PICTURE (Helmert contrast empty scope vs. controls), QUANTIFIER TYPE and MONOTONICITY ($\hat{\beta} = -46.4; SE = 20.8; t = -2.23; p < .05$). This interaction resulted from a stronger empty-scope effect in comparative than in Aristotelean DE quantifiers relative to their UE counterparts. We also investigated general answer-polarity effects in interaction with the monotonicity of quantifiers. To this end, we additionally analyzed the contrast between true vs. false controls in combination with ANSWER POLARITY and MONOTONICITY as predictors. This analysis revealed a significant two-way interaction between ANSWER POLARITY and MONOTONICITY ($\hat{\beta} = 172.8; SE = 42.1; t = 4.11; p < .01$) originating from '*yes*' responses being faster than '*no*' responses for UE quantifiers and the reverse effect for DE quantifiers replicating the well-established finding from Just and Carpenter (1971). Note that for the Aristotelean quantifier *no*, this effect counteracts the empty-scope effect which seemed to have levelled it out: the empty-scope condition requires a '*yes*' response, which is due to the DE property of the quantifier should exhibit longer RT than a monotonicity-congruent '*no*' response in the false control

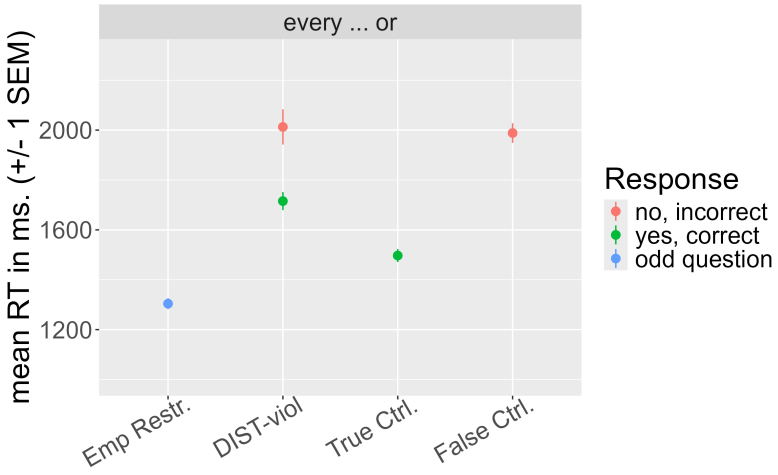


Figure 3: Mean answer RTs of legitimate responses in Exp. 2. Note: DIST-viol. = violation of DIST; ctrl. = control.

condition. For *less than n* we found the empty scope condition to be reflected by the highest RTs, an effect well exceeding the general monotonicity effect as becomes obvious by comparison with the true control condition, which also required a '*yes*' response.

Finally, we investigated the existence of answer polarity effects in the empty-scope target conditions, that is RTs of the quantifiers *less than n* and *no* for '*yes*' and '*no*' responses. For both quantifiers, there was no reliable difference between '*yes*' and '*no*' responses (pairwise comparisons, both $|t| \leq 1$).

2.2.2. Experiment 2 – NE-RESTRICTOR and DIST effects in Embedded Disjunction

The distribution of answers in Exp. 2 investigating disjunction embedded under universals was as follows. The DIST violation condition was rejected 22.7% of all cases in which no '*odd question*' response was given (abs. '*yes*': N=524, '*no*': N=154, '*odd question*': N=22). In the empty restrictor condition, again overwhelmingly '*odd question*' responses were observed in 89% of all cases (abs. '*yes*': N=56, '*no*': N=22, '*odd question*': N=632). The true and false control conditions received 92% '*yes*' and 79% '*no*' responses, respectively. A first GLMER analysis contrasted '*odd question*' responses with all other responses. The analysis again revealed a significant contrast between the empty restrictor condition versus the other three situation types ($\hat{\beta} = 5.79; SE = .60; z = 9.69; p < .01$) but no difference between the DIST violation condition and the controls. A second GLMER analysis on *yes* vs. *no* responses contrasted the DIST violation condition with the control conditions. This contrast revealed a significant DIST effect in the target condition relative to the true control condition ($\hat{\beta} = 6.72; SE = 1.41; z = 4.78; p < .01$). Taken together, the size of the DIST violation effect turns out to be comparable with existing research (Crnič et al., 2015; Ramotowska et al., 2022). DIST inferences thus seem to be largely unaffected by the embedding in a question environment, displaying robustness similar to NE-SCOPE in Exp. 1.

