Putting plural definites into context*

Jacopo Romoli
Heinrich Heine Universität Düsseldorf

Yasutada Sudo
University College London

Petra Augurzky
Goethe-Universität Frankfurt

Marion Bonnet
Georg-August-Universität Göttingen

Richard Breheny
University College London

Alexandre Cremers
Vilniaus Universitetas

Cornelia Ebert
Goethe-Universität Frankfurt

Kurt Erbach
Goethe-Universität Frankfurt

Markus Steinbach
Georg-August-Universität Göttingen

Clemens Mayr
Georg-August-Universität Göttingen

Abstract. There is a disagreement at the empirical level about the availability of *non-maximal readings* of plural definites in semantically negative grammatical environments. Some theories of definite plurals are built on the assumption that everything else being equal, non-maximal readings are equally available in positive and negative environments, while others assume that they are unavailable or less available in negative environments. What makes it difficult to settle this debate with introspective judgments is the fact that non-maximal readings are inherently context sensitive, and positive and negative sentences are typically used in different kinds of conversational contexts, so direct comparison of different sentences is not conclusive for adjudicating between the two views. This paper reports on three experiments that tested the availability of non-maximal readings in the scope of different quantificational expressions, *every*, *no*, *not every*, and *exactly two*. Crucially, we tested how two different background contexts affect truth-value judgments, instead of a direct comparison of truth-value judgments. Our results show that context has a robust effect for definite plurals under *every* and *not every* but a much smaller effect for definite plurals under *no* and *exactly two*. We claim that these

* We would like to thank Sam Alxatib, Moysh Bar-Lev, Keny Chatain, Milica Denic, Patrick Elliot, Matt Mandelkern, Paul Marty, Paolo Santorio, Todd Snider, Benjamin Spector, Sonia Ramotowska, Magdalena Roszkowski, and audiences in Cabo, Tel Aviv, Sinn und Bedeutung 27, XPRAG, HNM2. We would also like to thank Anna Nalini Kratzin and YaGao for their help with programming the studies. The research in this paper was supported by an AHRC-DFG grant (AHRC: AH/V003526/1; DFG: EB 523/2-1, STE 958/12-1, STE 2555/3-1). Data files associated with our experiments, as well as the codes relevant to data treatment and statistical analyses that we report on are available open access on the OSF: xxxx.

findings pose different challenges for different theories of non-maximality in the current literature.

Keywords: plural definites, homogeneity, non-maximality, context sensitivity, relevance

Contents

1 Introduction **3**

2 Symmetric and asymmetric theories **7**
2.1 Symmetric approach: the trivalent theory 8
2.1.1 The theory 8
2.1.2 Predictions 10
2.2 Asymmetric theory: implicature theory 12
2.2.1 The theory 12
2.2.2 Predictions 15

3 The experiments **18**
3.1 Experiment 1 21
3.1.1 Methods 21
3.1.2 Results 23
3.1.3 Discussion 25
3.2 Experiment 2 26
3.2.1 Methods 26
3.2.2 Results 27
3.2.3 Discussion 29
3.3 Experiment 3 29
3.3.1 Methods 29
3.3.2 Results 30
3.3.3 Discussion 31

4 General discussion **32**
4.1 Main results 32
4.2 The symmetric approach 33
4.2.1 The challenges 33
4.2.2 A possible response 33
4.3 The challenges for the asymmetric approach 35
4.3.1 The challenges 35
4.3.2 A possible response 35

5 Conclusion 37

1 Introduction

Sentences containing a plural definite expression like *his presents* in (1a) exhibit two peculiar interpretive properties: *homogeneity* and *non-maximality* (see Križ 2019 for an overview). Homogeneity has to do with truth-value gaps: all individuals who together constitute the maximal plurality denoted by the definite plural either make the main predicate true or they all make it false. Specifically, while (1a) tends to be interpreted similarly to its universal counterpart in (1b), its negation in (2a) is best paraphrased by the meaning of (2b) instead of the negation of (1b), as one might expect given the meaning of (1a).

(1) a. Frank opened his presents.
b. Frank opened all of his presents.

(2) a. Frank didn't open his presents.
b. Frank didn't open any of his presents.

For different approaches to homogeneity see Schwarzschild 1994, Löbner 2000, Gajewski 2005, Križ 2016, among others.

Non-maximality, on the other hand, refers to a particular kind of context sensitivity. Depending on the conversational context, the truth-conditions of sentences containing plural definites are perceived to be not as rigid as those containing quantifiers. For example (1a), repeated in (3c), is acceptable in the context in (3b) but not in the one in (3a). The difference between the contexts is that while in the former what matters is if Frank opened any of his presents, in the latter it is if he opened every single one of them. This kind of context sensitivity is not shared by the universal (1b), repeated in (3d), which is unacceptable in both contexts.

(3) a. CONTEXT 1: Frank's birthday cake will not be served until he has opened all of his ten presents. His father observes that Frank opened eight of the presents and says to Frank's mother who is waiting to serve the cake:
b. CONTEXT 2: The guests at Frank's birthday party are allowed to start eating their piece of cake once Frank has started opening his ten presents. His father observes that Frank has already opened eight of the presents and no one is eating yet. So he says to the guests:
c. Frank opened his presents. Go ahead. ✗in (3a), ✓in (3b)
d. Frank opened all his presents. Go ahead. ✗(3a), ✗(3b)

For further discussion of this phenomenon and different approaches along the general lines just sketched see Krifka 1996, Lasersohn 1999, Malamud 2012, Križ 2016, Križ & Spector 2021, Bar-Lev 2018, Bar-Lev 2021, among others.

In the current literature there is a disagreement at the empirical level about the availability of non-maximal readings for definite plurals in semantically negative grammatical environments like (2a), and different theories are built upon different assumptions about it. Some theories (Križ 2015, 2016, Križ & Spector 2021, among others) assume that everything else being equal, the availability of non-maximal readings should not be affected by the polarity of the grammatical environment the definite plural occurs in, so non-maximal readings should be in principle equally available for (1a) and (2a). We call such theories *symmetric theories*. By contrast, others (Bar-Lev 2018, Bar-Lev 2021, among others) stand on the premise that non-maximal readings are unavailable, or at least much less available, when definite plurals occur in semantically negative grammatical environments. We call these theories *asymmetric theories*.

A terminological remark is in order: We call these two sets of theories *symmetric* and *asymmetric*, based on their predictions for simple positive and simple negative sentences. As we will discuss in detail later on, however, their predictions are not simply symmetric or asymmetric when it comes to plural definites occurring in non-monotonic grammatical environments.

One major obstacle in settling this debate is the fact that the availability of non-maximal readings is intrinsically context dependent, and that positive and negative sentences are typically used in different conversational contexts (see Horn 1989, Tian & Breheny 2019, for example). As a consequence, it is difficult to asses the prediction of symmetric theories by directly comparing positive and negative sentences, because any difference that might be found in interpretation could potentially be attributed to these sentences favouring different conversational contexts, rather than a semantic difference with respect to the availability of non-maximal readings. To see this problem more concretely, consider the following pair of sentences.

(4) a. Frank skipped our lectures.
    b. Frank didn't attend our lectures.

These sentences are truth-conditionally equivalent, but nonetheless, intuitively, they cannot be used entirely interchangeably, and in certain contexts one is more preferable than the other. The problem is exacerbated for sentence pairs that are not even truth-conditionally equivalent, like (1a) and (2a), as it would be even easier for such pairs to significantly differ with respect to typical conversational contexts in which they are uttered.

These concerns apply to introspective judgments reported in the literature, as well as to previous experimental studies that tested the availability of non-maximal readings in positive and negative sentences (Križ & Chemla 2015, Tieu, Križ & Chemla 2019). Križ & Chemla 2015 collected truth-value judgments of sentences analogous to (1a) and (2b) against a 'non-maximal scenario' in which Frank opened some but not all of his presents, as well as their quantified counterparts like (5), against a non-maximal scenario in which some of the boys opened some but not all of their presents (while the others opened all of their presents for (5a) and none of them for (5b)).[1]

(5) a. Every boy opened his presents.
    b. No boy opened his presents.

The results indicate that all the sentences they tested receive more intermediate judgments than the control sentences, which do not license non-maximal readings. This means that non-maximal readings are available for all the target sentences. In addition, they observe more intermediate judgments for definite plurals in positive sentences like (1a) and (5a) than for those in negative sentences like (2a) and (5b). This *prima facie* seems more in line with the asymmetric theories, but it is also amenable to an alternative explanation that is consistent with the symmetric theories. That is, since the experiments reported in Križ & Chemla 2015 did not specify conversational contexts against which the target sentences are to be evaluated, beyond the visual scenes depicting how many presents are open and how many are closed, it is possible that participants had to imagine what the conversational context is meant to be like for each trial. As remarked above, positive and negative sentences are typically used in different kinds of conversational contexts, so the participants might have evoked different kinds of contexts, which in turn could have

1 The actual sentences they tested were sentences of the following kind.

(i) a. The triangles are blue.
    b. The triangles are not blue.

(ii) a. In all the cells, the symbols are blue.
    b. In none of the cells, the symbols are blue.

(iii) a. Every boy found his presents.
    b. No boy found his presents.

In addition, they also tested definite plurals under a non-monotonic quantifier, *exactly two*, as in (iv).

(iv) a. In exactly 2 of the 4 cells, the triangles are blue.
    b. Exactly 2 of the 4 boys found their presents.

We will discuss non-monotonic quantifiers shortly.

affected the availability of the non-maximal reading. According to this explanation the asymmetry is driven essentially pragmatically, and as such has little implication on the underlying semantic (a)symmetry.

Similarly, Tieu, Križ & Chemla 2019 tested children and adults' truth-value judgments of plural definites occurring in simple positive and negative sentences analogous to (1a) and (2a) against non-maximal scenarios.[2] In their results, child participants accepted positive sentences more often than negative sentences, suggesting that non-maximal readings are easier for positive sentences than for negative ones, but adult participants accepted positive sentences more often than negative sentences, suggesting the opposite. In other words, they observed asymmetric results for both children and adults, but adults' response pattern was different from the results of Križ & Chemla 2015, and in fact inconsistent with both symmetric and asymmetric theories.[3] Importantly, furthermore, the experiments reported in Tieu, Križ & Chemla 2019 did not provide to participants much information about the intended conversational context, making the asymmetric results amenable to a similar explanation to what we suggested above.

The upshot of the above summary of previous findings is that direct comparisons of definite plurals occurring in positive and negative sentences do not provide conclusive evidence for adjudicating between the symmetric and asymmetric theories. In this paper, we report on three experiments that, rather than directly compare truth-value judgments of different sentences in non-maximal scenarios, examined how the truth-value judgments change when the sentences are used in different conversational contexts, and compared the sentences in terms of how easily they were affected by context. We tested the four sentence types exemplified in (6).

(6) a. Every boy opened his presents.
  b. No boy opened his presents.
  c. Not every boy opened his presents.
  d. Exactly two boys opened their presents.

2 The actual sentences they tested were in French:

(i) a. Les cœurs sont rouges.
  the hearts are red
  'The hearts are red.'
 b. Les cœurs ne sont pas rouges.
  the hearts NE are not red
  'The hearts are not red.'

3 One way to make sense of the difference between the two sets of results, is the fact that Tieu, Križ & Chemla 2019 did not control for the scope of negation with respect to the plural definites, unlike Križ & Chemla 2015. A wide scope reading of the definite in the negative case could be the source of the adults' acceptance in the negative case.

