To appear in Cognition

Testing theories of plural meanings*

Lyn Tieu[1], Cory Bill[2], Jacopo Romoli[3], Stephen Crain[4]

[1] *Western Sydney University, Office of the Pro Vice-Chancellor (Research and Innovation), School of Education, MARCS Institute for Brain, Behaviour and Development*
[2] *Universität Konstanz, Department of Linguistics*
[3] *University of Bergen, Department of Foreign Languages*
[4] *Macquarie University, Department of Linguistics*

Abstract

Plural morphology in English is associated with a multiplicity inference. For example, "Emily fed giraffe**s**" is typically interpreted to mean that Emily fed *multiple giraffes*. It has long been observed that this inference disappears in downward-entailing linguistic environments, such as in the scope of negation. For example, "Emily didn't feed giraffes" does not merely suggest that she didn't feed multiple giraffes, but rather that she didn't feed any. There are three main approaches to explaining this puzzle: the first proposes that the plural is *ambiguous*, and invokes a preference for stronger meanings; the second derives multiplicity inferences as implicatures; and the third provides a homogeneity-based account. These different approaches can all account for the interpretation of the plural across upward- and downward-entailing environments. They differ, however, in what they predict for three further aspects of the plural: the status of positive and negative plural sentences in singular contexts, children's acquisition of plural meanings, and the relationship between plural meanings and scalar implicatures. In this paper, we report on three experiments investigating adults' and preschool-aged children's interpretation of plural morphology in English. The experiments reveal that participants distinguish positive and negative plural sentences presented in singular contexts, and that adults assign a different status to these positive and negative sentences. It is also observed that children, unlike adults, tend to accept underinformative *positive* plural sentences in singular contexts — in parallel with their behavior on standard scalar implicatures — while they are relatively more adult-like when it comes to negative plural sentences in the same contexts, showing a tendency to reject the negative sentences. We discuss how the findings of the three experiments are expected on a scalar implicature approach to multiplicity inferences, and the open challenges they pose for the ambiguity and homogeneity approaches.

For helpful feedback and discussion, we would like to thank Scott AnderBois, Emmanuel Chemla, Alexandre Cremers, Giorgio Magri, Uli Sauerland, Florian Schwarz, Jesse Snedeker, and Benjamin Spector. We would also like to thank Dorothy Ahn for allowing us to use her illustrations in our stimuli. Finally, we are grateful to the families who participated in the study at the CCD at Macquarie University. The research leading to these results was supported by the Australian Research Council Centre of Excellence in Cognition and its Disorders (CE110001021), the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610, ANR-10-IDEX-0001-02 PSL, ANR-10-LABX-0087 IEC, the Leverhulme Trust, grant RPG-2016-100, and Western Sydney University through the University's Research Theme Champion support funding.

Keywords: semantics; plural; multiplicity; implicatures; linguistic ambiguity; child language development

1 Introduction

1.1 The puzzle of multiplicity inferences

The plural-singular distinction is the source of a long-standing puzzle (see Sauerland 2003; Sauerland, Andersen & Yatsushiro 2005; Spector 2007; Zweig 2009; Farkas & de Swart 2010 and Magri 2014 for discussion). Consider the English sentence with a plural noun phrase (1a). This sentence seems equivalent in meaning to one in which the plural noun phrase is replaced by "more than one giraffe", as in (1b), and it differs in meaning from a sentence containing a singular noun phrase, (1c). These simple observations suggest that English plural morphology is associated with the meaning 'more than one' (Lasersohn 1995; Chierchia 1998, among others).

(1) a. Emily fed giraffes.
b. Emily fed more than one giraffe.
c. Emily fed a giraffe.

The paradox arises when the sentence with the plural noun (1a) appears under negation, as in (2a). This is because we would expect the negative sentence with the plural noun to yield a 'not more than one' reading, as in (2b). Instead, the sentence is better paraphrased as the negation of the corresponding sentence with the singular noun, (1c). That is, the better paraphrase of (2a) is (2c).

(2) a. Emily didn't feed giraffes.
b. Emily didn't feed more than one giraffe.
c. Emily didn't feed a (single) giraffe.