The mean answer RTs of legitimate responses are plotted in Figure 3. As in Exp. 1, the empty restrictor condition turned out to be answered faster than all other conditions. This is again

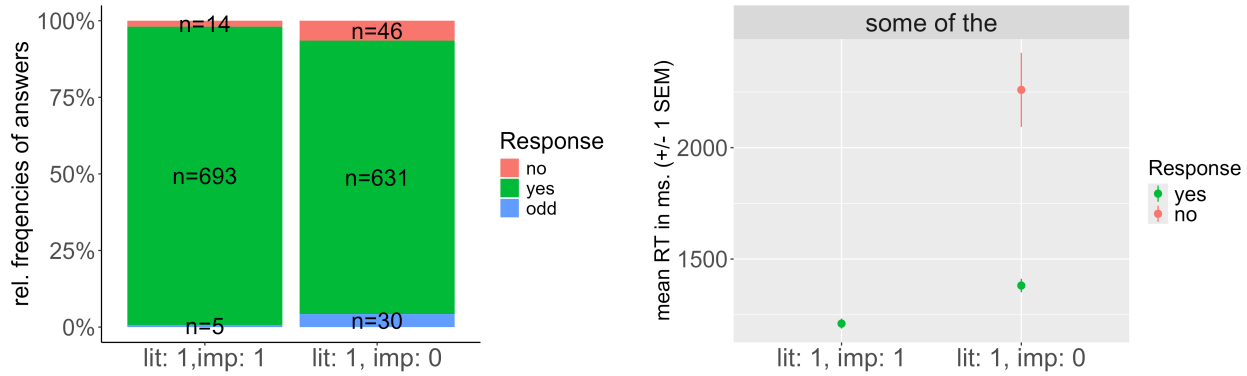


Figure 4: Distribution of answers and RTs of legitimate responses in Exp. 3. Note: lit: 1/0 = literal meaning is true/false; imp: 1/0 = UB inference is true/false.

in accordance with a presuppositional analysis under the assumption that the presupposition check is a first prerequisite before actual truth evaluation can proceed. For the analysis of the DIST violation condition, an LMER analysis was conducted including the fixed effects of PICTURE (contrast 1: DIST vs. true control, contrast 2: DIST vs. false control) and ANSWER POLARITY (positive vs. negative). The analysis revealed two main effects in the absence of interactions. Answers in the DIST violation condition took longer than in the true control condition ($\hat{\beta} = 224.6; SE = 37.6; t = 5.98; p < .01$) suggesting that the suppression of the DIST inference was in fact costly. However, in contrast to the NE-SCOPE effect in Exp. 1, there was also an answer polarity effect which was due to the fact that logically correct '*yes*' responses were generally faster than '*no*' responses indexing the pragmatic violation ($\hat{\beta} = -219.7; SE = 49.5; t = -4.44; p < .01$). On first sight, this corresponds to the pattern predicted by Negate Alternative accounts. However, since *every* is an UE quantifier in its right argument, we may just see the usual polarity effect observed for positive vs. negative answers in UE environments. We will come back to this point in a cross-experimental comparison in Section 2.2.4 below.

2.2.3. Exp. 3 – UB inferences for *some of the*

Experiment 3 served as a control study testing for scalar UB effects in terms of their robustness as well as with regard to RT differences between logical and pragmatic responses in the present paradigm. We therefore compared a condition in which only the literal reading was true to a condition in which both the literal reading and the UB inference were true. Figure 4 shows the distribution of answers and their RT as a function of answer polarity in the critical condition.

The analysis of acceptance rates confirmed a small but reliable UB effect, that is less acceptance in pictures in which the implicature was false than in pictures in which it was true (97.3 vs. 89.3% *yes*; $\hat{\beta} = -1.53, SE = 0.34, z = -4.51, p < .01$). Crucially, this UB effect was much weaker than the effects reported in the literature on unembedded scalars (with only ≈ 60% acceptance). The embedded UB effect thus turned out to be less less robust than the other effects reported in the other experiments on NE-RESTRICTOR, NE-SCOPE and DIST above. However, the implicature rates fit well with other studies testing embedded implicatures.