The primary reason why we tested quantified sentences, rather than simple sentences like (1a) and (2a), is because the bound pronouns allow us to better control the relative scope of the definite plural and the relevant operators that control the polarity. For instance, in (2a), it is possible that the definite plural outscopes negation, in which case it is interpreted in a semantically positive context, as in (1a). The same point was made by Križ & Chemla 2015, whose experimental results from quantified and non-quantified sentences suggest that this in fact mattered.

The case of non-monotonic contexts like (6d) is particularly interesting given the consideration above about context-sensitivity, as it can be seen as semantically positive and negative at the same time. Križ & Chemla 2015 tested sentences similar to (6) (see fn. 1 for details). However, the theoretical predictions of the two approaches for sentences like (6) are rather involved and intricate, so we will discuss them when we review the details of the theoretical proposals.

For now, let us briefly summarise our main findings: The results of our three experiments indicate that the truth-value judgments of all four sentences against non-maximal scenarios change in different contexts, indicating that definite plurals may receive non-maximal readings in all of these grammatical environments. However, we also observe that truth-value judgments of different sentences reacted to context manipulation to different degrees. That is, we found a large effect of context manipulation for (6a) and (6c), but a relatively small effect for (6b) and (6d). We will claim that these observations pose different challenges for symmetric and asymmetric theories.

This paper is organised as follows. In the next section, we will review representative symmetric and asymmetric theories, especially highlighting their predictions about the sentences in (6). The following three sections present our three experiments. Experiment 1 tested (6a) and (6b), Experiment 2 (6a) and (6c), and Experiment 3 (6d). Then we turn to theoretical implications of the experimental results in Section 4 and conclude in Section 5.

2 Symmetric and asymmetric theories

In this section, we will review the symmetric and asymmetric theories in detail, especially with respect to their predictions for the sentences we tested in our experiments. For concreteness, we will only discuss one symmetric theory, the trivalent theory of Križ 2015 and Križ 2016 and one asymmetric theory, the implicature theory of Bar-Lev 2018 and Bar-Lev 2021.[4] It should be reminded here again that the names

4 Other versions include the presupposed implicatures account, both within the asymmetric approach (Guerrini & Wehbe 2024) and the symmetric one (Paillé 2023), as well as the cognitive bias approach by Sbardolini (2023), which is a very different take on the symmetric approach. We leave an exploration of how our results with respects to these accounts for future work.

of the approaches, symmetric and asymmetric, are based on their predictions for simple positive and negative sentences, and their predictions for non-monotonic contexts will not be simply symmetric or asymmetric, as will be explained in detail below.

2.1 Symmetric approach: the trivalent theory

2.1.1 The theory

Križ (2015, 2016) argues that homogeneity is a lexical property of plural predication, (see also the overview paper, Križ 2019). A homogeneous predicate that is not true of a plurality a receives the third value # of a if it is true of some plurality b that has parts in common with a. This lexical assumption has the effect of creating a gap between truth and falsity conditions and capturing the homogeneity pattern of sentences with plural definites. In the case of a simple positive sentence like (1a), the effect is that the sentence receives the third truth value if Frank opened some but not all of his presents.

(7) $[\![\textsf{Frank opened his presents}]\!]^w$

$$= \begin{cases} \textsc{true} & \text{if Frank opened all of his presents in } w \\ \textsc{false} & \text{if Frank opened none of his presents in } w \\ \# & \text{otherwise} \end{cases}$$

Negation just flips the truth and falsity conditions, leaving the gap intact, so (2a) is also mapped to the third truth-value if Frank opened only some of his presents.

(8) $[\![\textsf{Frank didn't open his presents}]\!]^w$

$$= \begin{cases} \textsc{true} & \text{if Frank opened none of his presents in } w \\ \textsc{false} & \text{if Frank opened all of his presents in } w \\ \# & \text{otherwise} \end{cases}$$

In order to account for non-maximal readings, Križ 2016 proposes that perceived truth-values can differ from the semantic truth-values in certain contexts. That is, when a sentence that has a (non-trivially) trivalent meaning is used against a context where it denotes #, it can be judged as effectively true—or as 'true enough' in Križ's terminology—if # is enough to achieve the immediate conversational goal.

By way of illustration, consider CONTEXT 2 in (3b), in which we are trying to answer the question of whether Frank opened any of his presents. In other words, we would like to know if the possible world we are in is one where Frank opened none of his presents or one where he opened some or all of his presents. Frank's father knows that Frank opened some but not all of the presents. Križ's theory allows him to use the sentence in (1a) to inform the interlocutors that the true answer to the question under consideration is positive, meaning Frank opened at least one of the presents, despite the fact that the semantic truth-value the sentence denotes is #. This is because, in CONTEXT 2, we do not really care about the difference between the possible worlds in which (1a) denotes TRUE and those in which it denotes #. Said differently, the question under discussion (QuD, cf. Roberts 1996, among many others) introduces a relevant partition of epistemically live worlds.[5] In some cases, the partition associated with the QuD of the context lumps the worlds where the sentence receives # with those where it receives TRUE by the truth-conditions in (7). In such cases, the proposition can be judged as effectively true, i.e., 'true enough'. This is the case in CONTEXT 2 for (1a) given its truth-conditions in (7) as the worlds where Frank opened some but not all his presents are lumped with those where he opened all of them.

Consider now CONTEXT 1 from (3a) again. Here the QuD is whether Frank opened all of his presents. In this case, a sentence like (1a) is semantically # but cannot be judged as effectively true. This is so because now the partition of the QuD lumps together the possible worlds in which the sentence denotes # and those in which it denotes FALSE. That is, the QuD lumps the worlds where Frank opened some but not all of the presents with those where he opened none.

Importantly for us, the same holds for the negative case in (2a) but in the opposite direction: in both CONTEXT 1 and CONTEXT 2, (2a) receives the semantic value #. This time, however, a judgement of 'effectively true' is predicted to be available in CONTEXT 1 but not in CONTEXT 2. Given the truth-conditions in (8), the QuD in the former context lumps worlds where the sentence receives # with those where it receives TRUE. Recall that the worlds where Frank opened some but not all presents are in the same partition class as those where he opened none of them. For CONTEXT 2 a judgement of 'effectively true' is not predicted. Since the QuD here lumps worlds where Frank opened some but not all with those where he opened all, the #-worlds share a partition class with the FALSE worlds given (8). In this sense, the theory makes symmetric predictions for positive vs. negative sentences with respect to non-maximality.

In sum, under the trivalent theory, definite plurals give rise to a non-bivalent meaning, which feeds pragmatics in such a way that non-classical truth-values can be deemed to be effectively true and effectively false, depending on the interlocutors' interests. It is important to note at this point that such a trivalent theory is not the only way of making symmetric predictions regarding non-maximality in positive and negative environments. The alternative- and supervaluationist theory by Križ & Spector 2021, for instance, makes parallel predictions for the cases considered here

5 Actually, the technical notion employed by Križ is the concept of a *current issue*, closely related to but strictly speaking different from that of the QuD. For present purposes the differences are immaterial.

but does not rely on trivalency. We group the theories therefore together for present purposes.[6]

2.1.2 Predictions

We illustrated the predictions of the trivalent theory for simple positive and negative sentences. For the more complex sentences like the ones we tested, repeated below in (9), we need an explicit account of how the homogeneity of predicates projects. In order to save space, we will not discuss different theoretical options for the projection algorithm in detail here and refer the reader to Križ 2015 for discussion. Instead, we will just give the gist of one of the most standard choices—namely, so-called Strong Kleene—and illustrate its predictions with respect to the sentences we tested in the target scenarios that we will describe below.

(9) a. Every boy opened his presents.
    b. No boy opened his presents.
    c. Not every boy opened his presents.
    d. Exactly two boys opened their presents.

According to the Strong Kleene projection algorithm, the projection of the third truth-value # is handled in a supervaluational fashion. Roughly: we consider the possibilities of mapping # to either true or false, and if these different possibilities agree on the overall truth value of the whole sentence, then the sentence denotes TRUE or FALSE; otherwise it denotes #.

It will be useful to illustrate the predictions with respect to the target scenarios we used in the experiments and to which we will return below. For example, consider (9a) in the top left scenario given in Figure 1: (9a) denotes # in this scenario. The reason is that, depending on how we resolve the #-cases of Frank and Nathan, the denotation of the quantified sentence changes. The same goes for (9b) in the top right scenario: here it is crucial how we resolve the #-cases of Leo and Mike. Again, the different options do not yield the same result and therefore (9b) ends up denoting #. It can be easily verified that (9c) is predicted to denote # for parallel reasons in the target scenario on the left in Figure 1. Crucially, (9d) denotes # in both the scenario on the left and the on the one the right of that same figure, but for different reasons. In the scenario on the left, the positive component of (9d) saying there are two boys who opened their presents clearly denotes TRUE. Depending on how the #-cases of Frank and Nathan are resolved, the negative component saying that no more than two boys opened their presents is either TRUE or FALSE. Therefore (9d) as a whole

6 Križ & Spector 2021 argue that the proposal by Križ 2016 is too permissive, and overgenerates judgements of ‘effectively true’ in certain contexts, while their theory is more restrictive. For our immediate concerns the differences between the theory are immaterial.

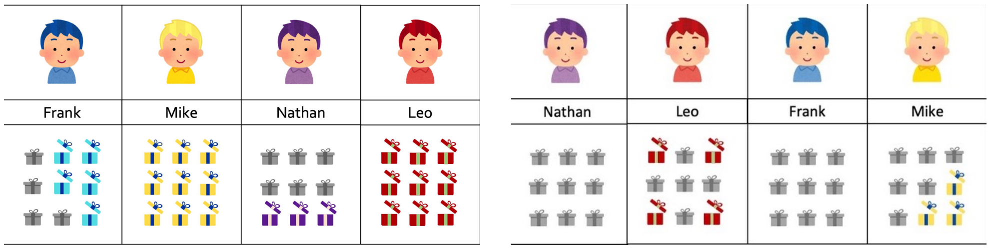

Figure 1 Examples of visual scenarios used in Experiment 1-3. The top left picture is the intended target scenario for *every* and the top right picture the one for *no*.

SENTENCE	NON-MAXIMALITY
Every boy opened his presents	✓
No boy opened his presents	✓
Not every boy opened his presents	✓
Exactly two boys opened their presents (*positive part*)	✓
Exactly two boys opened their presents (*negative part*)	✓

Table 1 Predictions of the trivalent theory for the tested sentences with respect to non-maximality. ✓ means that non-maximality is predicted to be possible, depending on what is relevant in the context, and $\times$ that it is not predicted.

receives #. Now, in the scenario on the right, the negative component is clearly TRUE. For the positive component, however, it depends on how the #-cases of Leo and Mike get resolved. Again, (9d) as a whole is therefore #. Taken together this means that on the symmetric approach the positive and the negative components of (9d) independently of each other are sensitive to the QuD and can therefore independently contribute non-maximality.

In sum, the underlying trivalent semantics predicts all sentences in (9) to denote # in our target scenarios. However, in the same way as above, Križ's theory predicts non-maximality to be possible in all of those cases, using the same ingredients, namely pragmatic re-interpretation of # dependent on what's relevant in the context. For instance, in a context in which, for all the boys, we want to know whether they opened any of their presents, (9a) will tend to be judged as true more than in a context in which we want to know whether they opened all of their presents. The predictions are summarised in Table 1: non-maximality is predicted for all of the sentences above, provided they are uttered in the right context.

