Empty restrictor and empty scope effects in questions with quantifiers

Sonia Ramotowska (sonia.ramotowska@ens.psl.eu)[1,3], Fabian Schlotterbeck[2], Tomasz Klochowicz[3], Maria Spychalska[4], Maria Aloni[3] & Oliver Bott[5]

[1]Institut Jean-Nicod, Département d'Études Cognitives, ENS-PSL, EHESS, CNRS
[2]Collaborative Research Center 1718: Common Ground, University of Tübingen
[3]Institute for Logic, Language and Computation, University of Amsterdam
[4]Institute of Philosophy II, Ruhr University Bochum
[5]Department of Linguistics, Bielefeld University

Abstract

One striking divergence between logical modeling and language processing emerges when evaluating the empty set in the interpretation of quantifiers. It has been shown that empty restrictor scenarios lead to presupposition failure, while empty scope scenarios lead to logically incorrect judgments and a processing delay of downward-entailing quantifiers. In this study, we tested whether empty restrictor scenarios would be evaluated as *true* or *false* rather than rejected as *odd* in a question-answering task. We contrasted two weak downward-entailing quantifiers (*fewer than n*, *at most n-1*) with two weak upward-entailing quantifiers (*more than m*, *at least m+1*) in two complex question embeddings that made the evaluation possible to various degrees. We found that in these embeddings, empty restrictors are less often rejected as odd as well as that the empty restrictors of downward-entailing quantifiers are often evaluated as *false*.

Keywords: quantifiers; empty set; scope; restrictor; presuppositions

Introduction

In quantified sentences, such as "Every square is white", the restrictor restricts the domain of quantification (here to squares), and the scope specifies the properties of the restrictor. Psycholinguistic studies (Knowlton et al., 2023) showed that the restrictor and scope of the universal quantifier are processed differently. Several studies have observed timing differences between the processing of the two arguments, with a highly incremental processing of the restrictor argument (Augurzky et al., 2017, i.a.) and (sometimes) delayed interpretation of quantifier scope (Bott & Schlotterbeck, 2015; Schlotterbeck & Bott, 2026; Urbach & Kutas, 2010, i.a.).

Moreover, quantifiers generally require that their restrictor and scope are not empty. Consider the sentence in examples (1) and (2) as description of scenarios of 1) **only triangles**; 2) **only black squares**. According to the truth conditions of *fewer than three*, the sentence is true in both scenarios: when there are no squares or when there are squares but none of them are white. Intuitively, however, using this sentence seems to be misleading because of the NON-EMPTY RESTRICTOR (there are some squares) and the NON-EMPTY SCOPE (some of those squares are white) inferences (1) and (2), which have been attested in a few experimental studies (Balbach et al., 2022; Bott et al., 2019, 2025).

(1) Fewer than three squares are white.
$\rightsquigarrow$ There are some squares. NON-EMPTY RESTRICTOR

(2) Fewer than three squares are white.
$\rightsquigarrow$ There are some white squares. NON-EMPTY SCOPE

The source of the inferences

Some linguistic theories claim that the inference in (1) is a presupposition (Bott et al., 2019; de Jong & Verkuyl, 1985). In particular, strong quantifiers, such as *every* or *most*, carry the presupposition of existential import (de Jong & Verkuyl, 1985; Geurts, 2008), which makes them infelicitous when they quantify over an empty domain. However, weak quantifiers, such as *some, no, fewer than n, more than m*, might be considered felicitous when the domain is empty and judged true or false (Lappin & Reinhart, 1988). Other theories proposed that the inference in (1) is an implicature (Abusch & Rooth, 2004).

Turning to the inference in (2), two non-implicature theories explain how this inference comes about.[1] According to the neglect zero account (Aloni & van Ormondt, 2023), (2) is due to a cognitive bias called neglect zero (as the (1) inference). Aloni (2022) claims that when verifying sentences, people tend to systematically neglect the verifiers that satisfy sentences by virtue of an empty set. For the sentence "Fewer than three squares are white" such verifiers are the abovementioned scenarios 1) and 2). The neglect zero tendency leads to difficulties in processing these scenarios as verifiers and logical mistakes in the verification or prolonged reaction times. The neglect zero account models this tendency with a special pragmatic enrichment function defined in a bilateral state-based modal logic (Aloni, 2022). In this framework, the quantifiers definable in first-order logic, such as *fewer than n*, *at most n-1*, *more than m*, *at least m+1*, are modeled with first-order logic formulas, lacking any encoding of the linguistic distinction between restrictor and scope. Thus, this account predicts symmetric effects for the inferences (1) and (2).