The analysis of RTs replicated the RT difference between logical and pragmatic responses in the critical condition with a difference in RT of 879 ms ('*yes*': 1381 ms vs. '*no*': 2260 ms). A

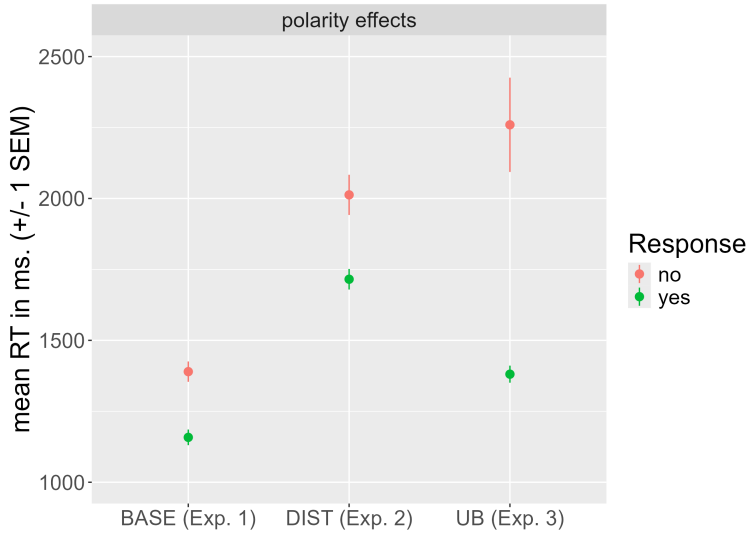


Figure 5: Answer polarity effects in the critical conditions of the DIST (Exp. 2) and the UB experiments (Exp. 3) in comparison with the control conditions for *every* in Exp. 1 (BASE).

planned pairwise comparison in this condition comparing an LMER model with the fixed effect of ANSWER POLARITY to one without in a likelihood ratio test revealed that this difference was significant ($\chi^2(1) = 5.75, p < .05$). Superficially, the RT effect for scalars was similar to the one observed for DIST in terms of their direction. However, they greatly differed in the absolute size of the RT differences. To do a direct comparison, we compared polarity effects in the target conditions of the DIST and UB experiments, the two experiments with UE quantifiers.

2.2.4. A Cross-Experimental Comparison of Polarity Effects

In the cross-experimental comparison RTs of '*yes*' and '*no*' responses in the target conditions in the DIST (Exp. 2) and the UB (Exp. 3) experiment were compared with respect to the effect sizes of the answer polarity effects. Since both involve UE quantifiers, '*yes*' responses should be faster than '*no*' responses independently of DIST or UB effects. As an estimate of polarity effects in UE quantifiers we therefore also included the control conditions for the Aristotelean UE quantifier *every* from Exp. 1, that is RTs of '*yes*' responses in the true control condition and RTs of '*no*' responses in the false control condition. Note, that this allows for a close comparison of sentences with universal quantifiers in a first contrast comparing RTs of legitimate responses to controls and DIST answers and a second contrast comparing RTs of controls with RTs of the UB target conditional also involving an Aristotetelean quantifier.

Mean RTs in the cross-experimental comparison are shown in Figure 5. The relative difference between '*yes*' and '*no*' answers was comparable in the comparison between baseline control and DIST responses, but much larger in the comparison involving UB inferences. This was corroborated in the LMER analysis. The analysis gave rise to an interaction between the contrast BASE/UB and POLARITY ($\hat{\beta} = 436.9, SE = 130.8, t = 3.34, p < .01$) but not for the contrast BASE/DIST and POLARITY ($|t| \leq 1$). Thus, in the DIST experiment an answer polarity was observed that was well in the range of general answer polarity effects observed also in the control

conditions. The UB effect, however, turned out to be substantially larger than the latter polarity effects supporting our plea for different mechanisms at work in DIST and UB inferences.

3. General Discussion

We presented experimental evidence for diversity between a number of phenomena that have received related explanations based on the negation of pragmatic alternatives of the expressions used. The experiments showed very clear interpretation differences between quantificational cases involving quantifiers with empty restrictor sets and others involving verification of quantifiers in situations with empty scope sets.