2.2 Asymmetric theory: implicature theory

2.2.1 The theory

The implicature theory (Magri 2014, Bar-Lev 2018, Bar-Lev 2021) captures homogeneity by appealing to a basic existential semantics for plural definites. To illustrate the idea, consider a context in which the relevant presents are commonly known to be a, b and c, and nothing else. In such context, the literal meaning of (1a) is analysed as in (10): Frank opened at least one of a, b and c. (10) is obviously too weak to capture the intuitively perceived interpretation of (1a), but its negation, in (11) immediately accounts for the observed reading of (2a).

(10) $\exists x \in \{a,b,c\}(\mathsf{frank's.present}(x) \wedge \mathsf{opened}(\mathsf{frank},x))$

(11) $\neg\exists x \in \{a,b,c\}(\mathsf{frank's.present}(x) \wedge \mathsf{opened}(\mathsf{frank},x))$

In order to capture the positive case, the implicature theory argues that the basic existential semantics is strengthened by an implicature, which gives rise to the maximal reading in (12).

(12) $\forall x \in \{a,b,c\}(\mathsf{frank's.present}(x) \rightarrow \mathsf{opened}(\mathsf{frank},x))$

To illustrate the theory in more concrete terms, we use the theory proposed by Bar-Lev 2021 (and we refer the reader to Bar-Lev 2021 for discussion of Magri's (2014) alternative implementation). The theory is based on a few main ingredients. Firstly, Bar-Lev (2021) assumes that plural predication is associated with an existential pluralization operator ∃-PL, which itself is associated with a domain of quantification D.[7] Assuming that the relevant presents are $\{a,b,c\}$, the meaning for (13) is the existential meaning indicated in (10) above.

(13) $[\![\mathsf{Frank}\ [\mathsf{his\ presents}]_1\ [\exists\text{-}\textsc{pl}_D\ \mathsf{opened}\ t_1]]\!] \Leftrightarrow$
$\exists x \in \{a,b,c\}(\mathsf{frank's.present}(x) \wedge \mathsf{opened}(\mathsf{frank},x))$

Secondly, Bar-Lev (2021) assumes that scalar implicatures are associated with the presence of a silent exhaustification operator, 'EXH', which combines with a sentence and strengthens it with its scalar implicatures. More precisely, EXH is an operator that takes a sentence, the so-called prejacent, and a set of contextually relevant

7 The meaning of the existential pluralisation operator is as in (i) (where PART(x) is the set of atomic parts of x). We are assuming that the plural definite moves from the object position to make the meaning composition work, but nothing crucial hinges on this assumption and other analytical options are possible.

(i) $[\![\exists\text{-}\textsc{pl}_D]\!] = \lambda P_{\langle e,t\rangle}\lambda x_e \exists y \in D \cap \textsc{part}(x)[P(y)]$

alternatives to the sentence as arguments, and returns the conjunction of that sentence with the negations of a subset of its relevant alternatives, namely those relevant alternatives that are 'innocently excludable.' The definition of innocent exclusion is given in (14), where 'C' stands for the set of salient alternatives. Moreover, in addition to conjoining the prejacent with the negation of its innocently excludable alternatives, the exhaustivity operator also conjoins the prejacent with a subset of other alternatives, those that are 'innocently includable'. Concretely, innocently includable alternatives, as defined in (15), are those alternatives that are in all maximal subsets of alternatives that can be conjoined consistently with the prejacent and the negations of all innocently excludable alternatives, (15).

(14) $$\mathsf{IE}(p,C) := \left\{ C' \;\middle|\; \begin{array}{l} C' \text{ is a maximal subset of } C \text{ such that for some } w, \\ p(w) = \text{TRUE and each member of } C' \text{ denotes FALSE at } w \end{array} \right\}$$

(15) $$\mathsf{II}(p,C) := \left\{ C'' \;\middle|\; \begin{array}{l} C'' \text{ is a maximal subset of } C \text{ such that for some } w, \\ p(w) = \text{TRUE} \\ \text{and each member of } \mathsf{IE}(p,C) \text{ denotes FALSE at } w \\ \text{and each member of } C'' \text{ denotes TRUE at } w \end{array} \right\}$$

Putting this together, the definition of EXH is shown in (16): EXH conjoins the prejacent with the negations of all innocently excludable alternatives and with all the innocently includable alternatives.

(16) $$[\![\text{EXH}]\!](C)(p)(w) \Leftrightarrow p(w) \wedge \forall \phi \in \mathsf{IE}(p,C)[\neg [\![\phi]\!](w)] \wedge \forall \phi \in \mathsf{II}(p,C)[[\![\phi]\!](w)]$$

With these definitions in place, we can show that the strengthened meaning of (1a) corresponds to its maximal universal reading indicated above. This is because the alternatives to (13) are assumed to be the corresponding sentence where ∃-PL is associated with subdomains of D.

(17) $$\left\{ \begin{array}{l} \text{Frank [his presents]}_1 \text{ [}\exists\text{-PL}_{\{a,b,c\}} \text{ opened t}_1\text{],} \\ \text{Frank [his presents]}_1 \text{ [}\exists\text{-PL}_{\{a,b\}} \text{ opened t}_1\text{],} \\ \text{Frank [his presents]}_1 \text{ [}\exists\text{-PL}_{\{b,c\}} \text{ opened t}_1\text{],} \\ \text{Frank [his presents]}_1 \text{ [}\exists\text{-PL}_{\{a,c\}} \text{ opened t}_1\text{],} \\ \text{Frank [his presents]}_1 \text{ [}\exists\text{-PL}_{\{a\}} \text{ opened t}_1\text{],} \\ \text{Frank [his presents]}_1 \text{ [}\exists\text{-PL}_{\{b\}} \text{ opened t}_1\text{],} \\ \text{Frank [his presents]}_1 \text{ [}\exists\text{-PL}_{\{c\}} \text{ opened t}_1\text{]} \end{array} \right\}$$

None of the alternatives in (17) is excludable, but all of them are includable, so we end up with the meaning in (18). In other words, the basic existential meaning of cases like (1a) is strengthened to a universal meaning through a 'maximality' implicature.

(18) $[\![\text{EXH}[\text{Frank [his presents]}_1\ [\exists\text{-PL}_D \text{ opened } t_1]]]\!]$
$\Leftrightarrow$
$\exists x \in \{a\}(\text{frank's.present}(x) \wedge \text{opened}(\text{frank}, x)) \wedge$
$\exists x \in \{b\}(\text{frank's.present}(x) \wedge \text{opened}(\text{frank}, x)) \wedge$
$\exists x \in \{c\}(\text{frank's.present}(x) \wedge \text{opened}(\text{frank}, x))$
$\Leftrightarrow$
$\forall x \in \{a, b, c\}(\text{frank's.present}(x) \rightarrow \text{opened}(\text{frank}, x))$

Importantly, no strengthening applies to negative sentences like (2a), because EXH is generally assumed not to be able to appear in the direct scope of negation (and exhaustifying above negation would be simply vacuous here, as the prejacent entails all of its alternatives). This restriction is generally formulated as a constraint on EXH's distribution that prevents it from appearing in the scope of negation in order to avoid meaning weakening, along the lines of (19) (see Chierchia, Fox & Spector 2012, Fox & Spector 2018 among many others).[8]

(19) **Constraint on the distribution of EXH**
An occurrence of EXH in a sentence S is not licensed if the resulting meaning of S with that occurrence of EXH is *entailed* by S without it.

The application of (19) in cases like (2a) accounts for the strong reading under negation, by blocking the LF shown in (20), the outcome of which would be entailed by the literal meaning of the sentence without EXH (see above in (11)).

(20) $[\![\text{not}[\text{EXH}[\text{Frank [his presents]}_1\ [\exists\text{-PL}_D \text{ opened } t_1]]]]\!] \Leftrightarrow$
$\neg\forall x \in \{a, b, c\}(\text{frank's.present}(x) \rightarrow \text{opened}(\text{frank}, x))$

In addition, in this theory, non-maximality is the by-product of contextual modulation of the maximality implicature. That is, while the strengthened meaning of the positive example above is obtained by quantifying over the alternatives in (17), in some contexts, some of these alternatives can be ignored, or pruned, and such pruned sets of alternatives can give rise to weaker truth-conditions, which amount to non-maximal readings. For instance, the strengthening of (1a) over the alternatives in (21)—a set from which the atomic alternatives have been pruned—gives rise to the non-maximal reading in (22); a reading we could paraphrase as: Frank opened at least two of *a*, *b*, and *c*.

8 Note that if one were to adopt Bassi, Del Pinal & Sauerland's (2021) version of EXH for the present needs, no strengthening would occur. While their EXH would freely occur under negation, it would not have a semantic effect here.

(21) $\left\{ \begin{array}{l} \text{Frank [his presents]}_1\ [\exists\text{-}\textsc{pl}_{\{a,b,c\}} \text{ opened } \text{t}_1], \\ \text{Frank [his presents]}_1\ [\exists\text{-}\textsc{pl}_{\{a,b\}} \text{ opened } \text{t}_1], \\ \text{Frank [his presents]}_1\ [\exists\text{-}\textsc{pl}_{\{b,c\}} \text{ opened } \text{t}_1], \\ \text{Frank [his presents]}_1\ [\exists\text{-}\textsc{pl}_{\{a,c\}} \text{ opened } \text{t}_1] \end{array} \right\}$

(22) $[\![\text{exh[Frank [his presents]}_1\ [\exists\text{-}\textsc{pl}_D \text{ opened } \text{t}_1]]]\!] \Leftrightarrow$
$\exists x \in \{a,b\}(\text{frank's.present}(x) \wedge \text{opened}(\text{frank},x)) \wedge$
$\exists x \in \{b,c\}(\text{frank's.present}(x) \wedge \text{opened}(\text{frank},x)) \wedge$
$\exists x \in \{a,c\}(\text{frank's.present}(x) \wedge \text{opened}(\text{frank},x))$

Finally, the mechanism of pruning alternatives has to be constrained so as to not overgenerate. Bar-Lev (2021), building and expanding on previous work on implicature (in particular Crnič, Chemla & Fox 2015), suggests that pruning is allowed only if it yields a meaning that is weaker than the one without pruning and that is relevant (given a QuD Q in the context), The constraint is formulated in (23).[9]

(23) **Constraint on pruning**:
$\textsc{exh}_C$ S is licensed for $C \subseteq Alt(S)$ given a contextually provided partition Q only if C is a maximal subset of $Alt(S)$, s.t.

a. $[\![\textsc{exh}_C \text{ S}]\!]$ is relevant to Q, and
b. $[\![\textsc{exh}_{Alt(S)} \text{ S}]\!] \Rightarrow [\![\textsc{exh}_C \text{ S}]\!]$

In sum, the implicature theory can account for the homogeneity pattern of definite plurals on the basis of an existential literal meaning, the maximality implicature, and independently justified constraints on the distribution of implicatures. Non-maximal readings arise by pruning alternatives for implicature computation, to the extent that such readings weaken the meaning of the sentence and that the resulting meaning is relevant in the context.

2.2.2 Predictions

As Bar-Lev (2021) points out, the fact that in the implicature approach non-maximality is tied to the presence of the maximality implicature makes the immediate prediction that non-maximal readings should only arise where implicatures are involved. As illustrated above, this predicts non-maximality to be possible in positive cases but not in negative ones. Let us now turn to the corresponding quantified sentences we tested in our experiments, repeated above in (9). Consider (9b)

9 The definition of relevance that Bar-Lev (2021) assumes in the standard one in (i): a proposition is relevant given a partition, if it corresponds to a union of cells of such partitions.