This is not an isolated fact. The absence of the expected 'more than one' meaning of a plural noun is also observed in other *downward-entailing* environments, that is, linguistic environments that license inferences from sets to subsets.[1] Again, in these environments, the plural morphology is better paraphrased using singular morphology than it is by substituting 'more than one'. That is, examples (3a) and (4a), containing plural morphology, are semantically equivalent to (3b) and (4b), containing singular morphology.[2]

[1]For example, the scope of the negative quantifier 'nobody' is downward-entailing and licenses subset inferences (i); the positive quantifier 'somebody' is instead upward-entailing on its nuclear scope and licenses superset inferences (ii).

(i) Nobody read books. ⇒ Nobody read linguistics books.

(ii) Somebody read linguistics books. ⇒ Somebody read books.

[2]The downward-entailing status of the antecedent of conditionals and of questions is admittedly controversial. What is relevant here, however, is that the antecedents of conditionals and questions generally pattern with downward-entailing environments with respect to the behavior of scalar inferences.

(3) a. If there are books on Stephen's desk, Robin should lock the door.
b. If there is a (single) book on Stephen's desk, Robin should lock the door.

(4) a. Are there books on Stephen's desk?
b. Is there a (single) book on Stephen's desk?

To restate the paradox: the interpretation of positive sentences with plural morphology gives rise to a 'more than one' inference, but this *multiplicity inference* typically disappears under negation, in the antecedent of a conditional, and in questions. This pattern is clearly problematic for a semantic account of multiplicity inferences, which simply encodes this inference in the literal meaning of plural morphology (e.g., Chierchia 1998; Lasersohn 1995).

Three main approaches to this puzzle have been proposed in the formal semantics literature. The first, defended in Farkas & de Swart (2010), is a semantic account that proposes a polysemous meaning for the plural and invokes a principle for selecting which meaning is preferred in a given context. The second is an implicature treatment of multiplicity inferences, implemented in different forms in Sauerland (2003), Sauerland, Andersen & Yatsushiro (2005), Spector (2007), Zweig (2009), Ivlieva (2013), and Mayr (2015). The third approach accounts for the different interpretations of the plural by appealing to homogeneity (Križ 2017). As we discuss below, these different approaches can all account for the interpretation of the plural across upward- and downward-entailing linguistic environments. They differ, however, in what they predict for the following three aspects of the plural: the status of positive and negative plural sentences in singular contexts, children's acquisition of the different interpretations of the plural, and the relationship between plural meanings and scalar implicatures.

In this paper, we report on three experiments investigating adults' and preschool-aged children's interpretation of plural morphology in English. The experiments reveal that adults assign a different status to positive and negative plural sentences presented in singular contexts. The experiments moreover show that children are relatively more adult-like in their interpretation of negative plural sentences, compared to their interpretation of positive plural sentences — a pattern similar to that observed for classical scalar implicatures. We discuss how the findings of the three experiments are more easily captured under a scalar implicature approach to multiplicity inferences than under ambiguity- or homogeneity-based approaches.

The rest of the paper is organized as follows. In the next section, we sketch the three accounts and discuss their predictions for adults' and children's behavior. In Section 2, we summarize two previous studies on the multiplicity inference and discuss how these previous results motivate the present experiments. We present our experiments in Sections 3, 4, and 5, and in Section 6, we discuss the results in the context of the theoretical predictions outlined in Section 1.3. Section 7 concludes the paper.

1.2 Theoretical approaches

1.2.1 The ambiguity approach

In response to the monotonicity puzzle described in the previous section, Farkas & de Swart (2010) enrich the semantic account in order to explain the sensitivity to monotonicity exhibited by multiplicity inferences. Simplifying, their approach is essentially based on two main ingredients: a polysemous meaning for plural morphology and a principle for choosing between two readings

of plural sentences in different contexts. In brief, the plural morpheme can have a weak *inclusive* meaning (compatible with both singular and plural entities) or a strong *exclusive* meaning (which excludes singularities). A plural sentence like (1a) will then correspondingly be associated with two possible readings, a weak reading along the lines of (5a) (compatible with Emily having fed a singularity or a plurality of giraffes) and a strong reading along the lines of (5b) (compatible only with Emily having fed more than one giraffe). The reading in (5b) is stronger than that in (5a), as it entails it — if (5b) is true, (5a) must necessarily also be true.

(5) a. Emily fed one or more giraffes. WEAK
b. Emily fed more than one giraffe. STRONG

Similarly, the negative (2a) is then associated with the two meanings in (6a) and (6b).