Questions about empty restrictor cases were uniformly perceived as odd no matter what kind of quantifier was involved. Moreover, rejection of questions about empty restrictors proceeded very quickly, as would be expected in a presuppositional framework in which a NE-RESTRICTOR presupposition first has to be checked and justified before the asserted content can be evaluated. Of course, this kind of processing behaviour was also supported by our experimental setup – as was the construal of all quantifiers as strong – which demanded participant to first identify the relevant set in the left- vs. the right-handside of the pictures presented to them by identifying the relevant shape. Empty-scope configurations were very well set apart from the NE-RESTRICTOR cases. Questions about empty-scope configurations with the German counterparts of the empty-set quantifiers *less than three* and *no* gave rise to empty-set effects of only about 35% and 15% non-logical responses. This is well in the range of earlier findings reported for declaratives showing the robustness of these effects but also the different effect sizes observed for these particular quantifiers. The empty-scope effect was also confirmed in the RT analysis showing longer response times in these conditions relative to all other conditions in the experiment. Last but not least, no answer polarity effect was observed in the empty-scope conditions. The observed lack of any difference fits well with proposals that treat NE-SCOPE as a default inference that does not take extra time or is in any way delayed.

A similar processing profile was observed for the DIST inference. It turned out to show up robustly in question environments in the previously reported size of about 20% rejections of DIST violations. We also observed a processing penalty for these violations in line with previous research and claims about DIST as a default inference. What's more, even though an answer polarity effect was observed, it turned out to be just of the size of general answer polarity effects also observed in the control conditions. Taken together, the processing of the NE-SCOPE and the DIST inferences were well in line with the neglect-zero proposals.

The UB inferences included as a control experiment turned out to behave rather differently from all the other inferences and interpretation preferences just discussed. UB violations were overwhelmingly accepted in about 90% of the time. This inference type thus turned out to be considerably less robust (cf., among others, Potts et al., 2016; Franke et al., 2016). In addition, the RT analysis revealed an answer-polarity effect in line with the literature on the verification of scalars reviewed in the introduction. Pragmatic responses had considerably longer RTs than logical responses and this effect was observed on top of a more general answer polarity effect.

Taken together the present study thus presents evidence for at least three mechanisms at work in the four interpretation phenomena discussed in the paper. Even though the *Negate Alternatives* accounts are on first sight a very attractive one-fits-all approach, they do not explain the diversity observed when it comes to the cognitive processing of these phenomena.

References

Abusch, D. and M. Rooth (2004). Empty-domain effects for presuppositional and non-presuppositional determiners. In H. Kamp and B. H. Partee (Eds.), *Context-Dependence in the Analysis of Linguistic Meaning*, pp. 7–27. Leiden: Brill.

Aloni, M. (2022). Logic and conversation: The case of free choice. *Semantics & Pragmatics 15*(5), 1–60.

Aloni, M. and P. van Ormondt (2023). Modified numerals and split disjunction: The first-order case. *Journal of Logic, Language and Information 32*, 539–567.

Balbach, N., F. Schlotterbeck, H. Wang, and O. Bott (2022). Disentangling semantic empty-set effects in quantifier comprehension from simple associations: Processing evidence from exceptive-additives. In *Proceedings of the 23rd Amsterdam Colloquium*, Amsterdam, pp. 15–22. ILLC.

Bar-Lev, M. E. and D. Fox (2020). Free choice, simplification, and innocent inclusion. *Natural Language Semantics 28*, 175–223.

Bar-Lev, M. E. and D. Fox (2023). On fatal competition and the nature of distributive inferences. *Natural Language Semantics 31*(4), 315–348.

Bassi, I., G. Del Pinal, and U. Sauerland (2021). Presuppositional exhaustification. *Semantics & Pragmatics 14*(11), 1–48.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1), 1–48.

Bott, L., T. M. Bailey, and D. Grodner (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language 66*(1), 123–142.

Bott, L. and I. Noveck (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language 51*(3), 437–457.

Bott, O., F. Schlotterbeck, and U. Klein (2019). Empty-set effects in quantifier interpretation. *Journal of Semantics 36*(1), 99–163.

Breheny, R. (2019). Scalar implicatures. In N. Katsos and C. Cummins (Eds.), *Oxford Handbook of Experimental Semantics and Pragmatics*, pp. 39–61. Oxford: Oxford Un iversity Press.

Breheny, R., H. J. Ferguson, and N. Katsos (2012). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes 28*(4), 443–467.

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

Chierchia, G., D. Fax, and B. Spector (2011). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn, and K. von Heusinger (Eds.), *Semantics: An International Handbook of Natural Language Meaning*, pp. 297–330. Berlin: de Gruyter.

Crnič, L., E. Chemla, and D. Fox (2015). Scalar implicatures of embedded disjunction. *Natural Language Semantics 23*(4), 271–305.

de Jong, F. and H. Verkuyl (1985). Generalized quantifiers: the properness of their strength. In J. van Benthem and A. ter Meulen (Eds.), *Generalized quantifiers in natural language*, pp. 21–43. Dordrecht: Foris.