(i) A proposition p is relevant given a partition Q iff $\exists Q' \subseteq Q[p = Q']$

and (9c) first: These are negative contexts, so the economy constraint above prevents EXH to appear in the scope of the quantifiers (and it is easy to show that when applied globally it would be just vacuous). Both are therefore analysed without EXH. The resulting meanings in (24) and (25) could be paraphrased as 'no boy opened any of their presents' and 'not every boy opened any of their presents', respectively. Given that no implicature is involved, these sentences should not be acceptable in non-maximal contexts (say, contexts in which some or all of the boys opened a few of their presents).

(24) $[\![[\text{no boy}]_2\ [t_2\ [\text{his presents}]_1\ [\exists\text{-}\textsc{pl}_D \text{ opened } t_1]]]]\!] \Leftrightarrow$
$\neg\exists x[\mathsf{boy}(x) \wedge \exists y \in \{a,b,c\}[\mathsf{present.of}.x(y) \wedge \mathsf{opened}(x,y)]]$

(25) $[\![[\text{not every boy}]_2\ [t_2\ [\text{his presents}]_1\ [\exists\text{-}\textsc{pl}_D \text{ opened } t_1]]]]\!] \Leftrightarrow$
$\neg\forall x[\mathsf{boy}(x) \rightarrow \exists y \in \{a,b,c\}[\mathsf{present.of}.x(y) \wedge \mathsf{opened}(x,y)]]$

The predictions are different when we move to (9a): this is because (9a) is a positive environment, and therefore EXH can occur in its scope, as in (26).[10] The resulting meaning can be paraphrased as 'every boy opened all of their presents'. In addition, the maximality implicature can be modulated by context as described above, making (9a) assertable also in certain non-maximal contexts (e.g. when some or all of the boys didn't open all of their presents).

(26) $[\![[\text{every boy}]_2\ \textsc{exh}[t_2\ [\text{his presents}]_1\ [\exists\text{-}\textsc{pl}_D \text{ opened } t_1]]]]\!] \Leftrightarrow$
$\forall x[\mathsf{boy}(x) \rightarrow \forall y \in \{a,b,c\}[\mathsf{present.of}.x(y) \wedge \mathsf{opened}(x,y)]]$

In sum, the implicature approach predicts an asymmetry between (9a), on one side, and (9b) and (9c), on the other, with respect to non-maximality.

Finally, the predictions for (9d) are more involved. Here we focus on the LF where EXH occurs in the scope of the non-monotonic quantifier (as allowed by the economy constraint above), which should be one of the main available parses of the

10 This raises immediately the question as to whether EXH can also occur in other positions or not occur at all. That is, whether there are other possible LFs for the sentence in (9a). These are non-trivial issues for this approach, but we put them aside here and follow Bar-Lev 2021 in assuming that at least one EXH has to be present, when allowed by the economy constraint above. We refer the reader to Bar-Lev 2021 for discussion.

sentence.[11] The meaning predicted can be paraphrased as 'two boys opened all of their presents and all of the other boys didn't open all of their presents'.

(27) $[\![$[exactly two boys]$_2$ EXH[t$_2$ [their presents]$_1$ $\exists$-PL$_D$ opened t$_1$]]$]\!]$ $\Leftrightarrow$
$\exists^2 x[\mathsf{boy}(x) \wedge \forall y \in \{a,b,c\}[\mathsf{present.of}.x(y) \wedge \mathsf{opened}(x,y)]] \wedge$
$\forall z[\mathsf{boy}(z) \wedge (z \neq x)] \rightarrow \neg[\forall y \in \{a,b,c\}[\mathsf{presents.of}.z(y) \rightarrow \mathsf{opened}(z,y)]]]$

Firstly, note that, without pruning, (27) is false in a non-maximal situation relative to the positive part (e.g. two out of the four boys opened some of their presents and the others didn't open any of theirs), while it is true in a non-maximal situation in the negative part (e.g. two out of the four boys opened all of their presents and the others open some but not all of theirs). In other words, local exhaustification, in this case, gives us a way to make the sentence compatible with non-maximal situations in the negative part in a way that is independent from pruning. This is a main prediction of the asymmetric approach, to which we will come back below.

Secondly, beyond the effect of local exhaustification in itself, pruning can also play a role as before. Interestingly, though, given the nature of non-monotonic contexts, in the parsing above, pruning has the opposite effect on the positive and negative parts. That is, while pruning weakens the meaning in the positive part, it strengthens it in the negative part.[12] More concretely, pruning the alternatives corresponding to the singleton of each present, we derive a meaning we could paraphrase as '*two boys opened two or more of their presents, and all of the others didn't open two or more*.' This pruned meaning is weaker in the positive and stronger in the negative part, than the meaning above without pruning and, as a consequence, can make the sentence compatible with a non-maximal situation in the positive part, and, in principle, could make it incompatible with a non-maximal situation in the negative part. In sum, given the parse above, (9d) is predicted to be straightforwardly compatible with non-maximal situations in the negative component, while the QuD-

11 We will focus on the prediction of this parse but we also note that there are in principle two other main parsing possibilities of the sentence, with global exhaustification with or without local exhaustification, as in (i) and (ii). These two parsing can be shown to either be equivalent to the meaning without EXH or false in both the positive and negative conditions, depending on whether we are considering alternatives to *exactly two* where the numeral is replaced with other numerals. We leave it to the reader to verify this; See Bar-Lev & Fox 2020, Gotzner, Romoli & Santorio 2020 for related discussion.

(i) $[\![$EXH[[exactly two boys]$_2$ EXH[t$_2$ [their presents]$_1$ $\exists$-PL$_D$ opened t$_1$]]]$]\!]$

(ii) $[\![$EXH[[exactly two boys]$_2$ [t$_2$ [their presents]$_1$ $\exists$-PL$_D$ opened t$_1$]]]$]\!]$

12 This is compatible with (23) as pruning is made with respect to the alternatives of the local EXH and the effect of strengthening the negative part only happens after the global non-monotonic quantifier is factored in.

SENTENCE	NON-MAXIMALITY
Every boy opened his presents	✓
No boy opened his presents	×
Not every boy opened his presents	×
Exactly two boys opened their presents (*positive part*)	✓
Exactly two boys opened their presents (*negative part*)	✓*

Table 2 Predictions of the implicature approach for the tested sentences with respect to non-maximality. ✓ means that non-maximality is predicted through pruning, ✓* means that non-maximality is predicted through local exhaustification, and × that is not predicted at all.

dependent version of non-maximality can make it compatible with non-maximal situations in the positive component (and the latter could in principle make the sentence false in a non-maximal situation in the negative part).

The predictions of the implicature approach for the sentences in (9) are summarised in Table 2.

In sum, the asymmetric approach differs from the symmetric one with respect to the availability of non-maximal readings in positive vs. negative environments. It predicts non-maximality in non-monotonic contexts, albeit in different ways with respect to their positive and negative parts. In the next section, we turn to discuss our experiments testing the different predictions of the two approaches.

3 The experiments

We report on three web-based experiments that use a truth-value judgment task with a five-point Likert scale.[13] The design of the experiments is largely based on Križ & Chemla 2015. As we are interested in the interpretation of definite plurals in environments with different polarity profiles, it is important to make sure that they do not outscope the operator that is responsible for the polarity of the local environment for the definite plural in question. As mentioned in the introduction, Križ & Chemla used bound pronouns for this purpose, and we followed suit.

We designed our target sentences as follows. They all contain a definite plural with a pronoun bound by one of the following four quantifiers, which have different polarity profiles with respect to their nuclear scope: *every*, *no*, *not every*, and *exactly two*. The actual sentences we tested looked as follows. The gender of the children was counterbalanced between participants in each experiment.

13 See the Appendix for a binary judgment version of Experiments 1 and 2.

(28) a. Every boy/girl opened his/her presents.
b. No boy/girl opened his/her presents.
c. Not every boy/girl opened his/her presents.
d. Exactly two boys/girls opened their presents.

Experiment 1 tested (28a) and (28b), and Experiment 2 tested (28a) and (28c). Experiment 3 tested different aspects of the meaning of (28d), as will be explained below.

One crucial difference from Križ & Chemla 2015 is that these sentences were evaluated against an explicit background context. We prepared two types of background context, which we call REQUIRED and PROHIBITED context, and varied them between participants in each experiment. Specifically, the context makes salient the question whether or not all the presents are open, while the PROHIBITED context makes salient the question whether any present was open at all. Each participant was randomly assigned one of these contexts, and read a description of the context at the beginning of the experiment. The two contexts were described as follows, in the same way in all three experiments.

Firstly, the participant was introduced to a family with four children (all boys or all girls) with pictures for the characters,[14] and told that each year, they give presents to each other at three different events: Christmas, Easter, and what they call the "Family Day". They celebrate these events with the children's grandparents, who usually arrive at noon. Then, the family rule was stated, which was different for the two contexts.

- In the PROHIBITED context, it is explained that the family opens the presents together with the children's grandparents, who come from a different city, so the children have to wait until they arrive before opening the presents they received. Then the family rule is stated on the next screen as "Opening the presents is prohibited before the guests arrive".
- In the REQUIRED context, it is explained that the family celebrate the events together with the children's aunt and uncle, who live close by, and the children open the presents early in the morning before their aunt and uncle arrive, so that the apartment is in an orderly state by then. Then on the next screen the rule is stated as "Opening the presents is required before the guests arrive".

We made sure that all the descriptions about the rules contained no quantifiers, so as to avoid encouraging participants to contrast the interpretations of the target definite

14 The pictures were constructed based on 12 free images downloaded from https://www.irasutoya.com/. The copyright of the images belongs to the designer.

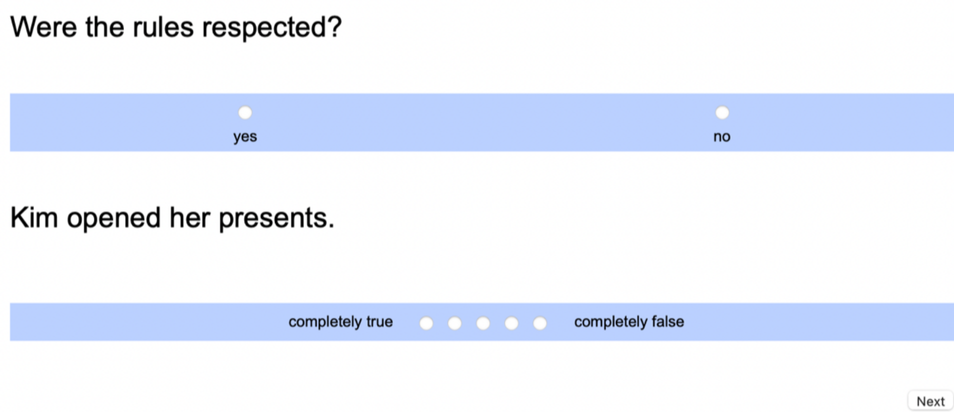

Figure 2 Response buttons used in the three experiments, with an example (filler) sentence "Kim opened her presents". A picture depicting who opened which presents is placed right above this, which in turn was placed below the rule statement accompanied by pictures of the childrens' parents.

plural *his/her presents* and quantified phrases like *all his/her presents*. Part of the instructions mentioned that the family rule is sometimes not obeyed.