(6) a. Emily didn't feed one or more giraffes. STRONG
b. Emily didn't feed more than one giraffe. WEAK

The proposed ambiguity of the plural morphology alone will not account for the fact that (1a) tends to be interpreted as (5b) while (2a) tends to be interpreted as (6a). To explain this pattern, Farkas & de Swart (2010) assume a version of the Strongest Meaning Hypothesis, formulated as in (7), which essentially regulates the choice between the two possible readings of plural sentences, favoring the strongest between the two (see Dalrymple, Kanazawa, Kim, Mchombo & Peters 1998; Winter 2001, among others).

(7) *The Strongest Meaning Hypothesis for Plurals*: for a sentence involving a plural nominal, prefer that interpretation of plural which leads to the stronger overall interpretation for the sentence as a whole, unless this interpretation conflicts with the context of utterance.

As can clearly be seen, (7) will favor the interpretation (5b) for the positive (1a) and (6a) for the negative (2a), as these correspond to the strongest possible interpretations of these sentences. (7) therefore allows the ambiguity approach to capture the monotonicity effect.

In addition, the principle in (7) allows for the weaker interpretations to re-emerge if the stronger ones are in conflict with the context of utterance. This predicts that we should be able to force a weak interpretation by adding a continuation that is in conflict with the strong counterpart. This is indeed the case, as illustrated in (8); here the negated plural receives the weak interpretation along the lines of (6b), rather than the strong interpretation in (6a).

(8) Emily didn't feed giraffe**s**, because she fed only one!

In sum, the ambiguity approach can explain the different readings of the plural and their distribution across upward- and downward-entailing contexts. In the next subsection, we turn to a different take on the puzzle, which involves treating multiplicity inferences as implicatures.

1.2.2 The implicature approach

A different response to the pattern associated with multiplicity inferences is to treat them as a type of implicature (Sauerland 2003; Sauerland, Andersen & Yatsushiro 2005; Spector 2007; Zweig 2009; Ivlieva 2013; Mayr 2015). This is because implicatures, and in particular scalar implicatures, exhibit a similar monotonicity pattern. To illustrate, a familiar example of a scalar implicature

is (9). In interpreting a positive sentence containing a disjunction phrase "A or B", we typically infer the *exclusive* 'A or B but not both.' For example, (9a) implies (9b).

(9) a. Leo ate the apple or the orange.
b. ⇝ Leo did not eat both the apple and the orange.

When disjunction is embedded under negation, however, it is typically interpreted inclusively, rather than exclusively: (10a) is typically interpreted as (10b) and not as (10c); the latter corresponds to the negation of the exclusive meaning of disjunction and is compatible with Leo eating both the apple and the orange. But (10a) doesn't seem compatible with such a situation.[3]

(10) a. Leo didn't eat the apple or the orange.
b. ⇝ Leo didn't eat the apple and he didn't eat the orange.
c. ⇝ Leo either ate both the apple and the orange or he ate neither of them.

Both multiplicity inferences and classical scalar implicatures, then, share the property of arising in upward-entailing environments and disappearing in downward-entailing environments. This parallelism in their sensitivity to monotonicity has led some researchers to capture the 'more than one' meaning of the plural in the same manner as classical scalar implicatures.

A standard approach to deriving scalar implicatures like (9) is to treat them as arising from the hearer's reasoning about what the speaker actually said compared to what she could have otherwise said, assuming she was being a cooperative conversational partner (see Grice 1975 and much subsequent work). Simplifying, the implicature above would arise from the comparison of the original assertion with alternative assertions that could have been uttered but were not. In particular, a sentence with the weak scalar term "or" is compared to the stronger alternative sentence containing "and":

(11) a. Leo ate the apple **or** the orange.
b. Leo ate the apple **and** the orange.

Assuming the speaker is being as informative as she can be, the fact that she uttered the assertion containing "or" rather than the stronger, more informative alternative containing "and", invites the listener to conclude that the speaker was not in a position to assert the stronger alternative. If the listener further believes that the speaker is well-informed with respect to the alternative, she will infer that the stronger alternative must therefore be false; the inference in (9b) is thus entered into the listener's mental model of the conversational context.

This reasoning also naturally explains why (10a) doesn't give rise to an implicature and, as we observed, is instead interpreted as the negation of an inclusive disjunction. The corresponding alternative that the speaker might have instead uttered would be (12), but (12) is not more informative than (10a); it is in fact weaker than (10a). The listener therefore does not draw an implicature from the speaker's use of (10a).

