

Putting bare plurals into context*

Yizhen Jiang

Massachusetts Institute of Technology

Yasutada Sudo

University College London

Abstract Bare plurals give rise to plurality inferences in positive grammatical environments but not in negative grammatical environments. There are two main approaches. The implicature approach analyzes plurality inferences as scalar implicatures, and the homogeneity approach likens them to homogeneity effects exhibited by definite plurals. These two approaches make divergent predictions regarding the interaction between context-sensitivity and polarity. We report on three experiments designed to test these predictions by manipulating contexts. Our results show that the effect of context is symmetric with respect to polarity in non-quantified sentences but asymmetric in quantified sentences with *every* and *no* such that a larger effect is observed with sentences with *every*; and that readings of quantified sentences that are not fully plural (‘partial plurality inferences’) are available in both positive and negative quantified sentences but their availability is not affected by context. We discuss challenges that these results pose for the two theoretical approaches, and argue that overall they are more straightforwardly explained by the implicature approach.

Keywords: bare plurals, plurality inferences, implicature, homogeneity, context sensitivity

1 Introduction

Bare plurals in English receive either existential or generic readings. On the existential readings, they often give rise to *plurality inferences*, which are the main focus of this paper. For example, (1) is read with a very robust plurality inference that Frank opened more than one present.

- (1) Frank opened presents.
⇒ Frank opened more than one present.

Plurality inferences of bare plurals have two important properties that pose challenges for a compositional semantic analysis. The first property is that they tend to

* We would like to thank Richard Breheny, Nathan Klinedinst, Viola Schmitt, organizers and audiences at HNM2 workshop, XPRAGX and LFRG at MIT for helpful feedback. This project is supported by an AHRC-DFG grant (AHRC: AH/V003526/1; DFG: EB 523/2-1, STE 958/12-1, STE 2555/3-1).

disappear when bare plurals are interpreted in negative grammatical environments.¹ For example, the meaning of the simple negative sentence in (2) is not the negation of (1), but the negation of a number-neutral existential statement.

- (2) Frank didn't open presents.
⇒ Frank opened no present.

The second property is that plurality inferences are not strictly speaking entailments and may be absent in certain conversational contexts (see Grimm 2011, Tieu et al. 2020 for relevant discussion and experimental data). For example, (3) can be judged as coherent.

- (3) Many people opened presents. I opened presents, you opened presents. And even Frank opened presents, namely, the only present he got.

Furthermore, negative sentences like (2) can actually be read with plurality inferences, at least in some contexts and perhaps with the help of marked intonation, as in (4).

- (4) Frank didn't open presents. He only opened one.

To summarise the core observations, generally, both positive and negative sentences receive stronger interpretations—with a plurality inference in the positive and without a plurality inference in the negative—but weaker interpretations are not impossible, at least in certain contexts. To facilitate the discussion to follow, let us introduce the following terminology.²

- (5) a. A bare plural occurring in a positive grammatical environment has a STRONG interpretation if it gives rise to a plurality inference. Otherwise, it has a WEAK interpretation.
b. A bare plural occurring in a negative grammatical environment has a STRONG interpretation if it does not give rise to a plurality inference. Otherwise, it has a WEAK interpretation.

¹ For the purposes of this paper, negative grammatical environments can be understood simply as Downward Entailing (DE) environments, as the negative environments we will discuss in this paper are two garden-variety DE environments, namely, the scope of *not* and *no NP*. Similarly, the positive grammatical environments we will discuss are simple sentences like (1) and the scope of *every NP*, so they can be simply understood as Upward Entailing environments.

² This terminology would not be precise enough to talk about bare plurals in sentences with non-monotonic operators and/or multiple operators that affect polarity, but we will not discuss such cases. See the previous footnote.

Broadly, there are two theoretical approaches to plurality inferences: The implicature approach (Spector 2007, Zweig 2009, Ivlieva 2013, Mayr 2015, Sudo 2023) and the homogeneity approach (Križ 2017). As we will discuss in greater detail below, these two approaches make divergent predictions about how context-sensitive WEAK interpretations should be with respect to polarity. Roughly, the implicature approach makes an asymmetric prediction that WEAK interpretations of bare plurals in negative environments should be much harder, if not completely unavailable, in comparison to WEAK interpretations of bare plurals in positive environments. By contrast, the homogeneity approach assumes that WEAK interpretations of bare plurals in positive and negative environments are brought about by one and the same mechanism, whereby making a symmetric prediction that everything else being equal, WEAK interpretations of bare plurals in positive and negative environments should be similarly context-sensitive.

In this paper, we will report on three experiments that are designed to test the divergent predictions of the two approaches by measuring how different conversational contexts affect truth-value judgments of positive and negative sentences containing bare plurals. Our main findings are as follows.

- Experiment 1 compared bare plurals in simple positive and negative sentences like (6) presented against scenarios where the subject (e.g., Frank) only opened one present.

- (6) a. Frank opened presents.
 b. Frank didn't open presents.

The effect of context on WEAK interpretations was largely symmetric between positive and negative sentences.

- Experiment 2 tested bare plurals in positive and negative quantified sentences like (7) against a scenario where each boy opened only one present.

- (7) a. Every boy opened presents.
 b. No boy opened presents.

This time, the effect of context on WEAK interpretations was asymmetric in that it was larger for positive sentences than for negative sentences. The effect size for positive sentences was comparable to what was observed in Experiment 1.

- Experiment 3 tested the truth-value judgments of the same quantified sentences as Experiment 2, (7), with respect to what we call *partial plurality*. It is known that the positive quantified sentence in (7a) is judged as true in

a situation where some of the boys opened more than one present, and the others only opened one. The homogeneity approach, but not the implicature approach, analyses such partial plurality judgments as a special case of WEAK interpretations. There are two testable predictions of this analysis. Firstly, the negative quantified sentence in (7b) should be similarly judged as true in a situation where some of the boys opened no presents, and the others opened one. Secondly, for both positive and negative sentences the availability of such partial plurality judgments should be context-sensitive in the same way as other cases of WEAK interpretations. The results of Experiment 3 showed no effect of context for either positive or negative sentences.

We will argue that over all these experimental results are more straightforwardly explained by the implicature approach than by the homogeneity approach, but the results of Experiment 1 pose certain challenges for the implicature approach. We will discuss a possible amendment.

2 Theories of plurality inferences

Bare plurals in English receive either existential or generic interpretations (Carlson 1977, Diesing 1992, Delfitto 2006, Dayal 2011). A number of compositional theories have been proposed to account for this ambiguity, many of which assume that bare plurals themselves are non-quantificational and only provide variables to be bound by a separate quantificational operator in the same sentence. Luckily for us, for the goals of this paper, we need not make a theoretical commitment about such compositional details, because we will exclusively focus on existential interpretations and also because our concern is only the sentence-level meaning, so it is immaterial for us where in the compositional derivation the existential force comes in.

It should be mentioned at this juncture that (simple) bare plurals on existential readings are believed to always take maximally narrow scope. Theories have been constructed to explicate this peculiar scope restriction, but for reasons just mentioned, this compositional question does not concern us. Nonetheless, it proved to be a practically convenient feature in designing our linguistic stimuli, as it let us not worry about scope ambiguity. That being said, however, in light of the experimental results, we will come back to this assumption at the end of the discussion section.

As outlined at the beginning, existential interpretations of bare plurals are often associated with plurality inferences. In this section, we will review relevant aspects of the two major approaches to be compared, the implicature approach and the homogeneity approach, and highlight different predictions they make.

2.1 Implicature approach

The implicature approach assumes that existential interpretations of bare plurals are simply semantically number-neutral. This immediately accounts for the STRONG interpretation in negative grammatical environments, but something needs to be said about the STRONG interpretation in positive grammatical environments. To this end, the implicature approach strengthens the meaning with a scalar implicature.

All versions of the implicature approach to plurality inferences make use of a singular alternative formed with a singular indefinite *a(n) NP* in place of the bare plural (Spector 2007, Zweig 2009, Ivlieva 2013, Mayr 2015, Sudo 2023). The theories differ with respect to the specific ways in which this singular alternative is used, but these theoretical details do not concern us here (see Sudo 2023 for a recent summary). Rather, we will focus on the following key predictions that all versions of the implicature approach commonly make.

- The distribution of plurality inferences should mirror the distribution of scalar implicatures. In particular in the scope of negation or a negative quantifier like *no NP*, scalar implicatures are generally absent by default, as illustrated by the lack of ‘not both’ implicatures of disjunction in the surface scope construal in (8).

- (8) a. The boy does not speak German or French.
b. No boy speaks German or French.

Similarly, plurality inferences should be absent by default in these negative environments. However, to the extent that scalar implicatures can be generated in such negative environments (which may require marked intonation), bare plurals should also be able to have plurality inferences.

- Scalar implicatures are sometimes absent in positive grammatical environments, as in (9).

- (9) The girl speaks German or French. The boy speaks German or French, too. In fact, he speaks both.

This is often understood in terms of contextual relevance: In conversational contexts where *and* is not relevant, the scalar implicature based on it can (and possibly must) be absent (see, e.g., Magri 2009, Crnič, Chemla & Fox 2015). Correspondingly, plurality inferences of bare plurals in positive sentences can (or must) be absent, if the singular alternative is contextually irrelevant.

- Lastly, in a positive sentence with a scalar item occurring in the scope of *every NP*, two readings are considered to be available, depending on whether the scalar implicature is computed above or below the quantifier.