Each trial of the three experiments consisted of a picture, depicting what happened in one of the events in one of the years from 2015–2019, and a sentence. Participants performed two tasks. The first task asked them to judge if the family rule was obeyed in the event depicted by the picture, by choosing between two radio buttons that were labeled "yes" and "no". This was done to make sure that participants understood and remembered the conversational context throughout the experiment. To remind them of the rule, the rule was displayed again on top of the screen in the same phrasing as in the instructions. The second task was to judge how well the sentence describes the scenario depicted in the picture on a five-point Likert scale ranging from 'completely true' to 'completely false'. Each trial was described as a single web page, whose response area looked as in Figure 2.

At this point, let us go back to the predictions of the two theories, with respect to the context manipulation described above. The predictions of the symmetric approach are summarised in Table 3 and those of the asymmetric approach in Table 4; the divergent predictions are boxed. The symmetric approach predicts non-maximality for all of the sentences tested, with an equal effect of context manipulation in positive versus negative grammatical environments, albeit going in opposite directions. The asymmetric approach predicts non-maximality and an

SENTENCE	CONTEXT	
	PROHIBITED	REQUIRED
Every boy opened his presents	✓	×
No boy opened his presents	×	$\boxed{✓}$
Not every boy opened his presents	×	$\boxed{✓}$
Exactly two boys opened their presents (*positive part*)	✓	×
Exactly two boys opened their presents (*negative part*)	×	✓

Table 3 Predictions of the symmetric approach for the tested sentences with respect to non-maximality and context manipulation. As before, ✓ means that non-maximality is predicted and × that it is not predicted.

SENTENCE	CONTEXT	
	PROHIBITED	REQUIRED
Every boy opened his presents	✓	×
No boy opened his presents	×	$\boxed{\times}$
Not every boy opened his presents	×	$\boxed{\times}$
Exactly two boys opened their presents (*positive part*)	✓	×
Exactly two boys opened their presents (*negative part*)	×	✓*

Table 4 Predictions of the asymmetric approach for the tested sentences with respect to non-maximality and context manipulation. As before, ✓ and ✓* means that non-maximality is predicted in different ways and × that it is not predicted.

effect of context manipulation only in positive grammatical environments. The non-monotonic case is more complicated in this approach, as non-maximality is predicted in two different ways and the predicted effect of context manipulation is unclear on the non-maximality in the negative aspect of the meaning.

3.1 Experiment 1

3.1.1 Methods

In our first experiment, we compared sentences involving the quantifiers *every* and *no* with a plural definite in their scope:

(29) a. Every boy/girl opened his/her presents.

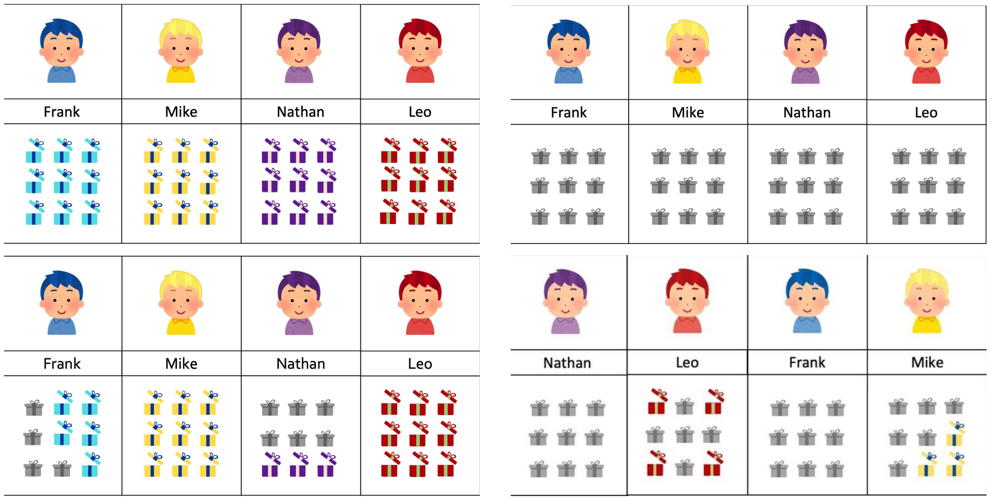

Figure 3 Examples of pictures used in Experiment 1. The top two pictures are items in the CONTROL conditions. The bottom left picture was a target item for *every* and the bottom right picture was a target item for *no*.

b. No boy/girl opened his/her presents.

Each sentence was presented in the context of one of three types of pictures. Each picture showed four children, each with nine presents, as in the examples in Figure 3. The pictures in the CONTROL conditions (e.g., the top two pictures in Figure 3) depicted uniform situations, either with all the presents closed (which were greyed out in the pictures), or with all the presents open. The sentences should receive straightforward true or false responses in these conditions. In particular, the control pictures that should elicit true responses for sentences with *every* should elicit false responses for sentences with *no* and vice versa.

The pictures in the TARGET condition depict non-maximal scenarios, where two of the four children opened some but not all of their presents. We prepared two types of TARGET pictures that differ with respect to what the other two children did, namely, pictures where the other children opened all of their presents (e.g., the bottom left picture in Figure 3) and pictures where they opened none of their presents (e.g., the bottom right picture in Figure 3). We tested both sentences, with *every* and with *no*, against both types of target pictures. Given that their non-maximal readings are true in different target pictures, however, only data from those pictures that could make the sentences true under a non-maximal reading were included in the analysis (i.e., for sentences with *every*, pictures similar to the bottom left one in Figure 3; for sentences with *no*, pictures similar to the bottom right one in Figure 3).

As explained above, each participant was assigned either the PROHIBITED context or the REQUIRED context, and after reading the instructions did one practice item. The practice item displayed the same four children, one of whom opened all of their presents, one of whom opened none of their presents, and the other two of whom some but not all of their presents. The sentence for the practice item was a non-quantified sentence mentioning one of the children by name, e.g., "Leo opened his presents". No feedback was given.

The experimental factors manipulated were CONTEXT (PROHIBITED or REQUIRED), PICTURE (TARGET, CONTROL-TRUE, CONTROL-FALSE) and POLARITY (POSITIVE = *every* or NEGATIVE = *no*). Among these, CONTEXT was a between-participant factor, and the other two were within-participant. We counterbalanced the position of the *yes* and *no* buttons for the first task and the order of the end points of the Likert scale, such that the *yes* button was always on the same side as 'Completely true', as well as the gender of the children (boys or girls).

Consequently, there were 8 lists (2 contexts $\times$ 2 genders $\times$ 2 ways of labeling the response buttons). Each list contained 24 experimental items, half with the positive quantifier *every* and the other half with the negative quantifier *no*. For the two tested sentences, with *every* and with *no*, there were 4 CONTROL-TRUE items, 4 CONTROL-FALSE items, and 4 TARGET items. In other words, each list contained 24 experimental items, half with the positive quantifier *every* and the other half with the negative quantifier *no*. As mentioned above, 2 of the TARGET items would not make the sentence true regardless of non-maximality, so we excluded them from the analysis. Consequently, for each participant, we analysed the data of 2 items for each Target condition. The experimental items appeared in a randomized order for each participant.

192 native speakers of English were recruited on Prolific (www.prolific.co) and paid £1.50 for their participation. 4 participants were excluded for low accuracy (75% or less correct) on the main judgment task on the CONTROL items, where responses above 3 for CONTROL-FALSE and responses below 3 for CONTROL-TRUE were considered to be errors. For the remaining participants, the mean accuracy of the CONTROL conditions was 98.1%.

3.1.2 Results

Figure 4 shows the mean acceptability ratings for the experimental conditions. We have recoded CONTEXT into LAX (PROHIBITED for *every*, REQUIRED for *no*) and STRICT (REQUIRED for *every*, PROHIBITED for *no*), because the two contexts should have different effects on the two sentences. LAX contexts should improve the acceptability in the TARGET condition (because they should make non-maximal

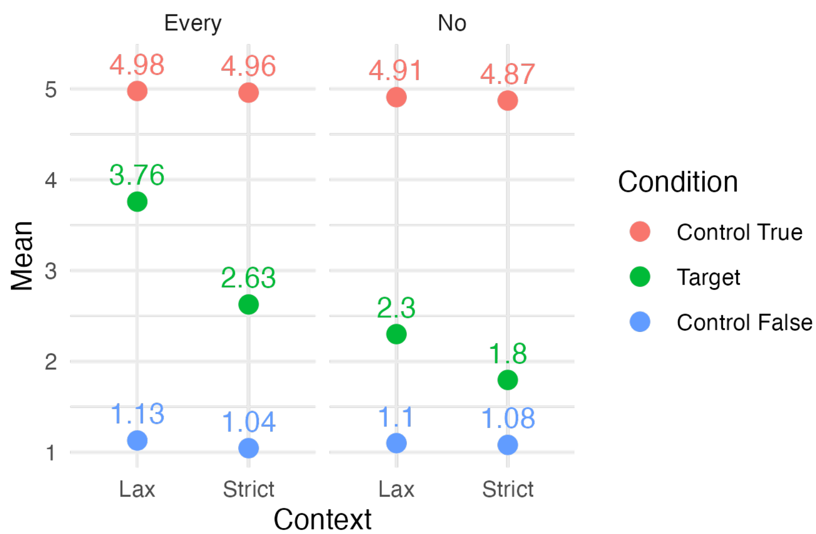

Figure 4 Mean acceptability ratings in Experiment 1. CONTEXT is recoded as LAX (favors non-maximality) or STRICT (favors maximality).

readings easier to access), while STRICT contexts should not (because they should make non-maximal readings harder to access).

We fitted a proportional-odds cumulative logit mixed effects model to the data from the TARGET condition, using `clmm()` of the `ordinal` package (Christensen 2022) on R (R Core Team 2022).[15] CONTEXT was sum-coded (STRICT = −0.5; LAX = 0.5), while POLARITY was treatment-coded with *no* as the reference level. These variables and their interaction were the fixed effects of the model. The random effect structure was maximal in the sense of Barr et al. 2013: by-participant variation for the intercept and POLARITY, and their correlation, as well as by-item variation for the intercept.[16]

The model estimates are summarised in Table 5. The *p*-values were obtained via likelihood ratio tests with simpler models using `anova()` on R. The model revealed a significant simple effect of CONTEXT (i.e., CONTEXT had an effect on *no*; $\chi^2(1) = 12.7$; $p < 0.001$) and a significant main effect of POLARITY (*every* is more acceptable than *no*; $\chi^2(1) = 52.2$; $p < 0.001$), as well as a significant interaction (*every* is more sensitive to CONTEXT than *no*; $\chi^2(1) = 10.9$; $p < 0.001$).

15 While we would have ideally included the CONTROL conditions in the model, floor and ceiling effects led to convergence issues in the ordinal logistic regression. However, since performance on the CONTROL conditions is near-perfect, very little information is lost from their exclusion.

16 As shown in Table 5, the by-item random intercept term does not explain any variance. Leaving it out makes no difference in the estimates of the other coefficients.

Fixed effects	β	s.e.	p
θ_1	-0.49	0.21	
θ_2	1.34	0.23	
θ_3	3.07	0.28	
θ_4	7.23	0.51	
CONTEXT	1.41	0.40	< 0.001***
POLARITY	3.36	0.37	< 0.001***
CONTEXT×POLARITY	2.43	0.74	< 0.001***
Random effects	σ^2	σ	ρ
By-participant intercept	4.62	2.15	
By-participant POLARITY	13.36	3.65	-0.42
By-item intercept	< 0.001	< 0.001	

Table 5 Estimates of the cumulative logit mixed effects model for Experiment 1. The p-values were obtained via likelihood ratio tests.