In contrast, the W-quantifier theory (Bott et al., 2019) explains the (1) inference as presuppositional, while the inference in (2) is taken to be an error that arises due to the complexity of the specific verification procedures required in these cases. They call the quantifiers (e.g., downward-entailing quantifiers, such as *fewer than three*), which have an

[1]For implicature of modified numerals see e.g., Cummins (2013) and Mayr (2013). For experimental research comparing scalar implicature interpretation to the processing of the inferences in (1) and (2) see Bott et al. (2025) and Klochowicz et al. (2025).

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

empty set as a verifier, empty-set quantifiers. Empirically, the W-quantifier account predicts that the scenarios in which the restrictor of the quantifier is empty should be processed fast, as presupposition failures, and uniformly across quantifiers, while the scenarios with empty scope require extra processing efforts in the case of empty set quantifiers.

To test these predictions against predictions from implicature-based accounts, Bott et al. (2025) embedded the quantifier sentences in questions and used a verification task with three response options: *no* (rejection of the scenario), *yes* (acceptance of the scenario), and an *odd question* (indicator of presupposition violation). They showed that both strong and weak quantifiers uniformly give rise to presupposition failure (but see Mankowitz, 2023). More specifically, questions were overwhelmingly rejected as odd in response to scenarios with empty restrictor sets instead of providing a response based on the semantics of the quantifiers.

Moreover, Bott et al. (2025) identified differences between the inferences in (1) and (2) in terms of their robustness and processing profile (i.e., judgments and RT). They found that empty-scope scenarios were treated as false in about 25% of the cases. Moreover, regardless of being rejected or accepted, the processing time for verification of these scenarios was longer than for control scenarios. In contrast, the empty restrictor scenarios were verified quickly and associated with *odd question* reactions. This finding indicated an asymmetry between the processing of empty-scope and empty-restrictor, not predicted by the neglect-zero account (Aloni, 2022), and that both inferences likely have a different mechanism of computation than scalar implicatures (cf. Klochowicz et al., 2025).

Assuming that the inference in (1) is a presupposition, one open question is whether, under certain circumstances, weak quantifiers can be evaluated even though the presupposition of a non-empty restrictor is violated. The literature on non-referential expressions (e.g., "the present king of France") offers some testable hypotheses (Abrusán & Szendroi, 2013).

Presuppositions and reference failure

Strawson (1964) observed that reference failure is more acceptable in cases where it does not concern the topic (subject) of a sentence. While the sentence "The king of France is bald" leads to the feeling of "squeamishness", the sentence "Yesterday's concert was attended by the king of France" feels false. This is because in the former sentence, "the king of France" is the likely topic, while in the latter, "the concert" is. In addition to whether the non-referential expression is the sentence topic, the verifiability of the sentence reduces the "squeamishness" (Lasersohn, 1993; von Fintel, 2004). Thus, even though the sentence "The king of France attended the concert yesterday at the Palais Garnier" has "the king of France" as the topic, it is more likely to be judged as false because one can verify who was among the attendees of that concert. These hypotheses were empirically confirmed by Abrusán and Szendroi (2013), who found that sentences are likely to be evaluated as *true* or *false* when the non-referring NP is not topicalized than when it is topicalized. Moreover, the availability of a referring expression as an anchor for verification facilitated the evaluation of a non-referring expression. This verifiability hypothesis is also compatible with the results by Mankowitz (2023), who showed that sentences with universal quantifier *every* are judged as false when the domain is empty (e.g., "Every American king lives in New York"). Although "American king" is a non-referential expression, one could verify the sentence by checking that among the inhabitants of New York, there is no American king.

The current study

Following the previous literature, we predicted that the empty restrictor scenarios would lead to a presupposition failure. In this study, we tested whether, despite this failure, weak quantifiers could be evaluated as *true* or *false* as in the case of non-referring NPs. Furthermore, we explore the asymmetry between empty-scope and restrictor processing. To ensure comparability and replicability of results, we also employed the question-answering task employed in Bott et al. (2025).

Firstly, we used a complex question (3), which allowed us to test to what extent participants reject questions as *odd* in situations where the existential presupposition is known to be false. Bott et al. (2025) tested direct polar questions, such as "Are fewer than three squares blue?", where the restrictor was the topic of the question. We used constructions, which make the card the topic of the question and hence should promote evaluation despite the existence presupposition not being satisfied.

(3) Do you have a card
   a. on which fewer than three squares are white?
   b. with fewer than three white squares?