- (10) Every boy speaks German or French.
- Local: Every boy speaks only German or only French.
 - Global: Every boy speaks at least one of German and French, and not every boy speaks both.

Correspondingly, two readings are predicted for bare plurals under *every NP*, and the global reading accounts for partial plurality.³

- (11) Every boy opened presents.
- Local: Every boy opened more than one present.
 - Global: Every boy opened at least one present, but not every boy opened only one.

No such ambiguity is predicted for bare plurals under *no NP*, simply because the global option will not lead to strengthening. Recall also that the local reading under *no NP* is very marked, if possible at all.

To summarise, the implicature approach accounts for STRONG interpretations of bare plurals in the same way as scalar implicatures triggered by scalar items like *or*, i.e., the base semantic meaning accounts for the negative case and an additional scalar inference accounts for the positive case. At the same time, the approach has some leeway for admitting WEAK interpretations as non-default readings: In the negative, a reading with a scalar implicature is not entirely impossible, although highly marked (often requiring marked intonation); in the positive, when the alternative is contextually irrelevant, WEAK interpretations should become possible.

It should be stressed that the assumed mechanisms behind WEAK interpretations are distinct for positive and negative cases. Furthermore, it is generally considered that finding a context where a scalar implicature of a scalar item occurring in a positive environment is absent due to the alternative being contextually irrelevant is easier than generating a scalar implicature in negative grammatical environments, so much so that it is part of the definition of scalar implicatures that they are cancellable and that they generally disappear when the scalar item is put in negative environments.

³ The derivation of the global interpretation requires further assumptions under certain theories, e.g., Spector 2007.

The first two experiments to be reported below tested this asymmetric prediction, whilst the third tested the final point above about partial plurality.

2.2 Homogeneity approach

The homogeneity approach to plurality inferences proposed by [Križ 2017](#) starts with the assumption that the semantics directly generates the STRONG interpretations for both positive and negative sentences. Specifically, a simple positive sentence with a bare plural is given a trivalent denotation along the lines of (12).

$$(12) \quad \llbracket \text{Frank opened presents} \rrbracket^w = \begin{cases} 1 & \text{if Frank opened more than one present in } w \\ 0 & \text{if Frank opened no present in } w \\ \# & \text{if Frank opened exactly one present in } w \end{cases}$$

Negation swaps the truth and falsity conditions, which results in the following semantics for the negative simple sentence.

$$(13) \quad \llbracket \text{Frank didn't open presents} \rrbracket^w = \begin{cases} 1 & \text{if Frank opened no present in } w \\ 0 & \text{if Frank opened more than one present in } w \\ \# & \text{if Frank opened exactly one present in } w \end{cases}$$

In both cases, the STRONG interpretation is entailed. In other words, the homogeneity approach captures the truth-value gap between the positive and negative sentences directly in the core semantics using trivalency.

The important core intuition that underlies the homogeneity approach is that bare plurals are parallel to definite plurals. Indeed, definite plurals also give rise to a similar truth-value gap, as illustrated in (14).

- (14) a. Frank opened his presents.
 \approx Frank opened all of his presents.
 b. Frank didn't open his presents.
 \approx Frank opened none of his presents.

[Križ \(2015, 2016\)](#) puts forward a trivalent theory of definite plurals that is parallel to [Križ's \(2017\)](#) theory of bare plurals.⁴ Specifically, the positive sentence (14a) is

⁴ In light of the recent surge of theoretical interests in homogeneity, it is an analytical option to extend other theories of homogeneity developed for definite plurals to bare plurals, e.g., [Bar-Lev 2021](#), [Križ & Spector 2021](#), [Guerrini & Wehbe 2024](#). Since such theories of bare plurals based on

given the following denotation, and the negative sentence (14b) is given a similar denotation except that the truth- and falsity-conditions are swapped.

$$(15) \quad \llbracket \text{Frank opened his presents} \rrbracket^w = \begin{cases} 1 & \text{if Frank opened all of his presents in } w \\ 0 & \text{if Frank opened none of his presents in } w \\ \# & \text{if Frank opened some but not all of his presents in } w \end{cases}$$

One aspect of the analogy to definite plurals that is particularly relevant for the current paper is that definite plurals are known to receive *non-maximal readings* in certain contexts. Normally, sentences with definite plurals receive interpretations as predicted by the truth-condition in the above trivalent analysis, but in certain contexts, they sound true even if the predicted truth-value is $\#$. Concretely, consider a context where Frank has received 9 presents from various relatives for Christmas, but has been told by his parents to keep all the presents closed for now until all the guests come. Suppose further he nonetheless opened 6 of them. In this context, one can truthfully utter (14a), even though the sentence should denote $\#$. Križ (2015, 2016) proposes that such non-maximal readings are not to be accounted for in the semantics, but in the pragmatics. Specifically, he claims that in contexts where the current conversational issue does not distinguish possible worlds in which the sentence denotes 1 and certain possible worlds in which the sentence denotes $\#$, the sentence may feel ‘true enough’ with respect to the latter possible worlds.⁵

Križ 2017 extends this idea to bare plurals in order to account for their WEAK interpretations. As remarked above, under the homogeneity approach, the trivalent semantic denotations directly encode the STRONG interpretations. On other hand, WEAK interpretations arise by the same pragmatic mechanism that derives non-maximal readings of definite plurals. That is, in contexts where the current conversational issue does not distinguish possible worlds in which the sentence denotes 1 and certain possible worlds in which the sentence denotes $\#$, the sentence may feel ‘true enough’ with respect to the latter possible worlds, even though the denotation of the sentence is $\#$.

By way of illustration, imagine a scenario where the current conversational issue at hand is whether Frank kept all presents closed, as he should, or opened any present

alternative theories of homogeneity were not available when the present project began in 2022, we do not consider them in this paper.

⁵ Note that generally, possible worlds in which the sentence denotes $\#$ are allowed to be heterogeneous with respect to the conversational issue. For example, in the same context as above, consider a possible world in which Frank opened 6 of the 9 presents he received but they happened to be all presents from ‘not so important relatives’, e.g., his parents and siblings, with whom he lives together. This possible world might not make the sentence ‘true enough’ and the sentence might be judged as neither true nor false.

at all. With this issue in mind, let us further suppose that we've found out that Frank did open one present, although he should not have done so. In this context, 'Frank opened presents' can be judged as true, even though it should denote #, according to the trivalent semantics. According to the homogeneity approach, this is because this particular possible world is pragmatically not distinguished from the possible worlds in which the sentence is true, in resolving the current conversational issue.

Importantly, the homogeneity approach makes a symmetric prediction for the negative simple sentence, which likewise has a trivalent denotation. That is, imagine a context with a different issue, say, a context where we are concerned with whether Frank opened more than one present or not, and suppose that we have found out that Frank opened exactly one present. The homogeneity approach predicts that such a context should be able to render the negative sentence 'Frank didn't open presents' 'true enough' with respect to this scenario to an extent similar to the above previous example with a positive sentence.

To summarise the discussion so far, Križ's homogeneity approach directly encodes the STRONG interpretations in the semantics, and accounts for WEAK interpretations using the pragmatic wiggle room created by conversational issues that render certain cases of # practically true for the purposes of resolving the current issue. Crucially, the predictions are symmetric with respect to polarity in that one and the same pragmatic mechanism governs the availability of WEAK interpretations of bare plurals in positive and negative sentences. This one crucial difference in prediction from the implicature approach, and Experiment 1 to be presented below tests this very difference.

Let us now turn to quantified sentences. A trivalent framework offers multiple potential denotations for each logical operator that converge with respect to classical truth-values (1 and 0) but diverge with respect to the third truth-value (#). Following his earlier work on definite plurals (Križ 2015, 2016), Križ (2017) uses what is called Strong Kleene projection as a general recipe for determining the denotations of logical operators. To save space, we will not explain what it is in the general case or his arguments for favouring it over alternatives, as what matters for the present paper is the predictions of the theory. Specifically, his theory assigns the following denotations for sentences containing bare plurals under *every NP* and *no NP*.

$$(16) \quad a. \quad \llbracket \text{Every boy opened presents} \rrbracket^w = \begin{cases} 1 & \text{if every boy opened multiple presents in } w \\ 0 & \text{if at least one boy opened no present in } w \\ \# & \text{otherwise} \end{cases}$$

$$\begin{aligned}
 \text{b. } & \llbracket \text{No boy opened presents} \rrbracket^w \\
 & = \begin{cases} 1 & \text{if no boy opened any present in } w \\ 0 & \text{if at least one boy opened multiple presents in } w \\ \# & \text{otherwise} \end{cases}
 \end{aligned}$$

Analogously to the case of non-quantified sentences, the theory predicts that intuitive truth-value judgments of quantified sentences may deviate from the semantic denotations, depending on the conversational issue. For instance, if the current issue is whether or not any boy opened any present at all, the universally quantified sentence may be judged as ‘true enough’, against a scenario where every boy opened only one present. Similarly, if the current issue is whether or not every boy opened more than one present, the negative sentence may be judged as ‘true enough’ against the same scenario.