3.1.3 Discussion

The main findings of Experiment 1 are the following:

i. Definite plurals under *every* were overall judged as more acceptable than definite plurals under *no* in mixed scenarios.

ii. CONTEXT had an effect on the acceptability of definite plurals under *no* in mixed scenarios.

iii. CONTEXT had a larger effect on definite plurals under *every* than those under *no*.

Recall that according to the asymmetric approach, definite plurals may receive non-maximal readings under *every* but not under *no*, so it predicts CONTEXT to have an effect in the former case but not in the latter. Under the symmetric approach, definite plurals may receive non-maximality readings under both *every* and *no*, so it predicts parallel effects of CONTEXT.

The first and last finding above are therefore in line with the predictions of the asymmetric approach, while the second one is unexpected. Conversely, the second finding is in line with the expectations of the symmetric approach, while the other two pose challenges.

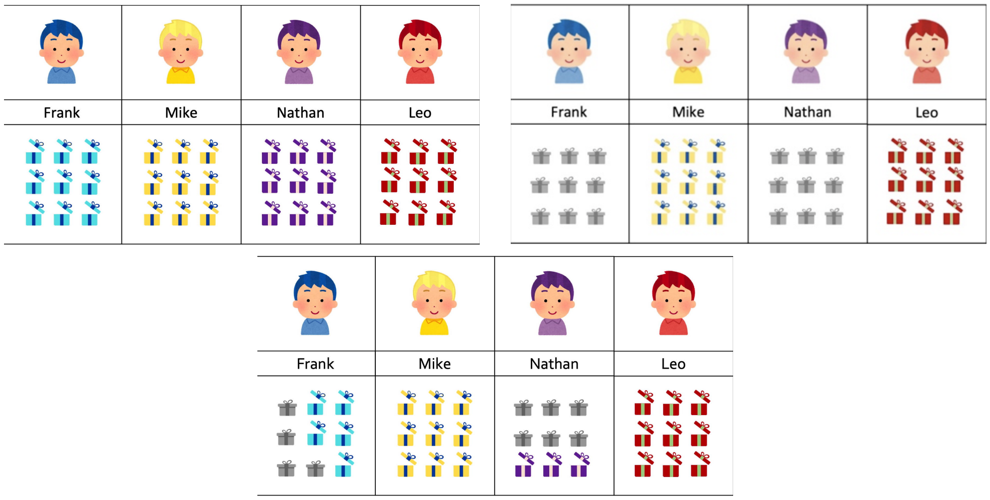

Figure 5 Examples of pictures used in Experiment 2. The top two pictures are items in the CONTROL condition and the bottom picture is an item in the TARGET condition.

3.2 Experiment 2

3.2.1 Methods

The design of Experiment 2 is essentially parallel to that of Experiment 1, with the negative quantifier being replaced by *not every*. One immediate consequence of this change is that the same TARGET pictures could be used for both sentences. That is, for all the items in the TARGET condition, two children opened some but not all of their presents, while the other two opened all of their presents. The CONTROL and TARGET pictures are illustrated in Figure 5.

The experimental sentences were spread over 8 lists (2 contexts $\times$ 2 genders $\times$ 2 ways of labeling the response buttons). Each list contained 32 items as a whole, with 24 experimental and 8 filler items. Half of the experimental items contained the positive quantifier *every* and the other half the negative quantifier *not every*. For each experimental sentence, there were 4 CONTROL-TRUE items, 4 CONTROL-FALSE items, and 4 TARGET items. Filler items were included in this experiment, because in the PROHIBITED condition, the correct answer to the *yes/no* task was uniformly *no* for all the items in the TARGET and CONTROL conditions. The filler items consisted of four pictures in which none of the children opened any of their presents, so that the correct answer for the *yes/no* task was *yes* in the PROHIBITED condition. Pictures were accompanied by sentences containing proper names in place of the quantifiers. 4 of these sentences were positive and 4 of them were negative as in (30).

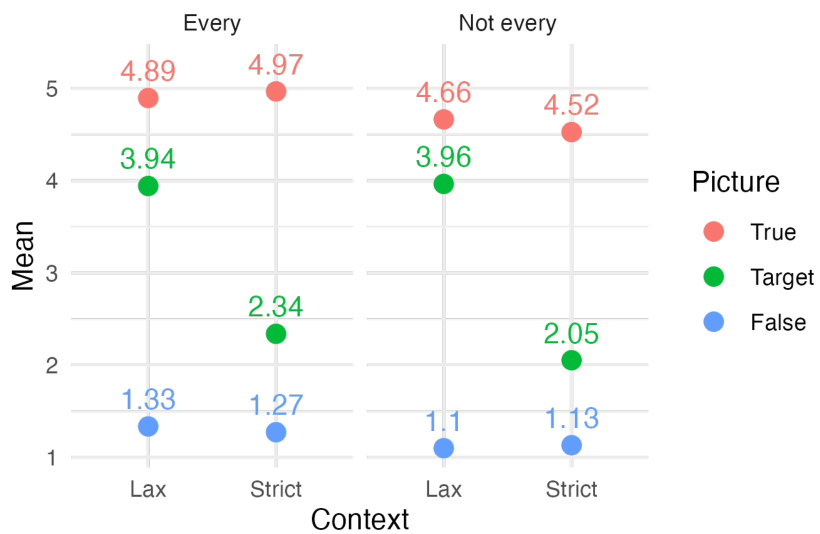

Figure 6 Mean acceptability ratings in Experiment 2. CONTEXT is recoded as LAX (favors non-maximality) or STRICT (favors maximality).

(30) a. Frank opened his presents.
b. Frank didn't open his presents.

The filler items were not used in the analysis.

192 native speakers of English were recruited from Prolific (www.prolific.co) and paid £1.50 for their participation. Six were excluded for low (75% or less) accuracy on the CONTROL items. The remaining participants' accuracy rate on the CONTROL items was 97.4%.

3.2.2 Results

Figure 6 shows the mean ratings of Experiment 2. As the figure indicates, some of the control conditions had less-than-perfect accuracy. We fitted a proportional-odds cumulative logit mixed effects model to the data from the CONTROL conditions. The model had CONTEXT, POLARITY, and PICTURE (CONTROL-TRUE vs. CONTROL-FALSE) and all their interactions as fixed effects. The mixed effect structure was the maximal one in the sense of Barr et al. 2013. Omitting the numerical details, it revealed effects of PICTURE and POLARITY, but crucially no effect of, or interaction with, CONTEXT.

Note that the unexpectedly high rating of *every* in the FALSE CONTROL condition is puzzling. Since this condition was essentially identical to Experiment 1, it might

Fixed effects	β	s.e.	p
θ_1	-1.88	0.10	
θ_2	-0.60	0.09	
θ_3	0.20	0.09	
θ_4	2.15	0.10	
CONTEXT	3.26	0.17	< 0.001***
POLARITY	0.13	0.17	0.47
CONTEXT$\times$POLARITY	-0.17	0.35	0.62
Random effects	σ^2	σ	
By-participant POLARITY	3.33	1.83	
By-item intercept	< 0.001	< 0.001	

Table 6 Estimates of the cumulative logit mixed effects model for Experiment 2. The p-values were obtained via likelihood ratio tests.

be that the *not every* quantifier is somehow responsible for this discrepancy between the two experiments. We have to leave it unexplained here, but importantly, the statistical model suggests that when analysing the TARGET conditions, we can be confident that any effect of CONTEXT is specific to the TARGET conditions and not an overall artifact of the task, although an effect of POLARITY, if any, may not be straightforwardly interpretable.

The data from the TARGET condition were analysed in a similar way to Experiment 1, with a proportional-odds cumulative logit mixed effects model with the same variable coding (i.e., sum-coding for CONTEXT with STRICT as -0.5 and LAX as 0.5 and treatment-coding for POLARITY with NEGATIVE as reference) and the same fixed effect structure. The random effects consisted of by-participant variance for POLARITY and by-item variance for the intercept. We left out by-participant variance for the intercept, as it shows a high collinearity with the by-participant random slope (with $\rho = -0.968$). Also, as in the case of Experiment 1, the by-item random intercept does not explain additional variance in the data and leaving it out would make no difference, but it is kept here.

The estimates of the model are summarised in Table 6. We observe a simple effect of CONTEXT (*not every* received a higher score in the LAX context; $\chi^2(1) = 97.3, p < .001$), no main effect of POLARITY (no evidence for difference in the acceptability of *every* and *not every* in the TARGET condition, but see the caveat above; $\chi^2(1) = 0.53, p = 0.469$), and no significant interaction between CONTEXT and POLARITY ($\chi^2(1) = 0.24, p = 0.624$).

3.2.3 Discussion

In Experiment 2, like in Experiment 1, we found a clear effect of CONTEXT. Importantly, however, this time we failed to observe a significant interaction effect between CONTEXT and POLARITY. The results therefore suggest that in this case CONTEXT had the same degree of robust effect on the two quantifiers *every* and *not every*.

The symmetry in the contextual effect and the overall similar acceptance of the positive and negative quantifiers is very much in line with the symmetric approach, and challenging for the asymmetric approach. Moreover, the difference from Experiment 1 is an issue for both approaches. We will discuss this in more detail in Section 4.

3.3 Experiment 3

Experiment 3 investigated the interpretation of plural definites in the scope of the non-monotonic quantifier *exactly two*. As discussed in the previous section, the theoretical predictions for this case are more involved, especially for the asymmetric approach. Our design mainly targets the predictions of the symmetric approach, but as we will discuss, the results end up having significant implications for both approaches.

3.3.1 Methods

The overall design of Experiment 3 is essentially the same as the previous two experiments, except that the target sentence was always (31).

(31) Exactly two boys/girls opened their presents.

The sentence was presented with one of four types of pictures, TRUE CONTROL, FALSE CONTROL, and two TARGET conditions, as illustrated in Figure 7. The idea is that in the POSITIVE TARGET condition, we tested if *their presents* can be read non-maximally with respect to the positive part of the meaning, while in the NEGATIVE TARGET condition, we tested if *their presents* can be read non-maximally with respect to the negative part of the meaning.

As before, the symmetric approach would expect a similar effect of context across the two conditions. The asymmetric approach, on the other hand, while still predicting an effect of context, also expects the negative condition to be accepted more (at least in the strict context where pruning should be less available). This is because, given the parse of (31) assumed above, repeated in (32), the predicted reading (i.e. exactly two boys opened all of their presents and all others didn't open all of theirs) is directly compatible with in our negative condition, while it is false in

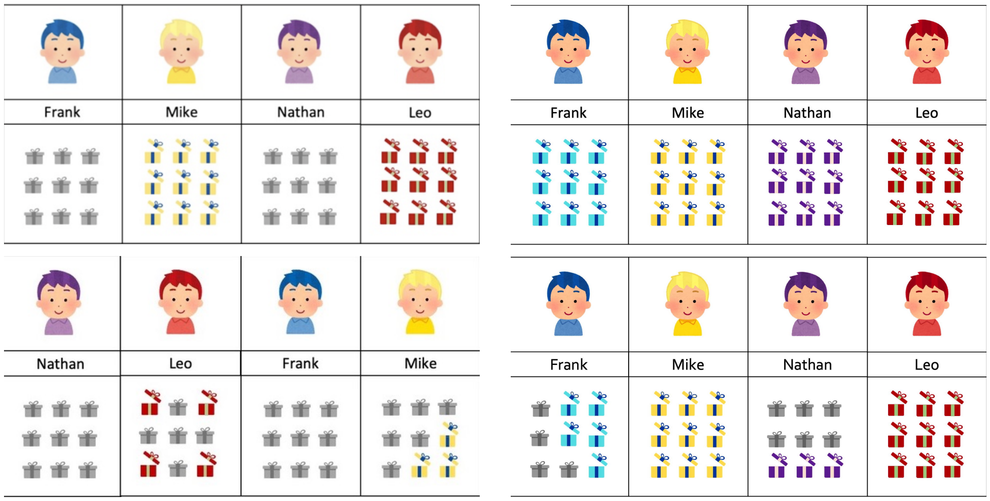

Figure 7 Examples of pictures used in Experiment 3. The top left picture is a TRUE CONTROL item and the top right picture is a FALSE CONTROL item. The bottom left picture was a target item for POSITIVE and the bottom right picture was a target item for NEGATIVE.

the positive one. So, assuming this parse is equally available in each condition, the sentence should be judged more true in the negative than the positive conditions.