De Neys, W. and W. Schaeken (2007). When people are more logical under cognitive load: dual task impact on scalar implicature. *Experimental Psychology 54*, 128–133.

Degano, M., P. Marty, S. Ramotowska, M. Aloni, R. Breheny, J. Romoli, and Y. Sudo (2025). The ups and downs of ignorance. *Natural Language Semantics*.

Degen, J. and M. K. Tanenhaus (2015). Processing scalar implicature: a constraint-based approach. *Cognitive Science 39*(4), 667–710.

Diesing, M. (1992). *Indefinites*. Cambridge, MA: MIT Press.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland and P. Stateva (Eds.), *Presupposition and Implicature in Compositional Semantics*, pp. 71–120. London: Palgrave Macmillan.

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics & Pragmatics 4*(1), 1–82.

Franke, M., F. Schlotterbeck, and P. Augurzky (2016, 08). Embedded scalars, preferred readings and prosody: An experimental revisit. *Journal of Semantics 34*(1), 153–199.

Geurts, B. (2008). Existential import. In I. Comorovski and K. von Heusinger (Eds.), *Existence: Semantics and Syntax*, pp. 253–271. Dordrecht: Springer.

Geurts, B. (2010). *Quantity Implicatures*. Cambridge, MA: Cambridge University Press.

Geurts, B., N. Katsos, C. Cummins, J. Moons, and L. Noordman (2010). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes 25*(1), 130–148.

Geurts, B. and N. Pouscoulous (2009). Embedded implicatures?!? *Semantics & Pragmatics 2*, 1–34.

Grice, P. H. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and semantics 3: Speech acts*, pp. 41–58. Leiden: Brill.

Grodner, D. J., N. M. Klein, K. M. Carbary, and M. K. Tanenhaus (2010). "some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition 116*, 42–55.

Huang, Y. T. and J. Snedeker (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology 58*(3), 376–415.

Huang, Y. T. and J. Snedeker (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cognitive Psychology 102*, 105–126.

Johnson-Laird, P. N. (1983). *Mental Models. Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge, UK: Cambridge University Press.

Just, M. A. and P. A. Carpenter (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior 10*(3), 244–253.

Khorsheed, A., J. Price, and B. van Tiel (2022). Sources of cognitive cost in scalar implicature processing: A review. *Frontiers in Psychology 7*, 1–14.

Lappin, S. and T. Reinhart (1988). Presuppositional effects of strong determiners: A processing account. *Linguistics 26*, 1021–1037.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Mankowitz, P. (2023). Experimenting with every American king. *Natural Language Semantics 31*, 349–387.

Marty, P., J. Romoli, Y. Sudo, and R. Breheny (2024). What makes linguistic inferences robust? *Journal of Semantics 41*(1), 1–52.

McNally, L. (2019). Strong and weak nominals. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, and T. E. Zimmermann (Eds.), *The Wiley Blackwell companion to semantics*, pp. 1–24. Hoboken, NJ: John Wiley & Sons.

Nieder, A. (2016). Representing something out of nothing: The dawning of zero. *Trends in Cognitive Sciences 20*(11), 830–842.

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

Potts, C., D. Lassiter, R. Levy, and M. C. Frank (2016). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics 33*(4), 755–802.

Ramotowska, S., P. Marty, J. Romoli, Y. Sudo, and R. Breheny (2022). Diversity with universality. In *Proceedings of the 23rd Amsterdam Colloquium*, Amsterdam, pp. 251–257. ILLC.

Reinhart, T. (2004). Topics and the conceptual interface. In H. Kamp and B. H. Partee (Eds.), *Context-Dependence in the Analysis of Linguistic Meaning*, pp. 275–305. Leiden: Brill.

Sbardolini, G. (2023). Homogeneity and the illocutionary force of rejection. In *Proceedings of SALT 33*, Yale University, pp. 399–420. Linguistic Society of America.

Strawson, P. F. (1952). *Introduction to logical theory*. London: Methuen.

Strawson, P. F. (1964). Identifying reference and truth-values. *Theoria 30*(2), 96–118.

Tieu, L., J. Romoli, P. Zhou, and S. Crain (2016). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics 33*(2), 269–298.

Tomlinson Jr, J. M., T. M. Bailey, and L. 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, B., E. Van Miltenburg, N. Zevakhina, and B. Geurts (2016). Scalar diversity. *Journal of Semantics 33*(1), 137–175.