Therefore, just as in the case of simple sentences, the predictions of the homogeneity approach about WEAK interpretations are symmetric with respect to polarity: Everything else being equal, both positive and negative quantified sentences should be able to receive WEAK interpretations in the pragmatics, when certain possible worlds in which the sentence denotes # are regarded as pragmatically equivalent to possible worlds in which the sentence denotes 1 with respect to the current conversational issue. Recall that the implicature approach makes an asymmetric prediction with respect to polarity such that such a WEAK interpretation is harder, if possible at all, to access, in comparison to a WEAK interpretation of a positive quantified sentence. Experiment 2 targets these diverging predictions.

In addition, [Križ \(2017\)](#) maintains that the same pragmatic mechanism accounts for what we referred to as partial plurality above. That is, the universally quantified sentence is judged as true with respect to a scenario where some boys opened more than one present whilst others only opened only one. According to the trivalent semantics in (16a), the sentence should denote # in such a scenario, but if the current conversational issue renders this particular possible world indistinguishable from possible worlds in which the sentence denotes 1 for the purposes of resolving the issue, then the sentence may be judged as true. Here, the homogeneity approach makes markedly different predictions from the implicature approach, so let us highlight some key differences.

- For the homogeneity approach, partial plurality is not a separate reading of the universally quantified sentence in the truth-conditional sense, but is a special case of pragmatic truths that arise from the interaction between the trivalent denotation and the pragmatics of conversational issues. For this reason, the truth-value judgment against a partial plurality scenario where some boys opened more than one presents and the rest of the boys only one

is expected to be as context-sensitive as other cases of WEAK interpretations. For the implicature account, on the other hand, partial plurality is not a case of WEAK interpretation, but rather of a STRONG interpretation, and no particular effect of context is expected.

- Unlike the implicature approach, the homogeneity approach predicts that the negative quantified sentence also should give rise to what corresponds to a partial plurality reading. That is, it is predicted that with the right kind of conversational issue, the negative quantified sentence should be judged as ‘true enough’ against a scenario where some boys opened no presents and the others opened exactly one, and also that the truth-value judgment here should be as context-sensitive as other cases of WEAK interpretations. The implicature approach does not predict the negative quantified sentence to be true in such a scenario, unless the plurality inference is generated in the scope of the negative quantifier, which is a marked option.

Experiment 3 is designed to test these divergent predictions.

2.3 Summary of the divergent predictions

The two approaches make different predictions for the availability and context-sensitivity of WEAK readings. Let us summarise below the key differences that our experiments are designed to test.

Firstly, for simple sentences, the implicature approach makes an asymmetric prediction with respect to polarity such that it should be easier to judge (17a) as true against a scenario where Frank opened only one present than to judge (17b) to be true against the same scenario. The homogeneity approach makes a symmetric prediction here such that the truth-value judgments of both sentences should be equally context-sensitive.

- (17) a. Frank opened presents.
 b. Frank didn’t open presents.

Secondly, for quantified sentences like (18), the two approaches make similarly different predictions against a scenario where every boy opened only one present.

- (18) a. Every boy opened presents.
 b. No boy opened presents.

Note that the implicature approach predicts two STRONG readings for (18a), but for a ‘uniformly singular’ scenario where every boy opened only one present, neither STRONG reading is true, and only a WEAK interpretation is true.

Thirdly, against a ‘mixed’ scenario where some boys opened multiple presents and the others opened only one, the implicature approach predicts the positive quantified sentence (18a) to have a true *STRONG* reading as a possible reading, and as such it does not predict a particular effect of context, beyond the general effect on ambiguity resolution with respect to other possible readings (which include another *STRONG* reading where the implicature is generated below the quantifier and is truth-conditionally stronger). On the other hand, in a similarly mixed scenario where some boys opened no presents and the others opened only one, (18b) is only expected to be judged as true under the *WEAK* interpretation, which by assumption is very marked for the negative sentence, so the judgment pattern should be comparable to (18b) against the ‘uniformly singular’ scenario where every boy opened only one. The homogeneity approach, on the other hand, makes symmetric predictions here as well as in the previous two cases, namely, the truth-value judgments of the two quantified sentences with respect to these mixed scenarios should be parallel to the truth-value judgments with respect to the ‘uniformly singular’ scenario where every boy opened only one present.

3 Experiments

The divergent predictions we identified above are about context-sensitivity, and in order to directly evaluate them, truth-value judgments against different conversational contexts with different considerations about relevance and/or conversational issues will be useful. Since introspective judgments involving systematic manipulation of contextual factors tend to be unreliable, especially with respect to interaction effects, formal experiments are particularly useful here.

We modelled our experiments after Romoli et al.’s (2024) experiments, which manipulated conversational contexts in order to evaluate theoretical predictions about the effect of context on non-maximal readings of definite plurals. Since they successfully observed effect of context, we decided to use the exact same contexts they used, which are called *EXISTENTIAL* and *UNIVERSAL* contexts. In the *EXISTENTIAL* context, the conversational issue is whether or not there is any open present, and in the *UNIVERSAL* context, what matters is whether or not there is any closed present. In our visual stimuli, there are at most two presents for each individual (unlike in Romoli et al. 2024), so the *UNIVERSAL* context is about the issue as to whether each individual opened more than one present or not.

Following Romoli et al. 2024, we tested these contexts between participants. The rationale is that changing the context in the course of a single experiment would be cognitively demanding and might yield non-trivial interference effects on answers.

Putting bare plurals into context

Each context is introduced at the beginning of the experiment as a family rule for a specific family (see the Appendix for full details).⁶

- (19) a. EXISTENTIAL: You must keep each present you received closed!
b. UNIVERSAL: You must open each present you received!

3.1 Experiment 1: Simple sentences

The goal of Experiment 1 is to examine the availability and context-sensitivity of the WEAK interpretations of simple positive and negative sentences like (20).


- (20) a. Frank opened presents.
b. Frank didn't open presents.

We assessed truth-value judgments of these sentences on a 5-point Likert scale against singular scenarios where the relevant individual (e.g., Frank) only opened one of two presents.⁷ Crucially, the truth-value judgment task is performed with the EXISTENTIAL rule or the UNIVERSAL rule in mind.


Our linking hypothesis is that truth-value ratings indicate the extent to which target sentences are construed under the WEAK interpretation. Under both approaches under consideration, therefore, (20a) is expected to receive a higher rating in the EXISTENTIAL context than in the UNIVERSAL context. On the other hand, the two approaches make different predictions for (20b): According to the implicature approach, it should receive a low rating in both contexts; according to the homogeneity

⁶ The phrasing of family rules in our study is slightly different from Romoli et al. 2024. In particular, they used a definite plural *the presents*, as in ‘Opening the presents is prohibited before the guests arrive!’ in the EXISTENTIAL context and ‘Opening the presents is required before the guests arrive!’ in the UNIVERSAL context. They avoided using a quantifier in these statements, because they did not want the definite plural in the target sentence to be contrasted with a quantifier. This was not a concern in our experiment, but in our pilot study for Experiment 1, we followed them in matching the phrase in the rule and target sentence by stating the rules with a bare plural, as in ‘Opening presents is required before the guests arrive’ and ‘Opening presents is prohibited before the guests arrive’. However, we noticed in the data (of the context questions; see below) that some participants seemed to have interpreted the rule in the UNIVERSAL context (but not in the EXISTENTIAL context) with a plurality inference as in ‘Opening *more than one* present is required’, which is not intended here. Furthermore, under the intended interpretation of the rules, the bare plural should be read generically, rather than existentially, so it has a different interpretation in the rule and the target sentence. For these reasons, we decided to use quantifiers in stating the rules, as in (19), in all three experiments.


⁷ We followed Romoli et al. 2024 in using a 5-point Likert scale. As they discuss, this does not necessarily commit one to a non-bivalent view of truth-values. Rather, it provides a more sensitive measure than a binary judgment task, and is especially useful in cases like ours where multiple interpretations (STRONG vs. WEAK in our case) are expected. See Križ & Chemla 2015, Jasbi, Waldon & Degen 2019 for relevant discussion.



**Family rule: You must keep
each present you received
closed!**



Carrie



Was the rule respected?

No Yes

Carrie didn't open presents.

Completely
false

Completely
true

Figure 1: Example trial in Experiment 1

approach, it should receive a higher rating in the UNIVERSAL context than in the EXISTENTIAL context.

3.1.1 Data availability

Stimuli, data, and analysis code for Experiment 1 are accessible on the OSF at https://osf.io/zqanx/?view_only=48dbea9761f240c381a7045a82ce5684.

3.1.2 Methods

Participants 92 self-reported native English speakers were recruited in the experiment (mean age: 43.1 years old; age range: 20-73; 48 female, 44 male) online via Prolific (<https://www.prolific.co/>) (pre-screening criteria: UK/US IP addresses; access via computers). The average completion time was about 6 minutes and participants were paid £1.5 for their time. Data collection and storage follow the provisions of the Data Protection Act 2018 (<https://www.gov.uk/data-protection>). The experiment was approved by the Research Ethics Committee at the author's institution.

Materials Figure 1 shows an example display of trials in Experiment 1.⁸ As shown here, on each trial participants were presented with i) a context picture with a family rule (upper left); ii) a target picture (upper right); iii) a context question ‘Was the rule respected?’; and iv) a target sentence ‘[Name] opened presents’ or ‘[Name] didn’t open presents’ together with a 5-point Likert scale labelled ‘Completely false’ and ‘Completely true’ on the ends.