For instance, participants could also reason about different possible cards (in an imagined deck of cards) that satisfy the inferences (1) and (2). The complex question should thus further allow us to distinguish the potentially independent existence presupposition (that there is such a card) from the empty restrictor effect. Hence, in both configurations we expect a lower level of rejection of the questions, i.e., a smaller amount of *odd question* responses, compared to Bott et al. (2025). In the case of downward-entailing quantifiers, the *no* response would be compatible with the inferences in (1) and (2), and the response *yes* would be a correct response with respect to truth conditions. **The neglect-zero theory predicts a mixture of responses in both experiments**.

Moreover, the configurations in Experiment 1 vs. 2 allow us to test whether the observed processing asymmetry between empty restrictor and empty scope is, in fact, **due to a semantic/pragmatic asymmetry between quantifier arguments**. In Experiment 1, the quantifier was part of a subordinate clause, but the internal structure of the clause remained the same as in the study by Bott et al. (2025) with the restrictor as part of the DP and the scope in the embedded VP. In Experiment 2, the quantifier was used as an NP (3b) in

the non-topical clause as in the study by Abrusán and Szendroi (2013), and both noun and adjective were parts of the restrictor argument. Thus, if the asymmetry between scope and restrictor arguments affects the rates of *odd* response, we would observe this asymmetry in Experiment 1 for the empty restrictor and scope, and symmetry in Experiment 2. Moreover, if topicality affects the response rates, we predict that the *odd* response rates should be lower in the NP structure in Experiment 2 than in the VP structure in Experiment 1.

Finally, we tested whether the empty-scope effects found by Bott et al. (2025) for comparative quantifiers (e.g., *fewer than three*) would generalize to superlative quantifiers (e.g., *at most two*), which have a different, disjunctive structure (Büring, 2008, i.a.) and are known to be more difficult to process and acquired later (Geurts et al., 2010). Accordingly, we predicted **stronger empty scope effects, that is, more errors, for downward entailing superlative quantifiers than for comparative ones**. In particular, this should affect the error rates of empty-scope conditions, as suggested by the results on doubly quantified sentences in Bott et al. (2019, Exp. 3). However, a difference in empty-scope effects for superlative vs. comparative quantifiers has not yet been tested in simple quantified sentences.

Methods

Participants

80 participants were recruited on Prolific for each experiment. All participants self-reported Dutch as their first language and reported not having problems seeing colors. They were between 18 and 44 years of age and had at least 90 approval rate on Prolific. Participants were paid with a rate of 9£ per hour. Median times of the experimental session were 22 min (Experiment 1) and 20 min (Experiment 2). We used an average of 75% accuracy threshold in the control conditions to exclude participants. We also excluded participants with very long duration of the testing session. Together, 5 participants were excluded from the first experiment and 8 from the second, leaving 75 and 72 participants in the analysis, respectively.

Materials and Design

All experimental materials were constructed in Dutch[2]. The design was factorial: QUANTIFIER TYPE (superlative vs. comparative) × MONOTONICITY (upward vs. downward) × SCENARIO (empty restrictor, empty scope[3], false and true control). This resulted in 16 conditions. Each participant received 6 trials in each condition (in total 96 trials). We created 24 question items distributed in four lists in a Latin Square design. Each item was paired with 4 different pictures. A picture-item pair occurred only once in each list.

The same materials were used as in both experiments, except for the form of the questions. Each trial involved quantified questions: Experiment 1 "Do you have a card on which QUANTIFIER SHAPES are COLOR?"; Experiment 2 "Do you have a card with QUANTIFIER COLOR SHAPES?". The tested quantifiers were *more than m, fewer than n, at least m+1*, and *at most n-1*. The comparative and superlative quantifiers were chosen to be truth-conditionally equivalent rather than referring to the same numeral. Five shapes (squares, circles, triangles, hearts, and crosses) and 11 colors (red, blue, yellow, green, orange, purple, turquoise, brown, pink, black, grey) were paired randomly.

Each image, always presented following the question presentation, depicted objects both on the left-hand and on the right-hand side of the depicted card. In the empty-scope and the control conditions, the target shape mentioned in the question was depicted on one side and a filler shape different from the target shape on the other side of the card (counterbalanced across items). In the empty-restrictor condition, both sides depicted filler shapes of different types than the one mentioned in the question. The shapes were depicted in two different colors: the target color mentioned in the question and a filler color. In the empty-scope condition, all target shapes were of the filler color. The filler shapes always contained objects of both colors, thus even in the empty-scope condition, the target color was present on the display.

Procedure

The procedure in both experiments was the same as in Bott et al. (2025). Upon giving their consent for participation, participants received written instructions and completed a short training block of 9 trials (3 questions per each response type). The questions and pictures used in the training were not used in the remaining part of the experiment. No feedback was provided. The main part of the experiment was split into four blocks. The trials were presented in random order.