(32) $[\![$[exactly two boys]$_2$ EXH[t$_2$ [their presents]$_1$ $\exists$-PL$_D$ opened t$_1$]]$]\!]$ $\Leftrightarrow$
$\exists^2 x[\mathsf{boy}(x) \wedge \forall y \in \{a,b,c\}[\mathsf{present.of}.x(y) \wedge \mathsf{opened}(x,y)]] \wedge$
$\forall z[\mathsf{boy}(z) \wedge (z \neq x)] \rightarrow \neg[\forall y \in \{a,b,c\}[\mathsf{presents.of}.z(y) \rightarrow \mathsf{opened}(z,y)]]]$

As in the previous two experiments, the experimental sentences were spread over 8 lists (2 contexts $\times$ 2 genders $\times$ 2 ways of labeling the response buttons). Each list contained 4 items in each of the four experimental conditions, so 16 items in total. All of these items contained the same sentence in (32).

192 native speakers of English were recruited on Prolific (https://www.prolific.co), and paid £1.05 for their participation. Using the same exclusion criterion as before (the participants whose accuracy rate on the control items at or below 75% were excluded from the analysis), we are left with 180 participants, whose average accuracy rate on the control items is 98.3%.

3.3.2 Results

Figure 8 shows the mean acceptability ratings for the experimental conditions. Note that the contexts are not recoded in this figure, as the control items are shared

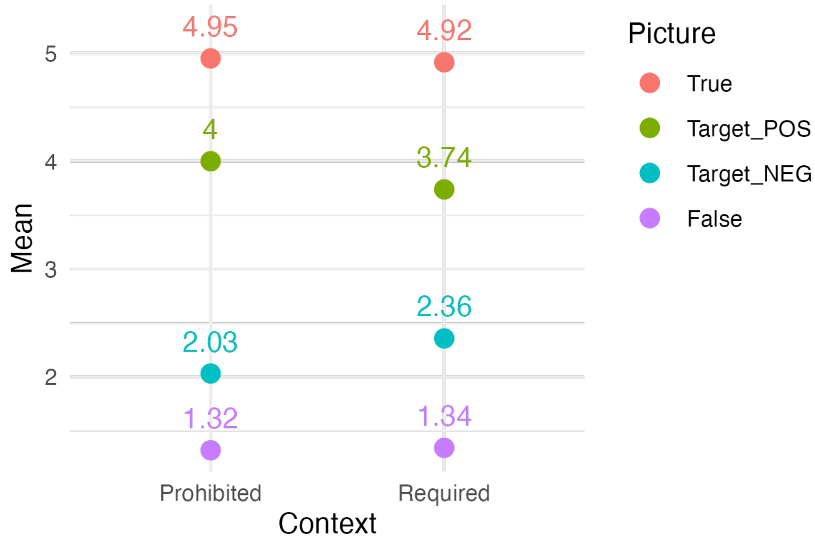

Figure 8 Mean acceptability ratings in Experiment 3.

between the two target conditions in this case, but in the statistical analysis reported below were conducted with recoding similarly to the previous two experiments. The two categorical variables, CONTEXT and POLARITY, were coded in the same way as in the previous experiments, i.e., sum-coding for CONTEXT (STRICT as -0.5 and LAX as 0.5) and treatment-coding for POLARITY with NEGATIVE as the reference level.

We fitted a proportional-odds cumulative logit mixed effects model to the data from the TARGET condition, using `clmm()`. This time, the model with a random by-participant intercept did not converge, so we did not include it.

The model estimates are summarised in Table 7. As before, the *p*-values were obtained via likelihood ratio tests with simpler models using `anova()`. We observe a significant simple effect of CONTEXT (CONTEXT had a significant effect on NEGATIVE; $\chi^2(1) = 15.8$, $p < 0.001$), and a significant main effect of POLARITY (POSITIVE was on average more acceptable than NEGATIVE; $\chi^2(1) = 34.2$, $p < 0.001$), but no significant interaction effect ($\chi^2(1) = 0.019$, $p = 0.892$).

3.3.3 Discussion

To summarize the main findings, we observe that non-maximality in the negative aspect of the interpretation of non-monotonic sentences containing plural definites is modulated by context and that non-maximality is judged as more acceptable with respect to the positive aspect of the meaning, in comparison to the negative aspect

Fixed effects	β	s.e.	p
θ_1	−0.11	0.26	
θ_2	0.73	0.26	
θ_3	1.50	0.27	
θ_4	2.93	0.28	
CONTEXT	0.61	0.15	< 0.001***
POLARITY	3.15	0.41	< 0.001***
CONTEXT×POLARITY	0.06	0.41	0.89
Random effects	σ^2	σ	
By-participant POLARITY	4.85	2.20	
By-item intercept	0.98	0.99	

Table 7 Estimates of the cumulative logit mixed effects model for Experiment 3. The p-values were obtained via likelihood ratio tests.

of the meaning. We should also note that the effect size of context manipulation is much smaller than in Experiment 2, and is numerically even smaller than the simple effect on *no* in Experiment 1.

The lack of a significant interaction effect suggests that non-maximality in the two aspects of the meaning of definite descriptions under *exactly two* is roughly equally context-sensitive. This is compatible with the predictions of both approaches. However, the difference in effect size from the previous experiment, especially the fact that among the negative contexts the context manipulation had the largest effect with *not every*, are puzzling for both approaches. We will come back to these points with more details in the next section.

4 General discussion

4.1 Main results

Let us first summarize the main experimental findings.

- We found evidence for an asymmetric effect of our context manipulation for non-maximal readings of plural definites in the scope of *every* and *no*, but found evidence for a symmetric effect in Experiment 2 and Experiment 3.

- In Experiment 1 and Experiment 3 the negative condition was accepted much less than the positive one.

Quantifier	Likert scale	Log-odds
Every (Exp. 1)	1.13	3.84
Every (Exp. 2)	1.60	3.09
Not every	1.91	3.26
No	0.50	1.41
Exactly two (positive)	0.26	0.66
Exactly two (negative)	0.33	0.61

Table 8 Effect size of context manipulation. The effect sizes on the Likert scale are the differences in the mean ratings, and the effect sizes in terms of logodds are computed from the regression models' estimates for the fixed effects.

- In all three experiments, context-sensitivity was observed in the negative condition, as well as in the positive condition.
- In addition, the effect size of context manipulation was different in the three experiments, as summarized in Table 4.1.

We will discuss the challenges these findings pose for the symmetric and asymmetric approaches in detail, and propose possible auxiliary assumptions and hypotheses that each approach could adopt to account for our results.

4.2 The symmetric approach

4.2.1 The challenges

The main challenges to the symmetric approach are as follows: Firstly, *no* gave rise to non-maximal readings much less than both *every* and *not every*. This needs an explanation, as the symmetric approach predicts that definite plurals should be context sensitive regardless of the semantic properties of grammatical environments they appear in. Secondly, and similarly, the difference between the positive and negative conditions in Experiment 3 is also not expected under this approach. Finally, the difference in the effect size we observed across the three experiments needs to be accounted for.

4.2.2 A possible response

We suggest here a hypothesis which would make the symmetric approach compatible with our experimental results. It should be noted that, as it stands, this hypothesis needs independent evidence, and we leave a systematic investigation of its predictions for future research.

The idea is as follows: our direct manipulation of the context allowed us to control its effects across the different quantifiers and showed that indeed context has a clear effect on the availability of non-maximality, which, however, depends on the quantifier involved. While this is a challenge for the symmetric approach, we cannot exclude that the effect size of context manipulation negatively correlates with the strengths of prior associations between context and sentences. Specifically, it could still be that sentences with *no* come with stronger prior associations with contexts that attenuate the effect of our manipulation and lower the availability of non-maximal readings of plural definites in its scope, compared to sentences with *every* or *not every*.

In order to test the hypothesis above, we will need a way to measure the priors associated with *no* and those of *every* or *not every*. While we cannot offer this here, for now, we would like to mention that the results of Experiment 3 are very much in line with this hypothesis. That is, given their nature, sentences with non-monotonic quantifiers allowed us to test the availability of non-maximal readings in the positive and negative part of the meaning, while keeping the prior bias associated with the sentence constant. In other words, given that we were testing the positive and negative entailments of the same sentence, we should find the symmetric behaviour that we in fact did find.

Finally, the symmetric approach still has to account for two remain aspects of the data. The first is the smaller effect size of Exp. 3. The second is the overall effect of polarity found in Exps. 1 and 3. One could, in principle, locate the source of the difference in the particular pictures we used, and given the theory under discussion, that seems to be the only reasonable way to explain the challenges. The idea would be that the target linguistic stimuli in the negative condition might have received low ratings, not (just) because of the truth-values they denote, but for other reasons, e.g., because they were judged as non-optimal expressions for describing the scenarios depicted in the pictures. In other words, the participants could have based part of their acceptability judgments on how likely they would use the sentence to describe the picture. However, it is at this point unclear if there is sufficient evidence to believe that such considerations for acceptability could affect truth-value judgments systematically. In addition, much more would have to be said in order to account for why the same consideration did not affect the positive conditions. For these reasons, we tentatively conclude that these remaining challenges are not fully explained under the symmetric approach.

4.3 The challenges for the asymmetric approach

4.3.1 The challenges

There are three challenges to the asymmetric approach: Firstly, the effect of context for negative sentences (both *no* and *not every*) is not expected, at least not by the theory as reviewed above. Secondly, it is not clear why, under this approach, definite plurals under *not every* gave rise to the non-maximal reading as much as those under *every* and were equally sensitive to contextual effects, in contrast with *no*. Finally, the smaller effect size with non-monotonic quantifiers with respect to that of the other quantifiers, and the effect of polarity in Experiment 3 are also puzzling for this approach.

4.3.2 A possible response

Bar-Lev 2021 acknowledges the first challenge for his asymmetric theory based on introspective judgments, and suggests that non-maximal readings in negative sentences should be explained by a different mechanism than pruning of alternatives, which has to do with (non-trivial) covers and only has truth-conditionally detectable effects in non-positive contexts. Putting aside the details (the interested reader is referred to Bar-Lev 2021), on the assumption that the availability of this mechanism is context-sensitive (as Bar-Lev 2021 in fact assumes), the effect of context on the availability of non-maximal readings can be explained.