There were two types of target sentences varying in their POLARITY: ‘[Name] opened presents’ in the POSITIVE condition and ‘[Name] didn’t open presents’ in the NEGATIVE condition. The [Name] term in the target sentences was selected from the following list: Frank, Mike, Leo, Nathan, Phoebe, Jenny, Carrie, and Kim. The name in the target sentence matches the image of the child in the picture. Participants were asked to judge how truly the target sentence describes the situation presented in the test picture on a 5-point Likert scale (without numeric labels). The endpoints of the Likert scale are counterbalanced between ‘Completely true’(left)-‘Completely false’(right) and ‘Completely false’(left)-‘Completely true’(right). The gender of children is counterbalanced between boys and girls.

Each visual stimulus features a child with their name and two presents they have received. There were three types of test pictures for each sentence. For POSITIVE sentences, the CONTROL TRUE pictures had both presents open and the CONTROL FALSE pictures had both presents closed; in the TARGET pictures, only one present was open and the other was unopened. For NEGATIVE sentences, the CONTROL TRUE pictures had both presents closed, and the CONTROL FALSE pictures had both presents open; the TARGET pictures were the same as those in the POSITIVE condition. Thus, the only difference between the two polarity conditions is that the control pictures were switched. Figure 2 provides examples.

The context question ‘Was the rule respected?’ is used to make sure that participants paid enough attention to the context with the specific family rule and also to the visual scene, and to check if they understood them as we intended. Participants were asked to respond to this question by clicking the ‘Yes’ or ‘No’ buttons beneath the question.

Design POLARITY (NEGATIVE, POSITIVE) and CONDITION ([TARGET, CONTROL TRUE, CONTROL FALSE]) were within-participant factors, whilst CONTEXT (EXISTENTIAL, UNIVERSAL) was a between-participant factor. There were 4 repetitions of each experimental condition in each list. Therefore, there were 2 POLARITY \times 3 CONDITION \times 4 repetitions = 24 trials in total for each list and they were

⁸ The visual stimuli used in our three experiments were constructed based on 12 free images downloaded from <https://www.irasutoya.com/>, which were also used to construct the visual stimuli of Romoli et al. 2024. The copyright of the images belongs to the designer.





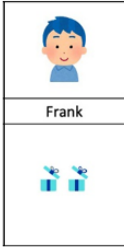
	CONTROL TRUE	CONTROL FALSE	TARGET
<p>POSITIVE Frank opened presents.</p>			
<p>NEGATIVE Frank didn't open presents.</p>			

Figure 2: Examples of CONTROL TRUE, CONTROL FALSE and TARGET pictures for POSITIVE and NEGATIVE conditions in Experiment 1.

presented in a random order for each participant. There were 8 lists differing in CONTEXT (EXISTENTIAL, UNIVERSAL), gender (boy, girl), and the label of the scale endpoints ('Completely true'(left)-'Completely false'(right) and 'Completely false'(left)-' Completely true'(right)). Participants were randomly assigned to one of the 8 lists.

Procedure The experiment was carried out by implementing an online survey using Gorilla Experiment Builder (<https://www.gorilla.sc>). Participants were first given general instructions at the beginning of the survey (see Appendix A for details). Before the experimental trials began, there were two practice trials featuring CONTROL TRUE pictures for POSITIVE sentences and CONTROL FALSE pictures for negative sentences. The first trial was accompanied by instructions (see Appendix B), whilst the second trial did not contain any instructions and was the same as control trials in the experimental phase. The purpose of practice trials was to help participants get familiar with the task as well as to provide an anchor for either the lowest point or the highest point of the scale, and hence minimize inter-participant variation in the use of the scale. Demographic information was collected at the end of the survey.

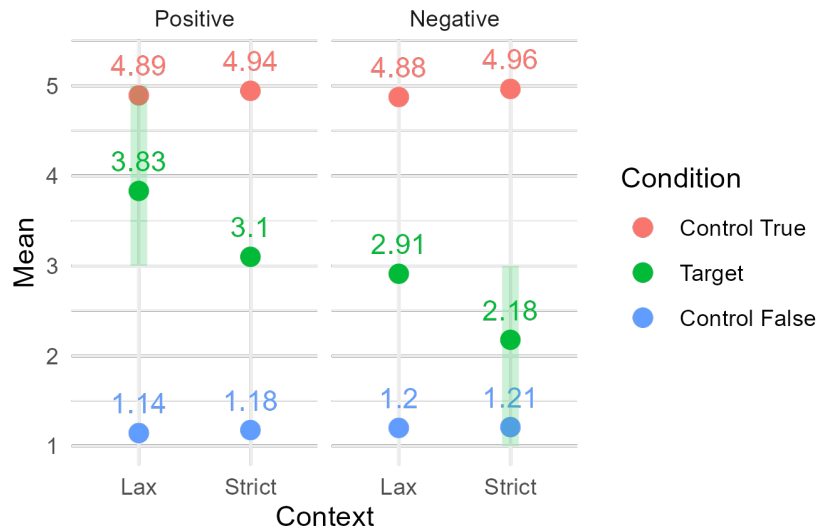


Figure 3: Mean ratings in Experiment 1. Error bars indicate interquartile range.

3.1.3 Results

Data treatment Responses above 3 (i.e., 4 and 5) for CONTROL TRUE and responses below 3 (i.e., 1 and 2) for CONTROL FALSE are considered correct responses. 2 participants were excluded from the data analyses as they didn't reach the pre-established 75% threshold for the accuracy rate of CONTROL trials. The mean accuracy rate for the remaining 89 participants is 98% (95%CI = [96.8, 98.1]).

Data analyses The mean ratings are given in Figure 5 ($n = 89$).⁹ Note that CONTEXT is recoded in terms of LAX and STRICT, as the two contexts should have opposite effects on the positive and negative sentences. Each context is coded as LAX if it is expected to facilitate the WEAK reading (EXISTENTIAL for POSITIVE, UNIVERSAL for NEGATIVE), and STRICT if it is expected to make it harder to access (UNIVERSAL for POSITIVE, EXISTENTIAL for NEGATIVE under the homogeneity approach).

⁹ The mean on a Likert scale does not receive a straightforward interpretation without the assumption of equal intervals, as in our case, but they are still useful in visually grasping the presence or absence of interaction effects. The statistical models we employed do not reason about means. For the same reason, we avoided representing standard deviation or standard error in the graphs, and used interquartile range to indicate variance. Note that when trivial interquartile ranges, which arise when half or more of the data points are at the median, are not made visible in the graphs.

We fitted a mixed-effects ordinal logistic regression model using the ordinal package (Christensen 2023) for R ({R Core Team} 2024) to the TARGET data with CONTEXT (sum-coded) and POLARITY (treatment-coded with NEGATIVE as the reference level) and their interaction as fixed effects and the maximal converging random-effects structure justified by the experimental design. In the present case, the maximal converging model included a random slope for POLARITY by subjects, with no random intercept. The model indicates a significant simple effect of CONTEXT ($\chi_2(1) = 70, p < .001$) and a significant main effect of POLARITY ($\chi_2(1) = 66, p < .001$), but no significant interaction effect between CONTEXT and POLARITY ($\chi_2(1) = 1, p = 0.23$).

3.1.4 Discussion

The results suggest that WEAK interpretations of bare plurals in simple sentences are context-sensitive, and the effect of context is symmetric with respect to polarity. When taken at face value, this is more in line with the homogeneity approach than with the implicature approach.

3.2 Experiment 2: Quantified sentences

Experiment 2 tested quantified sentences like (21) against a uniformly singular scenario where each individual opened exactly one present. All other aspects of the design are identical to Experiment 1.

- (21) a. Every boy opened presents.
b. No boy opened presents.

3.2.1 Data availability

Stimuli, data, and analysis code for Experiment 2 are available on the OSF at https://osf.io/zqanx/?view_only=48dbea9761f240c381a7045a82ce5684.

3.2.2 Methods

Participants We recruited 96 novel self-reported English speakers (mean age: 40.8 years old; age range: 18-75; 48 female, 48 male) online via Prolific using the same criteria as in Experiment 1. The average completion time was about 6 minutes and participants were paid £1.5. The consent and data collection procedures were consistent with those of Experiment 1.

Materials Each visual stimulus contained four children, each paired with two presents. There are three types of test pictures for each sentence. For POSITIVE sentences, the CONTROL TRUE pictures had each child with two open presents and the CONTROL FALSE pictures had each child with two closed presents; and the TARGET pictures had each child with one open present and one closed present. The exact same pictures were used for NEGATIVE sentences, except that the truth-value of the CONTROL pictures were reversed. Figure 4 contains example pictures used in Experiment 2.