Each trial consisted of two screens. On the first screen, participants read the question (self-paced). On the second screen, they saw a picture together with Dutch response options corresponding to "yes, true", "no, false" and "odd question". The response keys were arrow keys: down, right, and left. Key-response mappings were randomized between participants. At the end of the experiment, participants could complete an optional demographic questionnaire.

Results

To compare the rates of each response type ("no", "yes", "odd"), we ran a mixed-effects multinomial regression using the MCLOGIT R package (Elff et al., 2022) with three level dependent variable. In all statistical models, the "no" response was set as the reference level, meaning that the model compared "no vs. odd" and "yes vs. no" response rates[4]. The

[2]See stimuli in materials folder at OSF https://osf.io/euwgb. 88 filler questions involving: definite description, quantifiers *both* and *some*, and *maximum* expression were added to the experiments.

[3]For simplicity, we stick to the labels empty-restrictor and empty-scope conditions even though this is strictly speaking not appropriate for the construction used in Experiment 2.

[4]The data and analysis scripts are available at OSF. For an explanation of the random-effect structure, we refer therein.

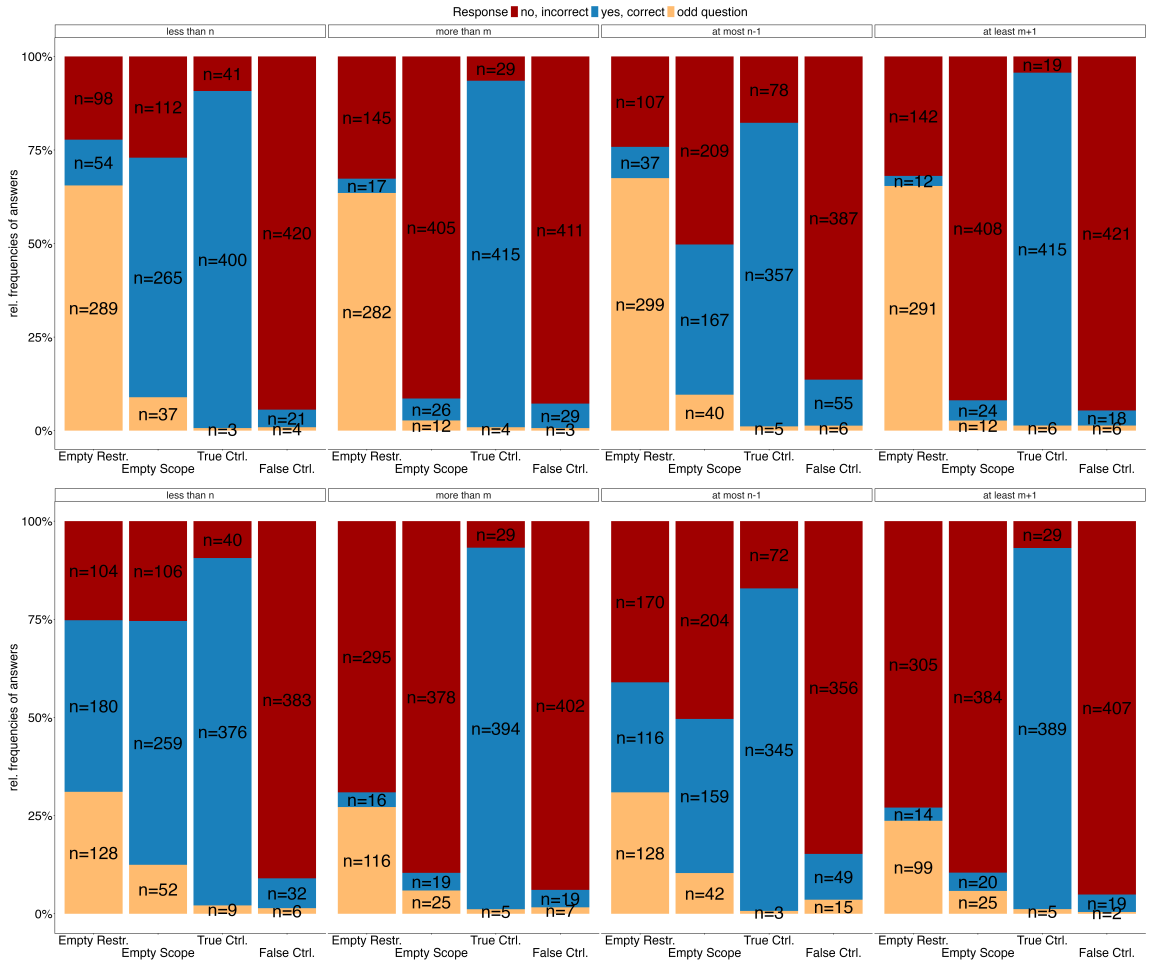

Figure 1: Distribution of responses to questions in Experiments 1 (top) and 2 (bottom).