In addition, if we also assume that the additional mechanism in question is more restricted and less available than pruning of alternatives, we could explain the asymmetry between *every* and *no*. The symmetric results of Experiment 2 remain puzzling, however, given that *not every* would be predicted to pattern with *no*.

In order to explain the difference between the two negative quantifiers, *no* and *not every*, a natural hypothesis to explore is that *not every* is not a ‘genuine’ negative environment. That is, when evaluating whether a sentence is negative, for the purpose of the distribution of implicatures, we should also take into account the implicatures the sentence may give rise to.[17] In fact, (33) does very robustly suggest that some boys did open their presents. Once this implicature is added to the literal meaning the resulting overall environment is not downward entailing, but non-monotonic.

(33) EXH[Not every boy opened his presents].
$\rightsquigarrow$ *some boys opened their presents*.

17 Compare the so-called ‘intervention effects’ of universal quantifiers for NPI licensing in sentences like **I didn’t give every boy any present*, which are considered to be due to indirect scalar implicatures emanating from *every* under *not*, effectively turning the environment of NPI *any* non-monotonic (Chierchia 2004, 2013).

As implicatures are allowed in non-monotonic contexts given the economy constraint assumed above, the parse above would give Bar-Lev's implicature theory another route to make the sentence compatible with a non-maximal situation. This embedded implicature would make the sentence equivalent to (34), thus true in a non-maximal situation. Finally, this mechanism would correctly not extend to *no*, thus predicting the difference between *not every* and *no*.

(34) EXH[Not every boy$_x$ EXH[x opened his presents]].
$\rightsquigarrow$ *not every boy opened all of his presents*
$\rightsquigarrow$ *Some boys opened all of their presents*

This response for the implicature theory would capture our results of Experiment 2 and the difference with Experiment 1. As in the case above for the auxiliary assumptions we added to the symmetric approach, this hypothesis too needs independent evidence. In particular, it makes an immediate prediction that implicatures should generally arise more easily in the scope of *not every* than that of *no*. For instance, (35) should more easily be read with the indicated implicatures than (36). Said differently, (35) should be accepted more easily in a context in which some of the boys opened both presents, (while the others opened only one of the two), than (36) in a context in which all the boys opened both.[18]

(35) Not every boy opened the present on their left or on their right.
? $\rightsquigarrow$ *not every boy opened one or the other but not both*
? $\rightsquigarrow$ *some boys opened one or the other but not both*

(36) No boy opened the present on his left or on his right.
? $\rightsquigarrow$ *no boy opened one or the other but not both*

We leave testing this prediction for future research, but let us note, for now, that even if the hypothesis above were supported, the challenges relating to Experiment 3 would remain. In particular, it is not clear, under the asymmetric approach, why the effect size of the contextual manipulation was much smaller in Experiment 3. Even more puzzling is the large difference between the positive and negative condition, with the latter being accepted much less. Recall that, as discussed, at least given the parse of (37) assumed above, repeated in (38), the predicted reading (i.e. exactly two boys opened all of their presents and all others didn't open all of theirs) is true in our negative condition above and false in the positive one. So, assuming this parse is

18 It should be remarked that provided that something like Bar-Lev's (2021) is necessary for the implicature theory and it will affect the interpretation of definite plurals under *not every*, meaning there will be two ways of getting non-maximal readings under *not every*, the symmetric results between *every* and *not every*, under the current hypothesis, should not carry over to the case of regular scalar implicatures.

equally available in each case, it is unclear why the sentence should be judged much less acceptable in the negative condition. If anything, we would expect a difference in the opposite direction.

(37) Exactly two boys opened their presents.

(38) $[\![\text{[exactly two boys]}_2 \text{ EXH}[\text{t}_2 \text{ [their presents]}_1 \; \exists\text{-PL}_D \text{ opened t}_1]]]\!] \Leftrightarrow$
$\exists^2 x[\text{boy}(x) \wedge \forall y \in \{a,b,c\}[\text{present.of}.x(y) \wedge \text{opened}(x,y)]] \wedge$
$\forall z[\text{boy}(z) \wedge (z \neq x)] \rightarrow \neg[\forall y \in \{a,b,c\}[\text{presents.of}.z(y) \rightarrow \text{opened}(z,y)]]]$

The asymmetric approach could, in principle, pursue a similar explanation as the one sketched above for the symmetric approach about alternative ways to describe the negative vs. positive picture, but as discussed there, the feasibility of such an explanation is unclear at this point. In sum, while the implicature theory can be made compatible with our results of Experiment 1 and 2, the results of Experiment 3 and their differences with the first two experiments remain challenging.

5 Conclusion

We reported on three experiments using a picture-sentence truth-value judgment task, investigating sentences with plural definites in the scope of positive, negative, and non-monotonic quantifiers, against different contexts. The experiments were designed to test divergent predictions of two main approaches in the literature with respect to non-maximal readings of plural definites: the symmetric and asymmetric approaches. Our results provided evidence for an asymmetry between plural definites in the scope of *every* and *no* sentences but a symmetry between plural definites in the scope of *every* and *not every*, and the positive and negative part of non-monotonic quantifiers such as *exactly two*. In addition, we found that non-maximal readings were possible and affected by context in all of these environments, and that the effect of context was larger in some cases than others. We argued that taken together, these results are challenging for both approaches to non-maximal readings of plural definites, and discussed possible directions each approach can explore in order to account for our results and the subsequent predictions these theoretical choices make, to be tested in future work.

Finally, we would like to remark that there are similar debates in the current theoretical literature between implicature and trivalent theories concerning other empirical domains, such as free choice inferences (Tieu, Bill & Romoli 2019), strong vs. weak donkey pronouns (Sun, Breheny & Rothschild 2020), and counterfactual conditionals (Marty, Romoli & Santorio 2020), and for these cases too, comparing positive and negative sentences with the explicit contextual manipulation will be informative in adjudicating between theoretical options.

References

Bar-Lev, Moshe. 2018. *Free choice, homogeneity and innocent inclusion*. The Hebrew University of Jerusalem dissertation.

Bar-Lev, Moshe E. 2021. An implicature account of homogeneity and non-maximality. *Linguistics and Philosophy* 44(5). 1045–1097. https://doi.org/10.1007/s10988-020-09308-5.

Bar-Lev, Moshe E. & Danny Fox. 2020. Free choice, simplification, and innocent inclusion. *Natural Language Semantics* 28. 175–223. https://doi.org/10.1007/s11050-020-09162-y.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bassi, Itai, Guillermo Del Pinal & Uli Sauerland. 2021. Presuppositional exhaustification. *Semantics and Pragmatics* 14(11).

Chierchia, G., D. Fox & B. Spector. 2012. Scalar implicature as a grammatical phenomenon. In C. Maienborn, K. von Heusinger & P. Portner (eds.), *Semantics: an international handbook of natural language meaning*, vol. 3, 2297–2331. Mouton de Gruyter.

Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Adriana Belletti (ed.), *Structures and beyond: The cartography of syntactic structures*, vol. 3, 39–103. Oxford: Oxford University Press.

Chierchia, Gennaro. 2013. *Logic in grammar: polarity, free choice, and intervention*. Oxford University Press.

Christensen, R. H. B. 2022. *ordinal—Regression Models for Ordinal Data*. R package version 2022.11-16. https://CRAN.R-project.org/package=ordinal.

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

Fox, Danny & Benjamin Spector. 2018. Economy and embedded exhaustification. *Natural Language Semantics* 26(1). 1–50. https://doi.org/https://doi.org/10.1007/s11050-017-9139-6.

Gajewski, Jon R. 2005. *Neg-raising: polarity and presupposition*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Gotzner, Nicole, Jacopo Romoli & Paolo Santorio. 2020. Choice and prohibition in non-monotonic contexts. *Natural Language Semantics* 28. 141–174. https://doi.org/10.1007/s11050-019-09160-9.

Guerrini, J. & J. Wehbe. 2024. Homogeneity as presuppositional exhaustification. Unpublished ms MIT and ENS.

Horn, Laurence. 1989. *A natural history of negation*. Chicago: Chicago University Press.

Krifka, Manfred. 1996. Pragmatic strengthening in plural predications and donkey sentences. In Teresa Galloway & Justin Spence (eds.), *Proceedings of SALT VI*, 136–153. Ithaca, NY: CLC Publications.

Križ, Manuel. 2015. *Aspects of homogeneity in the semantics of natural language*. University of Vienna dissertation.

Križ, Manuel. 2016. Homogeneity, maximality, and *all*. *Journal of Semantics* 33. 493–539.

Križ, Manuel. 2019. Homogeneity effects in natural language semantics. *Language and Linguistics Compass* 13(11). e12350. https://doi.org/10.1111/lnc3.12350.

Križ, Manuel & Emmanuel Chemla. 2015. Two methods to find truth value gaps and their application to the projection problem of homogeneity. *Natural Language Semantics* 23. 205–248. https://doi.org/10.1007/s11050-015-9114-z.

Križ, Manuel & Benjamin Spector. 2021. Interpreting plural predication: homogeneity and non-maximality. *Linguistics & Philosophy* 44(5). 1131–1178. https://doi.org/10.1007/s11050-015-9114-z.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 75(3). 522–551.

Löbner, Sebastian. 2000. Polarity in natural language: predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy* 23(3). 213–308.

Magri, Giorgio. 2014. An account for the homogeneity effects triggered by plural definites and conjunction based on double strengthening. In S. Pistoia-Reda (ed.), *Pragmatics, semantics and the case of scalar implicatures*, 99–145. Springer.

Malamud, Sophia A. 2012. The meaning of plural definites: a decision-theoretic approach. *Semantics and Pragmatics* 5. 1–58.

Marty, Paul, Jacopo Romoli & Paolo Santorio. 2020. Counterfactuals and Undefinedness: homogeneity vs supervaluations. In Joseph Rhyne, Kaelyn Lamp, Nicole Dreier & Chloe Kwon (eds.), *Proceedings of SALT 30*, 603–623.

Paillé, M. 2023. Trivalent exh and summative predicates. In *Semantics and linguistic theory*, 421–438.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Roberts, Craige. 1996. Information structure in discourse: towards an integrated formal theory of pragmatics. In *Osu working papers in linguistics 49: papers in semantics*, 91–136.

Sbardolini, G. 2023. Homogeneity and the illocutionary force of rejection. In *Proceedings of salt 33*.

Schwarzschild, Roger. 1994. Plurals, presuppositions and the sources of distributivity. *Natural Language Semantics* 2(3). 201–248.

Sun, Chao, Richard Breheny & Daniel Rothschild. 2020. Exploring the existential/universal ambiguity in singular donkey sentences. In Michael Franke, Nikola Kompa, Mingya Liu, Jutta L. Mueller & Juliane Schwab (eds.), *Proceedings of Sinn und Bedeutung 24*, vol. 2, 289–305.

Tian, Ye & Richard Breheny. 2019. Negation. In Chris Cummins & Napoleon Katsos (eds.), *The Oxford handbook of experimental semantics and pragmatics*, 195–207. Oxford University Press.

Tieu, Lyn, Cory Bill & Jacopo Romoli. 2019. Homogeneity or implicature: an experimental study of free choice. In Katherine Blake, Forrest Davis, Kaelyn Lamp & Joseph Rhyne (eds.), *Proceedings of SALT 29*, 706–726.

Tieu, Lyn, Manuel Križ & Emmanuel Chemla. 2019. Children's acquisition of homogeneity in plural definite descriptions. *Frontiers in Psychology* 10. 2329. https://doi.org/10.3389/fpsyg.2019.02329.