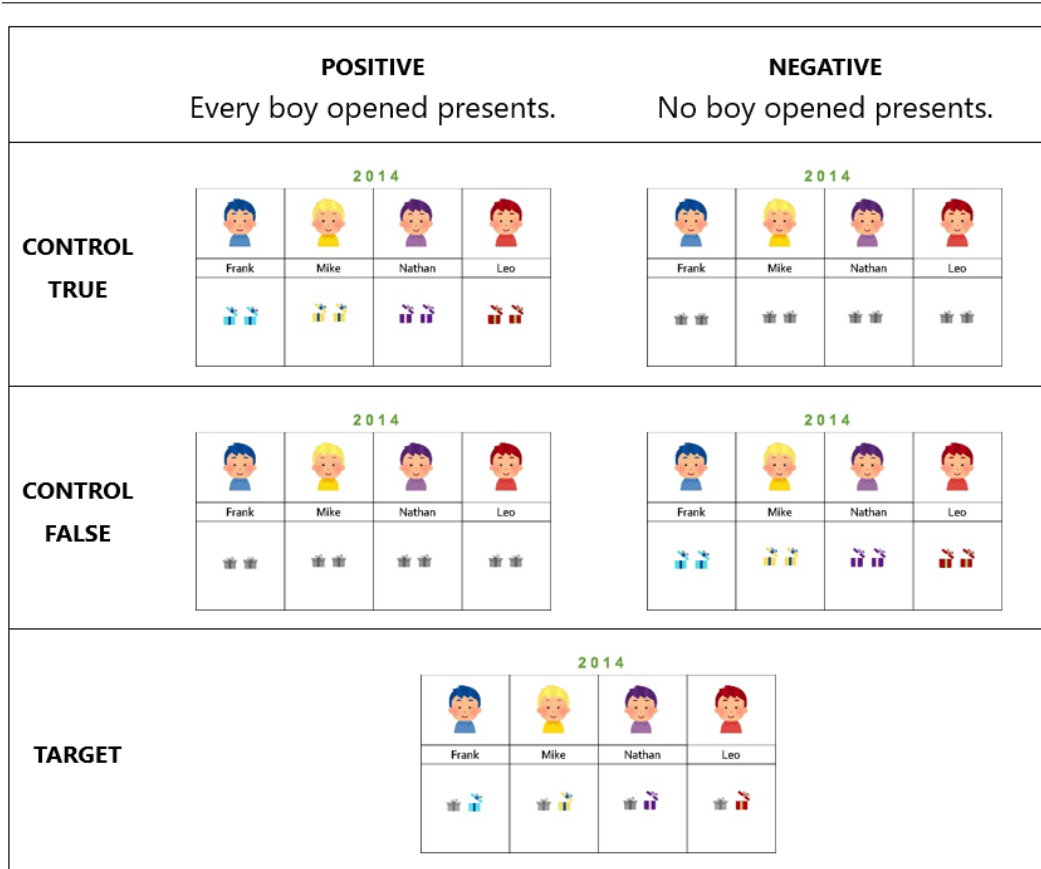


Figure 4: Examples of CONTROL TRUE, CONTROL FALSE, and TARGET pictures for POSITIVE and NEGATIVE conditions in Experiment 2.

The configuration of the attention question is similar to that in Experiment 1. There were two types of target sentences according to their POLARITY: ‘Every boy/girl opened presents’ in the POSITIVE condition; and ‘No boy/girl opened presents’ in the NEGATIVE condition. The configuration of the 5-point Likert scale is similar to that in Experiment 1.

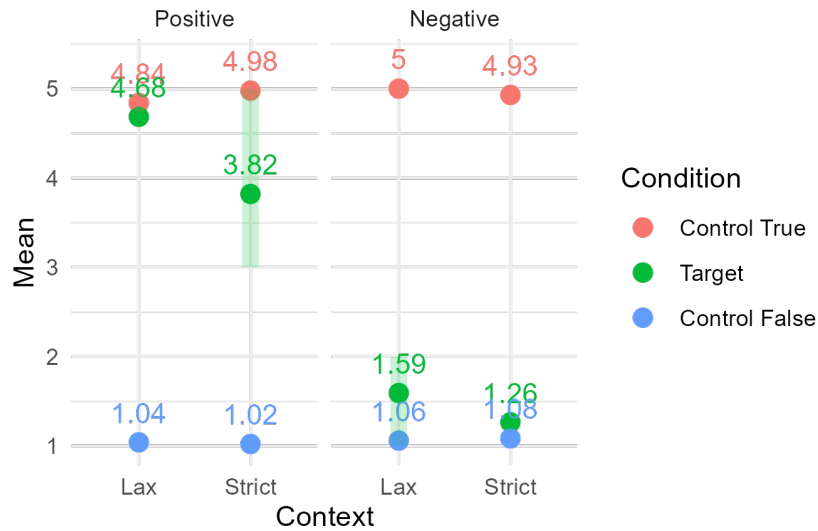


Figure 5: Mean ratings in Experiment 2. Errorbars indicate interquartile range.

Procedure The instructions and procedures were similar to those in Experiment 1. See Appendix A and Appendix B for instructions and practice trials.

3.2.3 Results

Data treatment As in Experiment 1, responses above 3 (i.e., 4 and 5) for CONTROL TRUE and responses below 3 (i.e., 1 and 2) for CONTROL FALSE are considered correct responses. 2 participants were excluded from the data analyses as they didn't reach the pre-established 75% threshold for the accuracy rate of CONTROL trials. The mean accuracy rate for the remaining 94 participants is 98% (95%CI = [96.8, 98.1]).

Data analyses The mean ratings are given in Figure 5 ($n = 94$). The analysis procedure is similar to Experiment 1. In R ([R Core Team](#) 2024), we fitted a mixed-effects ordinal logistic regression model to the TARGET data with CONTEXT (sum-coded) and POLARITY (treatment-coded with NEGATIVE as reference level) and their interaction as fixed effects with the maximal converging random-effects structure justified by the experimental design. In the current case, the maximal random-effects model that converged included random intercepts for subjects and random slopes for POLARITY grouped by subjects as well as their correlation. The model indicates a significant simple effect of CONTEXT ($\chi^2(1) = 15, p < .001$),

Putting bare plurals into context

a significant main effect of POLARITY ($\chi_2(1) = 393, p < .001$), and a significant interaction effect between CONTEXT and POLARITY (*every* is more sensitive than *no*; $\chi_2(1) = 10, p < 0.01$).

3.2.4 Discussion

Unlike in Experiment 1, the results revealed an asymmetry in the effect of context for WEAK interpretations of positive and negative quantified sentences. This is more in line with the predictions of the implicature approach.

3.3 Experiment 3: Quantified positive vs. negative sentences against mixed scenarios

The aim of Experiment 3 is to test the same quantified sentences as Experiment 2 against partial plurality scenarios. We prepared two types of such scenarios, which we call [o] and [xo].

- (22) [o]
- a. POSITIVE: Two children each received two presents and opened both; the other two children each received one and opened it.
 - b. NEGATIVE: Two children each received two presents and did not open either; two other children each received one and opened it.
- (23) [xo]
- a. POSITIVE: Each child received two presents. Two children opened both; the other two opened only one.
 - b. NEGATIVE: Each child received two presents. Two children opened neither; the other two opened only one.

Example pictures depicting these are given in Figure 6

The reason why we tested both of these cases of partial plurality inferences is because the homogeneity approach, but not the implicature approach, makes different predictions about them. In particular, notice that with respect to the UNIVERSAL context, every child obeys the family rule ('You must open each present you received!') in the [o]-POSITIVE condition but only two children do in the [xo]-POSITIVE condition. Similarly, every child flouts the family rule in the [xo]-NEGATIVE condition, but only two children do in the [o]-NEGATIVE condition. In all the other cases, the [o] and [xo] conditions are equivalent with respect to the number of children obeying or flouting the family rule.

Now, recall that the implicature approach predicts the positive sentence to have a (STRONG) reading that is simply true in such partial scenarios, and the negative

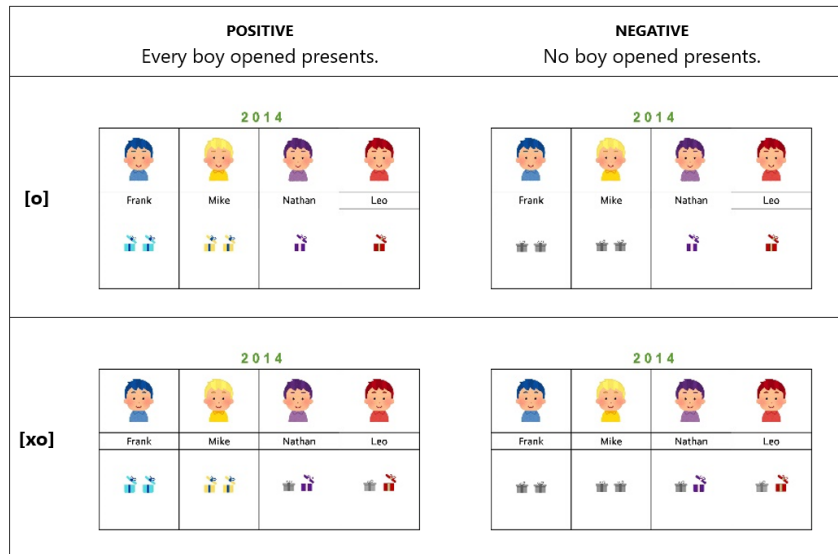


Figure 6: Examples of TARGET pictures in Experiment 3

sentence to lack such a reading. As such, no effect of context is expected, and no difference between [o] and [xo] is expected.

On the other hand, according to the homogeneity approach, both positive and negative sentences can be judged as pragmatically ‘true enough’ in such partial plurality scenarios. In particular, it is expected that it would matter whether or not the family rule is obeyed by every child in the positive case, and flouted by every child in the negative case, because these are the cases where the pragmatics of conversational issues should be able to render the sentence pragmatically ‘true enough’. Therefore, not only effects of context comparable to what was observed in Experiment 2 are expected, but also they should manifest themselves differently between the [o] and [xo] conditions.

3.3.1 Data availability

Stimuli, data, and analysis code for Experiment 3 are available on the OSF at https://osf.io/zqanx/?view_only=48dbea9761f240c381a7045a82ce5684.