binary predictors (MONOTONICITY, QUANTIFIER TYPE, EXPERIMENT VERSION) were contrast coded and the SCENARIO predictor was dummy coded with the reference level set to *empty scope*. We ran a statistical model for each experiment separately, which included all SCENARIOS, MONOTONICITY, QUANTIFIER TYPE, and their interaction as predictors; and one model to compare the response types between EXPERIMENT VERSIONS constrained to only *empty set* SCENARIOS (predictors: SCENARIO, MONOTONICITY, QUANTIFIER TYPE, EXPERIMENT VERSION, and their interactions) to test specifically whether empty scope and restrictors are evaluated differently.

Concerning reaction times (RT) in Experiment 1, we tested whether we replicate the delay in processing the *empty scope* scenario relative to the *control* scenarios in the case of downward entailing quantifiers reported by Bott et al. (2025). Only logically correct responses entered the analysis; however, for the *empty scope* scenario for downward-entailing quantifiers, we also included incorrect "no" responses. A linear mixed-effects model predicted logarithmically transformed RTs as a function of MONOTONICITY, QUANTIFIER TYPE, SCENARIO, and their interactions. The SCENARIO was encoded using Helmert's contrast (contrast-1: *empty scope* vs. *control* SCENARIOS; contrast-2: *false* vs. *true* control SCENARIOS). In additional analysis of downward entailing quantifiers, we tested whether acceptance and rejection of the *empty scope* scenario are equally fast relative to the *control* conditions. For Experiment 2, we did not plan any RT analysis because the choice of comparison would depend on the response rates in the *empty scope* scenario.

Experiment 1

In Experiment 1, in the empty-restrictor condition, the dominant response type for all quantifiers was "odd question" (average of 65%, see Figure 1). In contrast to Bott et al. (2025), we observed, however, that "no" responses were also frequent (average of 28%). The least frequent were "yes" responses. Moreover, there was a substantial proportion of logically incorrect "no" responses in the empty scope condition for downward entailing quantifiers (comparative: 27% and superlative: 50%). We found a significant three-way interaction between MONOTONICITY × QUANTIFIER TYPE × SCENARIO (*empty scope* vs. *false* control; $\beta = -0.65$; $z = -4.95$; $p < 0.0001$) only for

the "yes vs. no" response contrast. As expected, for upward monotone quantifiers, there were no significant differences of "yes" response compared to "no" between the *empty scope* and the *false* SCENARIO ($\beta = -0.13; z = -0.57; p = 0.57$), and no significant interaction with QUANTIFIER TYPE (all $p_s > 0.2$).

For downward monotone quantifiers ("yes vs. no"), we found a significant effect of QUANTIFIER TYPE ($\beta = 0.59; z = 7.3; p < 0.0001$) and significant interactions between QUANTIFIER TYPE and SCENARIOS (*empty scope* vs. *false* control: $\beta = -1.14; z = -7.22; p < 0.001$; *empty scope* vs. *empty restrictor*: $\beta = -0.35; z = -2.25; p = 0.02$). For both types of quantifiers, we replicated higher rates of logically incorrect rejections of *empty scope* scenarios compared to *true* control scenarios (all $\beta_s > 1.6; p_s < 0.001$). In fact, for superlative quantifiers, rejections of *empty scope* scenarios were not significantly less likely than acceptance ($\beta = -0.24; z = -1.68; p = 0.09$). In addition, *empty scope* scenarios were more often accepted than *empty restrictor* scenarios (both quantifier types: all $\beta < -0.8; p < 0.001$).

Concerning the ratios of "odd" to "no" responses, for upward entailing quantifiers, we found that "odd" responses were more frequent for *empty restrictor* scenarios than for *empty scope* scenarios ($\beta = 4.6; z = 13.0; p < 0.0001$) and more frequent for the *empty scope* than for the *false* scenarios ($\beta = -1.25; z = -2.99; p = 0.003$)[5]. For downward entailing quantifiers, we also found that the rate of "odd" response was higher in the case of *empty restrictor* than in the case of the *empty scope* scenarios ($\beta = 3.37; z = 17.82; p < 0.0001$) and was higher in the case of the *empty scope* than in *true* control scenarios ($\beta = -1.41; z = -3.47; p = 0.001$). Taken together, the distribution of responses to a large extend replicated the pattern found by Bott et al. (2025): That is, higher rates of "odd" question response choices for all quantifiers in the case of *empty restrictors* (however, not at ceiling), and a higher proportion of logically incorrect "no" responses in the case of the *empty scope* scenarios as a verifier of downward entailing quantifiers. In addition to their findings, we observed a significantly larger *empty scope* effect for superlative *at most n-1* relative to comparative *fewer than n*, which is in line with considerations on their semantic complexity (Geurts et al., 2010) and generalizes variability observed for empty-scope effects from doubly to simply quantified sentences (cf. Bott et al., 2019, Exp. 3).

Reaction times We found a significant MONOTONICITY × QUANTIFIER TYPE × SCENARIO contrast-1 (*empty scope* vs. *control*, $\beta = -0.02, t = -2.23, p = 0.03$) interaction. For both types of quantifiers there was a significant MONOTONICITY × SCENARIO contrast-1 ($\beta_s > 0.06, p_s < 0.001$) interaction, where for *fewer than n* and *at most n-1* the responses were significantly longer for *empty scope* SCENARIO than *control* scenarios (comparatives: $\beta = 0.21, t = 8.58, p < 0.001$; superlatives: $\beta = 0.18, t = 6.61, p < 0.001$). The additional analysis demonstrated that the response polarity effect ("yes" responses were slower than "no": $\beta = 0.09, t = 4.71, p < 0.001$) did not differ between scenarios ($\beta = -0.004, t = -1.13, p = 0.26$).

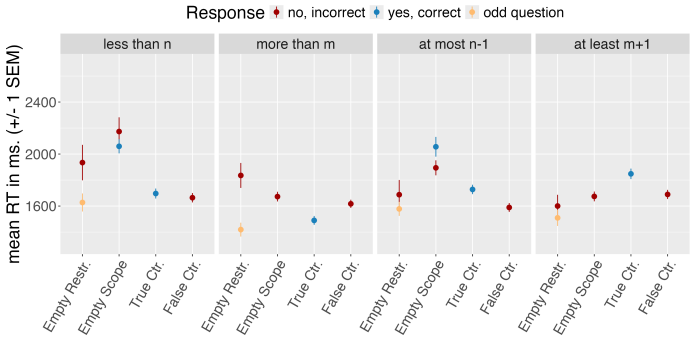

Figure 2: Reaction times of legit responses in Experiment 1.

These results replicate the empty-scope effect found by Bott et al. (2025); for downward entailing quantifiers, the mean logical "yes" responses in the empty scope condition were slower than "yes" responses in the true control condition (2058 ms vs. 1711 ms), this difference was of 363 ms for comparative quantifiers and 328 ms for superlative quantifiers.

As for the RTs in the *empty restrictor* conditions, consistently with findings by Bott et al. (2025), we observed that the "odd" responses were the fastest (mean of 1526 ms), the "no" responses were slower (mean of 1740 ms), and the "yes" responses were the slowest (mean of 2015 ms). An exploratory analysis of RTs associated with "odd" and "no" responses revealed that this difference was statistically significant for all quantifiers ($\beta = 0.07, t = 4.52, p < 0.001$), but larger for comparative quantifiers ($\beta = 0.02, t = 2.41, p = 0.02$)[6].

Experiment 2

In contrast to the previous experiment, "odd question" responses were not the dominant response for the *empty restrictor* SCENARIOS (see Figure 1, bottom). For upward entailing quantifiers, the most frequently chosen (70%) response was "no", and "odd" was only chosen 25% of the time. For the downward-entailing quantifier, the "odd" response was chosen with a 31% rate. For superlatives, the most frequently chosen response was "no" (41%) and for comparative, "yes" (44%). For the *empty scope* SCENARIO, in the case of upward entailing quantifiers, the dominant response was "no" and for downward entailing quantifiers, the dominant response was "no" for *at most n-1* (50%) and "yes" for *fewer than n* (62%).

Turning to the statistical model, for both response type contrasts, we found a significant three-way interaction between MONOTONICITY × QUANTIFIER TYPE × SCENARIO *empty scope* vs. *false* ("no vs. yes": $\beta = -0.47; z = -3.36; p = 0.0008$; "no vs. odd": $\beta = -0.77; z = -3.0; p = 0.003$). As in

[5] See explanatory notes in the analysis script at the OSF repository.

[6] RT as dependent variable and MONOTONICITY, QUANTIFIER TYPE, ANSWER TYPE ("odd" vs. "no") as sum-coded predictors. "Yes" responses were not analyzed because they did not constitute a sufficiently large portion of data.