3.3.2 Methods

Participants We recruited 94 novel participants (mean age: 40.5 years old; age range: 18-75; 49 female, 44 male, 3 other) online via Prolific using the same criteria as in Experiments 1 and 2. The average completion time was about 10 minutes and

Putting bare plurals into context

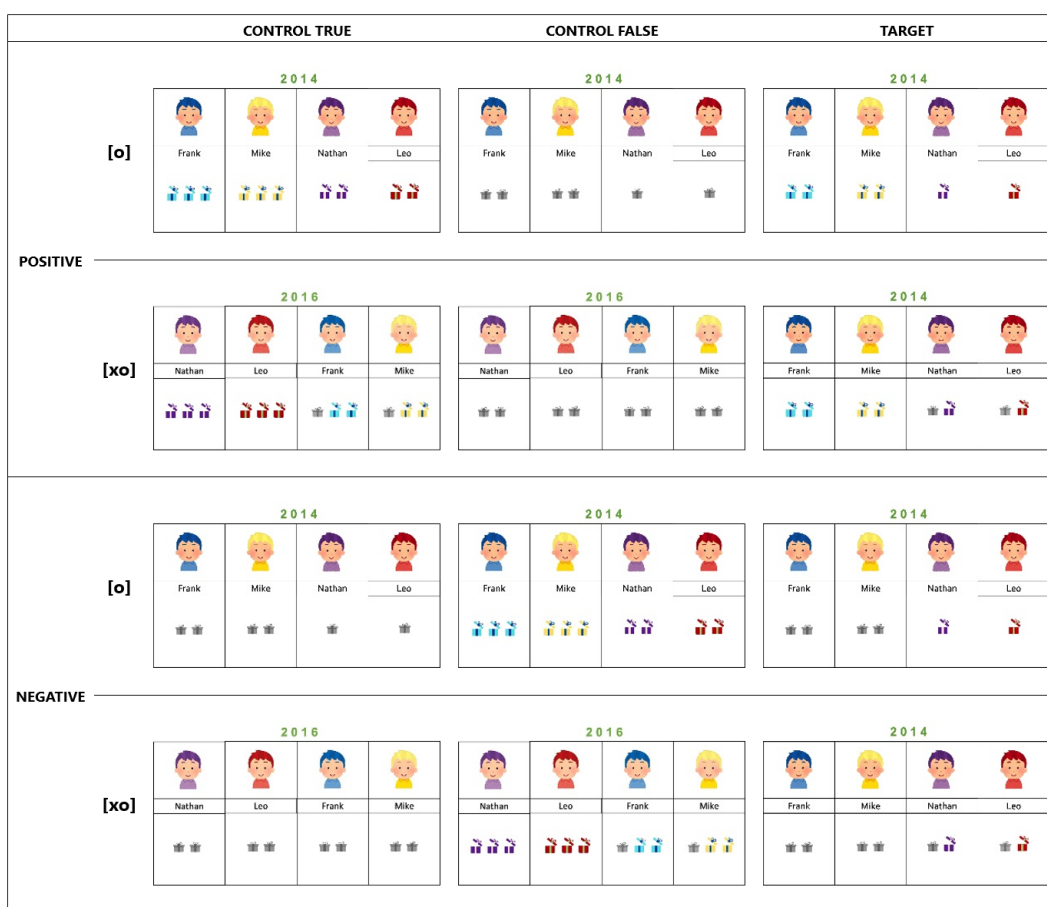


Figure 7: Examples of CONTROL TRUE, CONTROL FALSE, and TARGET pictures used in Experiment 3.

participants were paid £2.5. The consent and data collection procedures were the same as those in Experiments 1 and 2.

Materials The materials used in Experiment 3 were analogous to those in Experiment 2: the configurations of the context pictures, the attention questions, and the target sentences were identical to those in Experiment 2. Figure 7 contains example pictures used in Experiment 3.

Design We included SCENARIO as a within-subject factor. Consequently, there were $2 \text{ POLARITY} \times 3 \text{ CONDITION} \times 2 \text{ SCENARIO} \times 4 \text{ repetitions} = 48$ trials in total in each list. All the other aspects of the design were similar to Experiments 1 and 2.

Procedure The procedure was consistent with that in Experiments 1 and 2. Instructions and practice trials can be found in Appendix A and Appendix B, respectively.

3.3.3 Results

Data treatment As in the other experiments, responses above 3 (i.e., 4 and 5) for CONTROL TRUE and responses below 3 (i.e., 1 and 2) for CONTROL FALSE are considered correct responses. 3 participants were excluded from the data analyses in Experiment 3 because their accuracy rate in control trials failed to reach the pre-established threshold 75%. The mean accuracy rate of the remaining 91 participants was 96.4% (95%CI = [95.6, 97.1]).

Data analyses The results are shown in Figure 8 ($n = 91$). In R ([R Core Team 2024](#)), we fitted a model similar to Experiments 1 and 2 to the TARGET data for POSITIVE and NEGATIVE sentences separately. The model includes CONTEXT (sum-coded) and SCENARIO (treatment-coded with [o] as reference level) and their interaction as fixed effects within the maximal converging random effects structure justified by the design. In the current case, the converged maximal random effects structure includes random slopes for SCENARIO grouped by subjects without random intercepts. For POSITIVE conditions, the model indicates a non-significant simple effect of CONTEXT ($\chi_2(1) = 3, p = 0.08$) and a significant main effect of SCENARIO ($\chi_2(1) = 25, p < .001$), as well as a significant interaction effect between CONTEXT and SCENARIO ($\chi_2(1) = 6, p < .05$). For NEGATIVE conditions, the model indicates a significant simple effect of CONTEXT ($\chi_2(1) = 5, p < .05$) and a non-significant main effect of SCENARIO ($\chi_2(1) = 2, p = 0.13$), as well as a non-significant interaction effect between CONTEXT and SCENARIO ($\chi_2(1) = 0.38, p = 0.54$).

3.3.4 Discussion

The results show that positive quantified sentences received higher ratings against the [o] scenario against the [xo], whilst negative quantified sentences exhibited no such sensitivity to the [o] vs. [xo] distinction and generally received very low ratings. Thus, there is a stark positive vs. negative asymmetry, and this is contrary to the prediction of the homogeneity approach. The results are therefore more in line with the implicature approach. Yet, the difference between [o] and [xo] in positive sentences calls for an explanation.

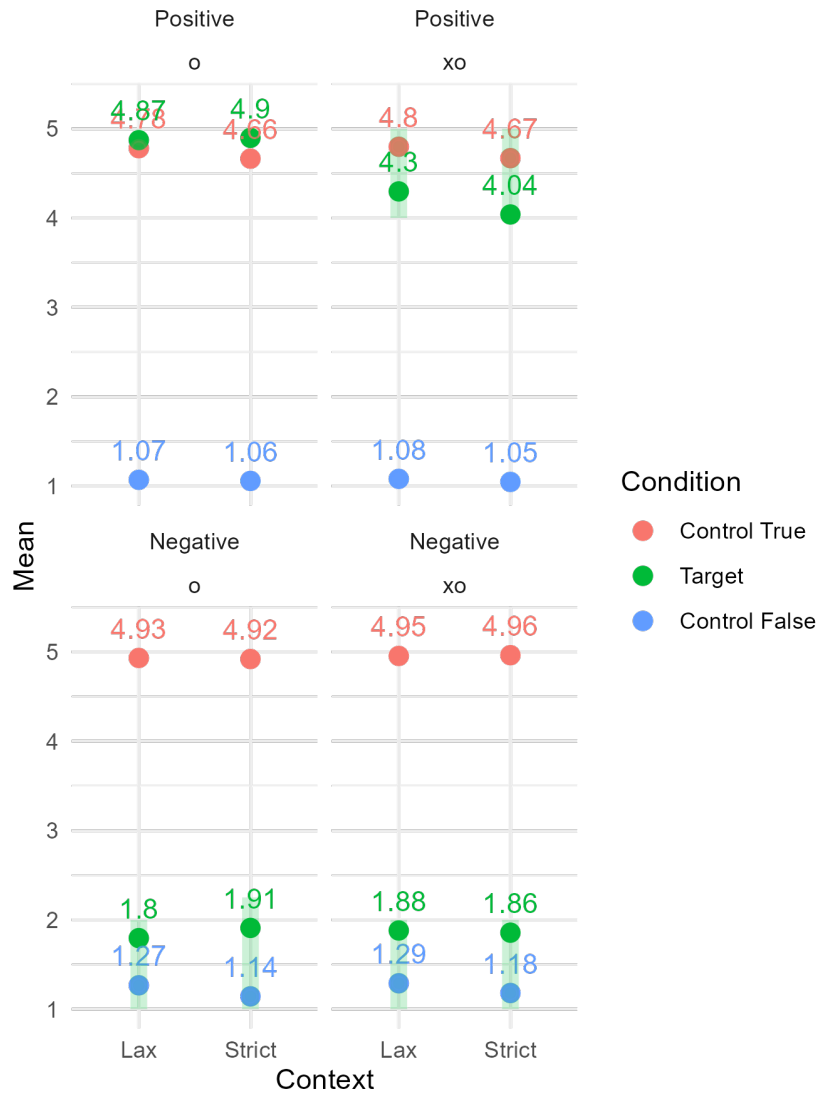


Figure 8: Mean ratings in Experiment 3. Error bars indicate interquartile range.