Experiment 1, for upward entailing quantifiers, we found that the rates of "yes" responses compared to "no" responses did not differ between *empty scope* and *false* scenarios ($\beta = -0.09; z = -0.36; p = 0.72$). For downward monotone quantifiers, in the comparison between "yes" and "no" responses, we found significant interactions between QUANTIFIER TYPE and SCENARIO (*empty scope* vs. *false*: $\beta = -0.89; z = -6.03; p < 0.0001$; *empty scope* vs. *true*: $\beta = -0.28; z = -2.06; p = 0.04$). For comparative downward-entailing quantifiers, we found that there were significantly fewer "yes" responses compared to "no" in the case of *empty scope* than in the case of *true* scenario ($\beta = 1.48; z = 7.09; p < 0.0001$). In the case of superlative downward-entailing quantifiers, this effect was even greater ($\beta = 2.11; z = 11.63; p < 0.0001$). In the *empty scope* condition, there were in fact significantly more rejections than acceptances ($\beta = -0.34; z = -2.04; p = 0.04$).

Turning now to the rates of the "odd" response compared to "no", for upward-entailing quantifiers, they were higher for *empty restrictor* than *empty scope* ($\beta = 1.64; z = 6.46; p < 0.0001$), and higher for *empty scope* than *false* scenarios ($\beta = -1.28; z = -2.81; p = 0.005$). For downward-entailing quantifiers, the QUANTIFIER TYPE – SCENARIO interaction between *empty scope* vs. *false* ($\beta = -0.94; z = -3.34; p = 0.0009$) was significant. For both QUANTIFIER TYPES, there were significantly more "odd" responses in the case of *empty restrictor* than *scope* (comparative: $\beta = 1.27; z = 5.30; p < 0.0001$; superlative: $\beta = 1.60; z = 6.96; p < 0.0001$) and significantly more "odd" responses in the case of *empty scope* than in the *true* scenario (comparative: $\beta = -1.01; z = -2.39; p = 0.02$; superlative: $\beta = -1.76; z = -2.83; p = 0.005$).

The results indicate that, as in Experiment 1, in Experiment 2 there were substantially more "no" responses to *empty scope* rather than in the *true* scenario for the empty set quantifiers. Moreover, we found asymmetry in choices of the "odd" responses for *empty scope* and *empty restrictor*. Thus, the pattern of results in Experiment 2 was similar to Experiment 1; however, the overall rates of inferences were different. These differences were investigated in a cross-experimental analysis.

Comparison between experiments

For a comparison between the two experiments, for both response type contrasts, we found a significant three-way interaction between MONOTONICITY, SCENARIO, and EXPERIMENT VERSION ("no vs. yes": $\beta = -0.19; z = -3.68; p = 0.002$; "no vs. odd": $\beta = -0.15; z = -2.98; p = 0.003$). For the upward entailing quantifiers, we found that there were more "odd" responses than "no" responses in Experiment 1 than in Experiment 2 for the *empty restrictor* scenario ($\beta = -1.4; z = -6.25; p < 0.001$) and reversely for the *empty scope* scenario ($\beta = 0.4; z = 3.15; p = 0.002$).

For downward entailing quantifiers, we found that the rate of "odd" responses relative to "no" decreased in Experiment 2 for the *empty restrictor* cases only ($\beta = -0.9; z = -4.68; p < 0.001$). No change in response rate was detected for the *empty scope* scenarios, neither in the rate of "odd vs. no" responses ($\beta = 0.09; z = 0.58; p = 0.56$) nor in the rate of "yes vs. no" responses ($\beta = -0.03; z = -0.22; p = 0.83$). Moreover, in the *empty restrictor* cases, there were more "yes" responses relative to "no" responses in Experiment 2 compared to Experiment 1 ($\beta = 0.5; z = 3.55; p < 0.001$).

To summarize, the change of the question form affected only the judgments of *empty restrictor* scenarios, but not the judgments about *empty scope* scenarios. In the case of *empty restrictors* the rates of "odd" question responses decreased for all quantifiers, and in the case of downward-entailing quantifiers, the rates of "yes" responses increased.

General Discussion

The results of Experiment 1 showed a similar pattern to previous findings (Balbach et al., 2022; Bott et al., 2019, 2025). We showed that scenarios with empty restrictors are mostly perceived as presupposition failure and that this effect is uniform across quantifiers. However, by changing the question structure, well in line with our predictions, less responses indicating rejection of the question due to presupposition failure were observed than in Bott et al. (2025). Nevertheless, these responses were still mostly negative evaluations in these scenarios (cf. Mankowitz, 2023). Moreover, responses indicating presupposition failure were faster than those related to evaluation. Furthermore, we found responses consistent with evaluation, even to a larger extent, in Experiment 2. This suggests that, similarly to the case of non-referential expressions (Abrusán & Szendroi, 2013), the empty restrictor cases can be verified as true or false when they are not topical. In particular, in Experiment 2, there was always a viable verification strategy. For example, when verifying "fewer than three blue squares" in the critical scenarios, participants could either verify that all blue shapes are not squares or all squares are not blue. Thus, they could consider the available information as sufficient to verify the sentence (Lasersohn, 1993).