4 General discussion

To summarize the main findings of the three experiments, the results of Experiment 1 indicated a symmetric effect of context for WEAK interpretations of bare plurals in simple sentences with respect to polarity, which is more in line with the homogeneity approach. In contrast, the results of Experiment 2 showed an asymmetric effect of context for WEAK interpretations of bare plurals under *every* and *no*, which is more in

line with the implicature approach. Thus, neither approach would straightforwardly explain the results of the first two experiments.

The results of Experiment 3 provide more decisive evidence for one approach. Whilst we observed some effects of context for the truth-value judgment of positive quantified sentences with *every* against partial plurality scenarios, no comparable effect was observed for negative quantified sentences with *no*. Focusing first on the NEGATIVE condition, we did observe an effect of context in the [o]-NEGATIVE condition, but the effect went in the opposite direction with respect to the prediction of the homogeneity approach such that the sentence received a higher reading in the EXISTENTIAL/STRICT condition (with a mean of 1.91) than in the UNIVERSAL/LAX condition (with a mean of 1.8). Since the effect size here is small, relative to the other contrasts we observed across the three experiments, we can consider this effect to be of no particular theoretical importance. Our statistical analysis of the NEGATIVE condition indicated no other significant effect. We should certainly be careful in drawing conclusions from the absence of significance, but given the magnitude of effects in other conditions across the three experiments, it is reasonable to conclude that bare plurals in negative quantified sentences with *no* do not exhibit the context-sensitivity predicted by the homogeneity approach against partial plurality scenarios. In contrast, the implicature approach simply predicts that the negative quantified sentences do not have what corresponds to a partial plurality reading of their positive counterparts with *every*, so it fares better in explaining the results.

Turning now to the POSITIVE condition of Experiment 3, we observed a significant main effect of SCENARIO ([o] vs. [xo]) such that the [o] condition received higher ratings, as well as a significant interaction effect such that a bigger effect of SCENARIO was observed for [xo]. The interaction effect here could well be due to the ceiling effect in the [o] condition: in both contexts, the mean rating was quite high, with the means at 4.87 in [o] and at 4.90 in [xo], with essentially no variance. The data therefore does not contradict the hypothesis that context has the same effect in [o] and [xo]. However, the implicature approach still has to explain the effect of context. Recall that the implicature approach postulates a STRONG interpretation of the universally quantified sentence that is simply true in both [o] and [xo] scenarios, which we called Global, along with another STRONG interpretation, Local, which is false in both [o] and [xo] scenarios.

- (24) Every boy opened presents.
- a. Local: Every boy opened more than one present.
 - b. Global: Every boy opened at least one present, and some boys opened more than one present.

We claim that by making recourse to these two STRONG interpretations, the implicature approach is capable of explaining the effect of context as follows. Previous experimental research on sentences that similarly have two (or more) options for scalar implicatures has observed that the truth-value judgments of such sentences sometimes degrade against scenarios that make one of these scalar implicatures true but the other one false, relative to scenarios that make both of them true (Chemla & Spector 2011). Given this, we further hypothesize that the magnitude of this degradation effect depends on to what extent the false scalar implicature is relevant in the context. In particular, with the assumption that the EXISTENTIAL/LAX context makes the local reading (24a) less relevant—which is intuitive, as the rule is ‘You must keep each of your presents closed!’—than the UNIVERSAL/STRICT context, the effect of context we observed. In addition, this hypothesis about the contextual relevance of the false scalar implicature explains the observed main effect of SCENARIO such that higher ratings were given in the [o] condition than in the [xo] condition. In the [o] condition, some boys received only one present so the local reading in (24a) could not have been made true no matter how the boys had behaved, allowing participants to reason that it is certainly not what is meant so is not contextually relevant, whilst in the [xo] condition, it could have been true had the boys opened the closed presents, so it is contextually more relevant.

If the above explanation of the results of Experiment 3 is on the right track, the effects of context observed in Experiment 3 should be of a different nature than the effect of context observed in Experiment 2, which is about WEAK interpretations. This in fact is not far-fetched as the effect size of context manipulation on *every* was much larger in Experiment 2 than in Experiment 3. Conversely, this difference in effect size would be a challenge to explain for the homogeneity approach. Recall that for the homogeneity approach Experiments 2 and 3 were both about WEAK interpretations, and given that the linguistic stimuli as well as the contexts were identical in the two experiments with the only difference being whether the visual stimuli contained children that opened two presents, the different effect size would have to be attributed to this difference in the visual stimuli. Such an explanation would arguably lack cogency.

The remaining observation to be accounted for under the implicature approach is the symmetry in the results of Experiment 1. One analytical possibility is that the bare plural *presents* (and the existential quantifier associated with it) in the negative simple sentence could actually outscope the negation, contrary to the assumption that (simple) bare plurals always obligatorily take narrow scope. For (25), for example, this reading would be true if there is at least one present that Frank didn’t open.

(25) Frank didn’t open presents.

Furthermore, if the bare plural does outscope the negation, it is in fact in a positive grammatical environment, so it should be able to generate a plurality inference as a scalar implicature via competition with the singular alternative (26), where *a present* takes scope over negation.

(26) Frank didn't open a present.

The resulting reading would then amount to that there is more than one present that Frank didn't open. In the target trials, the visual scene depicted a situation with one open present and one closed present, so if the scalar implicature is computed, the sentence should be judged as false. If, on the other hand, the singular alternative in (26) is somehow deemed contextually irrelevant, the scalar implicature becomes cancellable and the sentence should be accepted. Assuming that the wide scope interpretation of the bare plural was very easily accessed and also that the contexts had analogous effects on the singular alternatives of the positive and negative sentences, the symmetric effect of context could be explained under the implicature approach.

Note, however, if such a wide scope reading had also been available in Experiment 2, we would lose an explanation for the asymmetric results. Therefore, it would have to be assumed that the bare plural could not take scope over the subject negative quantifier in sentences like 'No boy opened presents.'

In sum, one analytical possibility we've offered so as to make the implicature approach compatible with the results of the first two experiments is by assuming that bare plurals can take scope over clausemate negation but not over the subject. This account makes a testable prediction, namely, we should observe a difference between simple negative sentences judged against scenarios where one present is open and one present is closed, which we tested in Experiment 1, and scenarios where one present is open and two presents are closed. In the latter case, the plurality inference of the wide scope bare plural would be true, so the negative sentence is predicted to receive a much higher rating, comparable to what the positive simple sentence should receive against scenarios where two presents are open and one present is closed, and since in both of these cases STRONG interpretations are true, we would expect no effect of context. We will leave an investigation of this prediction for future research.

5 Conclusion

To conclude, the implicature approach to plurality inferences surpasses the homogeneity approach in explaining the results of Experiment 2 and Experiment 3. The results of Experiment 1, however, pose a challenge for the implicature approach. We offered one possible account and a further experimental idea to test its prediction.

Before closing, we would like to briefly remark on the similarities and differences with respect to the results reported in Romoli et al. 2024, who tested the context-sensitivity of non-maximal interpretations of definite plurals in the scope of *every*, *no*, *not every*, and *exactly two*—which were the quantifiers tested in a previous experimental study, Križ & Chemla 2015—using the same experimental methodology as this paper, including the two contexts with family rules (but see fn. 6 for a slight difference). Let us zoom in on the first two quantifiers, whose results are more or less directly commensurable with the results of our experiments. Romoli et al. tested the following sentences against ‘non-maximal scenarios’ where each boy received 9 presents, two of them opened some but not all, whilst the other two opened all (for *every*) or none (for *no*).

- (27) a. Every boy opened his presents.
b. No boy opened his presents.

Similarly to the results from our Experiment 2, they observed an asymmetric effect of context manipulation such that the non-maximal interpretation of (27a) shows a greater discrepancy between the two contexts than (27b). This might at first seem to support a uniform view where the mechanism for non-maximal interpretations of definite plurals is the same as the mechanism for WEAK interpretations of bare plurals. Recall Križ (2017) proposed such a uniform view, but since we concluded that what is behind WEAK and STRONG interpretations of bare plurals is the mechanism for scalar implicatures, it will be different from Križ 2017.

Furthermore, there are certain details that point to the opposite conclusion, favouring a non-uniform view. Firstly, the observed effect size of context manipulation is larger for definite plurals than for bare plurals. Specifically, in our Experiment 2, the mean ratings in the *every* condition were 4.68 (LAX) vs. 3.82 (STRICT), whilst in Romoli et al. 2024, they were 3.76 (LAX) vs. 2.63 (STRICT) in their Experiment 1 and 3.94 (LAX) vs. 2.34 (STRICT) in their Experiment 2. Similarly, the mean ratings in the *no* condition were 1.59 (LAX) vs. 1.26 (STRICT) in our Experiment 2, whilst they were 2.30 (LAX) vs. 1.80 (STRICT) in their Experiment 1. Such comparisons across experiments must certainly be understood with a grain of salt, and difference in effect size between definite and bare plurals should ideally be tested in a single controlled experiment, but these numbers suggest the possibility that different mechanisms are responsible for non-maximal readings of definite plurals and WEAK interpretations of bare plurals.

Secondly, Romoli et al. (2024) argue that their experimental findings overall favour a view according to which it is *not* implicatures that are behind non-maximal readings of definite plurals. Although we cannot go into details of their findings and reasoning here, given our conclusion that the results reported in this paper

support the implicature theory of STRONG and WEAK interpretations of bare plurals, it appears reasonable to tentatively conclude that non-maximal interpretations of definite plurals and WEAK interpretations of bare plurals are brought about by distinct mechanisms, which runs counter to the core tenet of the uniform view.

A Instructions in Experiments 1–3

A.1 General instructions in Experiments 1–3

In the following, you will learn about the festive habits of the Smith family. Each year, they give presents to each other at three different events: Christmas, Easter, and what they call the ‘Family Day’. The Smiths have particular habits about when exactly during these days the children open presents.

During the session, you will see pictures that show for one boy whether he has already opened presents. Below the pictures, you will see sentences that are intended to describe the particular situations shown in the pictures.

Your task will be to **judge whether these sentences accurately describe the situations shown in the pictures, given the Smiths’ specific habits.**

Before the main session begins, you will now be familiarised with the Smiths and their habits. After that, you will perform a short practice trial, and then, we begin!

A.2 Introduction of the family rule for EXISTENTIAL conditions in Experiment 1–3

These are the Smiths. They live in Perth in Scotland, and have four boys. Usually, the boys’ grandparents arrive at noon from their journey from Pitlochry to celebrate together.

The Smiths always open presents together with their grandparents, and the boys have to wait until they arrive.

So this is the Smiths’ rule:

Family rule: You must keep each present you received closed!

Though this rule is clear, this is not what always happens. In the following, you will see different scenarios from recent years, in which the children sometimes obeyed the rule, and sometimes, they didn’t.

On the following screen, you will, for instance, see what happened at one of the festivities in 2018. You are reminded of the rule on the top left of the screen, and then, what one of the Smith boys actually did is shown on the right.

Putting bare plurals into context

Given this information, your task is the following: first, you will decide **whether the boy followed the family rule**. After that, you will be asked **how truly the sentence at the bottom describes the scenario in the picture**.

A.3 Introduction of the family rule for UNIVERSAL conditions in Experiments 1–3

These are the Smiths. They live in Perth in Scotland, and have four boys. Usually, the boys' grandparents arrive at noon from their journey from Pitlochry to celebrate together.

The Smiths always open the presents in the early morning before their grandparents arrive, so that the apartment is in an orderly state by then.

So this is the Smiths' rule:

Family rule: You must keep each present you received opened!

Though this rule is clear, this is not what always happens. In the following, you will see different scenarios from recent years, in which the children sometimes obeyed the rule, and sometimes, they didn't.

On the following screen, you will, for instance, see what happened at one of the festivities in 2014. You are reminded of the rule on the top left of the screen, and then, what each of the Smith boys actually did is shown on the right.

Given this information, your task is the following: first, you will decide **whether the boys followed the family rule**. After that, you will be asked **how truly the sentence at the bottom describes the scenario in the picture**.

B Practice trials with instructions in Experiments 1–3

B.1 Experiment 1

Was the rule respected? In the right picture, you see the state of affairs about Frank before the guests arrive. Now please indicate whether or not the family rule was respected by Frank in the preceding scenario. Please click on 'Yes' or 'No' below before answering the next question.

Frank didn't open presents. You now have to judge on the scale below how truly the sentence 'Frank didn't open presents' is describing this situation. Note that you can choose any value on the scale from 'Completely true' to 'Completely false' which you think is appropriate. Please click on one of the five values (boxes) on the scale below.

B.2 Experiments 2 and 3

Was the rule respected? In the right picture, you see the state of affairs in the Smith house before the guests arrive. Now please indicate whether or not the family rule was respected by each of the boys in the preceding scenario. Please click on ‘Yes’ or ‘No’ below before answering the next question.

Leo didn’t open presents. You now have to judge on the scale below how truly the sentence ‘Leo didn’t open presents’ is describing this situation. Note that you can choose any value on the scale from ‘Completely true’ to ‘Completely false’ which you think is appropriate. Please click on one of the five values (boxes) on the scale below.

References

- {R Core Team}. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. <https://www.R-project.org/>.
- Bar-Lev, Moshe E. 2021. An Implicature account of Homogeneity and Non-maximality. en. *Linguistics and Philosophy* 44(5). 1045–1097. <https://doi.org/10.1007/s10988-020-09308-5>.
- Carlson, G. 1977. *Reference to kinds in English*. University of Massachusetts at Amherst Ph.D. <https://www.semanticscholar.org/paper/Reference-to-kinds-in-English-Carlson/03076a488a0cb8aaef336130eb4b3108533bd75> (21 June, 2023).
- Chemla, Emmanuel & Benjamin Spector. 2011. Experimental Evidence for Embedded Scalar Implicatures. *Journal of Semantics* 28(3). 359–400. <https://doi.org/10.1093/jos/ffq023>.
- Christensen, Rune H. B. 2023. *ordinal—regression models for ordinal data*. manual. <https://CRAN.R-project.org/package=ordinal>.
- Crnič, Luka, Emmanuel Chemla & Danny Fox. 2015. Scalar implicatures of embedded disjunction. en. *Natural Language Semantics* 23(4). 271–305. <https://doi.org/10.1007/s11050-015-9116-x>.
- Dayal, Veneeta. 2011. Bare Noun Phrases. In Claudia Maienborn, Klaus von Stechow & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 1. Walter de Gruyter.
- Delfitto, Denis. 2006. Bare Plurals. en. In *The Blackwell Companion to Syntax*, 214–259. Section: 8_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470996591.ch8>. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470996591.ch8>.
- Diesing, Molly. 1992. Indefinites. en. *Journal of Linguistics* 30(2). 570–571. <https://doi.org/10.1017/S0022226700016832>.

- Grimm, Scott. 2011. Plurality is Distinct from Number Neutrality. In Yelena Fainleib, Nicholas LaCara & Yangsook Park (eds.), *Proceedings of NELS 41*, 247–258. Amherst, MA. <https://www.semanticscholar.org/paper/Plurality-is-Distinct-from-Number-Neutrality-Grimm-Condoravdi/f627ebad2789da46e74cc93d6cddf85523b6133a> (21 June, 2023).
- Guerrini, Janek & Jad Wehbe. 2024. *Homogeneity as presuppositional exhaustification*. LingBuzz Published In: <https://lingbuzz.net/lingbuzz/007957> (12 November, 2024).
- Ivlieva, Natalia. 2013. *Scalar Implicatures and the Grammar of Plurality and Disjunction*. en. Massachusetts Institute of Technology Ph.D.
- Jasbi, Masoud, Brandon Waldon & Judith Degen. 2019. Linking Hypothesis and Number of Response Options Modulate Inferred Scalar Implicature Rate. *Frontiers in Psychology* 10. 189. <https://doi.org/10.3389/fpsyg.2019.00189>.
- Križ, Manuel. 2015. *Aspects of Homogeneity in the Semantics of Natural Language*. en. University of Vienna Ph.D.
- Križ, Manuel. 2016. Homogeneity, Non-Maximality, and all. *Journal of Semantics* 33(3). tex.ids= kriz_homogeneity_2016-1, 493–539. <https://doi.org/10.1093/jos/ffv006>.
- Križ, Manuel. 2017. Bare plurals, multiplicity, and homogeneity. *École Normale Supérieure*. <https://semanticsarchive.net/Archive/2U3YzUwZ/Kriz%202017%20Bare%20Plurals.pdf>.
- Križ, Manuel & Emmanuel Chemla. 2015. Two methods to find truth-value gaps and their application to the projection problem of homogeneity. en. *Natural Language Semantics* 23(3). 205–248. <https://doi.org/10.1007/s11050-015-9114-z>.
- Križ, Manuel & Benjamin Spector. 2021. Interpreting plural predication: homogeneity and non-maximality. en. *Linguistics and Philosophy* 44(5). 1131–1178. <https://doi.org/10.1007/s10988-020-09311-w>.
- Magri, Giorgio. 2009. A theory of individual-level predicates based on blind mandatory scalar implicatures. en. *Natural Language Semantics* 17(3). 245–297. <https://doi.org/10.1007/s11050-009-9042-x>.
- Mayr, Clemens. 2015. Plural definite NPs presuppose multiplicity via embedded exhaustification. en. *Semantics and Linguistic Theory*. 204–224. <https://doi.org/10.3765/salt.v25i0.3059>.
- Romoli, Jacopo, Yasutada Sudo, Petra Augurzky, Marion Bonnet, Richard Breheny, Alexandre Cremers, Cornelia Ebert, Markus Steinbach & Clemens Mayr. 2024. Putting plural definites into context. <https://ling.auf.net/lingbuzz/007987>.
- Spector, Benjamin. 2007. Aspects of the Pragmatics of Plural Morphology: On Higher-Order Implicatures. en. In Uli Sauerland & Penka Stateva (eds.), 243–281. Book Title: *Presupposition and Implicature in Compositional Semantics*. London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230210752_9.

- Sudo, Yasutada. 2023. Scalar implicatures with discourse referents: a case study on plurality inferences. en. *Linguistics and Philosophy*. <https://doi.org/10.1007/s10988-023-09381-6>.
- Tieu, Lyn, Cory Bill, Jacopo Romoli & Stephen Crain. 2020. Testing theories of plural meanings. en. *Cognition* 205. 104307. <https://doi.org/10.1016/j.cognition.2020.104307>.
- Zweig, Eytan. 2009. Number-neutral bare plurals and the multiplicity implicature. en. *Linguistics and Philosophy* 32(4). 353–407. <https://doi.org/10.1007/s10988-009-9064-3>.