In Experiment 1, we replicated the empty scope effect (Balbach et al., 2022; Bott et al., 2019, 2025) and showed that this effect generalizes to superlative quantifiers, where it was even increased to about 50% non-logical responses in the *at most* empty-scope condition. Concerning verification latencies, we replicated the processing cost effect related to the empty scope scenario (Bott et al., 2025). Crucially, Experiment 2 demonstrated that the effect associated with the adjective does not change when the position of the adjective is moved from the scope to the restrictor. The persistence of this asymmetry between adjectives and nouns is challenging to the neglect zero account (Aloni, 2022), which distinguishes neither restrictors from scope nor adjectives from nouns. However, this asymmetry could also be explained by properties of the pictures presented in the present study. These pictures always had two kinds of shapes split into two sides of the picture, while the colors could be presented on both sides. This could offer a likely explanation for the observed asymmetry, but we have to leave this a question for future research.

Acknowledgments

This research was funded by the "Nothing is Logical" project (the Dutch Research Council (NWO) OC grant 406.21.CTW.023) and by the CRC 1646 "Linguistic Creativity in Communication" (DFG, German Research Foundation, project number 512393437). Sonia Ramotowska, Tomasz Klochowicz, and Maria Aloni were supported by the "Nothing is Logical" project (the Dutch Research Council (NWO) OC grant 406.21.CTW.023). Sonia Ramotowska was also supported by the French National Research Agency (ANR) grants "Communicative efficiency, cognitive constraints and lexical meaning" (ANR-23-CE28-0016) and "FrontCog" (ANR-17-EURE-0017). FS received funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 538184825. OB received funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – CRC-1646, projects B01 and Ö - 512393437.

References

Abrusán, M., & Szendroi, K. (2013). Experimenting with the king of france–topics, verifiability and definite descriptions. *Semantics and Pragmatics*, *6*(10), 1–43.

Abusch, D., & Rooth, M. (2004). Empty-domain effects for presuppositional and non-presuppositional determiners. *Context dependency in the analysis of linguistic meaning*.

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

Aloni, M., & van Ormondt, P. (2023). Modified numerals and split disjunction: The first-order case. *Journal of Logic, Language and Information*, *32*(4), 539–567.

Augurzky, P., Bott, O., Sternefeld, W., & Ulrich, R. (2017). Are all the triangles blue? — ERP evidence for the incremental processing of German quantifier restriction. *Language and Cognition*, *7*(9), 603–636.

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

Bott, O., & Schlotterbeck, F. (2015). The processing domain of scope interaction. *Journal of Semantics*, *32*(1), 39–92.

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

Bott, O., Schlotterbeck, F., Klochowicz, T., Ramotowska, S., & Aloni, M. (2025). Neglect-zero effects in the interpretation of quantifiers and disjunction. *Proceedings of Sinn und Bedeutung*, *29*, 214–232.

Büring, D. (2008). The least at least can do. *West Coast Conference on Formal Linguistics (WCCFL)*, *26*, 114–120.

Cummins, C. (2013). Modelling implicatures from modified numerals. *Lingua*, *132*, 103–114.

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

Elff, M., Elff, M. M., & Suggests, M. (2022). Package 'mclogit'. *R CRAN*.

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

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

Klochowicz, T., Schlotterbeck, F., Ramotowska, S., Bott, O., & Aloni, M. (2025). Neglect zero: Evidence from priming across constructions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *47*.

Knowlton, T., Pietroski, P., Williams, A., Halberda, J., & Lidz, J. (2023). Psycholinguistic evidence for restricted quantification. *Natural Language Semantics*, *31*(2), 219–251.

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

Lasersohn, P. (1993). Existence presuppositions and background knowledge. *Journal of semantics*, *10*(2), 113–122.

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

Mayr, C. (2013). Implicatures of modified numerals. *From grammar to meaning: The spontaneous logicality of language*, 139–171.

Schlotterbeck, F., & Bott, O. (2026). Scope as a source for non-incremental effects? *Language and Linguistics Compass*, *20*(3).

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

Urbach, T. P., & Kutas, M. (2010). Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, *63*(2), 158–179.

von Fintel, K. (2004). Would you believe it? the king of France is back! (Presuppositions and truth-value intuitions). In M. Reimer & A. Bezuidenhout (Eds.), *Descriptions and beyond*. Oxford University Press.