(12) Leo didn't eat the apple **and** the orange.

The implicature approach to multiplicity inferences extends this proposal to plural sentences like

[3] There is a possible marked reading of (10a) that appears to be compatible with such a situation; we return to this below.

(1a) and (2a). Roughly, the idea is the following. A sentence like (1a), repeated below in (13), is assumed to unambiguously mean that Emily fed one or more giraffes. Upon hearing (13), the listener reasons about why the speaker didn't instead say something to the effect of (14).[4]

(13) Emily fed giraffes.

(14) Emily fed **exactly one** giraffe.

Given that (14) would have been more informative than (13), the listener concludes that the speaker must believe that (14) is false. But if (13) is true and (14) is false, the result is then exactly the multiplicity inference we are after, as illustrated in (15):

(15) Emily fed one or more giraffes and it's not true that she fed exactly one giraffe =Emily fed more than one giraffe.

This approach can also explain why, as with the disjunction case above, the negative (2a), repeated below in (16), is not associated with any inference. This is because the listener will compare (16) to its alternative in (17). (17), however, is weaker than (16) and therefore the listener will not draw any implicature from (16).

(16) Emily didn't feed giraffes.

(17) Emily didn't feed **exactly one** giraffe.

Finally, the scalar implicature approach can also capture the additional reading of the negated plural in (8), repeated below in (18). (18) typically involves stress or emphasis on the syllable containing the "-s" morpheme, and results in a reading that excludes the singularity. Notice that this kind of reading also arises with standard scalar items like disjunction; for instance, (19), when pronounced with stress on "or", also has a marked reading compatible with Leo eating both the apple and the orange. The implicature approach unifies the two cases, extending the same mechanism from the standard scalar (19) to the plural (18). Generally, the way of generating these marked readings is to postulate a local scalar implicature that is computed under the scope of the negation. Given that implicatures tend not to arise under negation, the readings in (18) and (19) are correctly predicted to be marked.

(18) Emily didn't feed giraffe**s**, because she fed only one!

(19) Leo didn't eat the apple **or** the orange, he ate both!

[4]Different versions of the implicature approach differ in how they derive the alternative that we have paraphrased in (14). For instance, the account in Spector (2007) involves higher order implicatures, where (14) is the corresponding singular version of (1a) enriched with its own implicatures. In Mayr (2015), Zweig (2009), and Ivlieva (2013), multiplicity inferences are derived by calculating an implicature at an embedded level of the sentence containing the plural. Our experiments will not hinge on a particular version of the scalar implicature approach, since the experimentalal hypothesis is consistent with all current scalar implicature-based accounts. Finally, an account based on an implicature reasoning over the assumed presuppositions of the singular versus plural forms, rather than a scalar implicature mechanism, has also been defended in Sauerland et al. (2005). The predictions of this account are very similar to the ones of the scalar implicature accounts we have mentioned, with the exception of what they predict regarding the relationship between multiplicity inferences and scalar implicatures. We will focus here on the predictions of the general implicature approach, leaving to future research a more fine-grained study of the predictions of different *versions* of the implicature approach.

1.2.3 The homogeneity approach

Plural definite descriptions such as "the giraffes" typically give rise to so-called *homogeneity* effects: the positive *The giraffes are tall* and the negative *The giraffes aren't tall* are neither true nor false when the group of giraffes is not homogeneous with respect to the property of being tall, i.e. when only some of the giraffes are tall (Löbner 1987; Schwarzschild 1994; Križ 2015). Križ (2015) provides an account of this homogeneity effect, and Križ (2017) extends it to account for the multiplicity inference of bare plurals. The general idea under the homogeneity approach is that predicates applied to pluralities can not only be true or false, but also *undefined* under certain conditions.

On the homogeneity theory, homogeneity arises as a lexical property of predicates, in combination with a general homogeneity principle. These two ingredients together yield trivalent truth conditions for sentences involving predicates applied to pluralities. For example, the nominal predicate *giraffes* has the meaning in (20):

(20) $[\![\text{giraffes}]\!] = \lambda x \left\{ \begin{array}{l} 1 \text{ iff } x \text{ is a plurality of giraffes} \\ 0 \text{ iff } x \text{ doesn't contain any giraffe} \\ \# \text{ otherwise} \end{array} \right\}$

As Križ (2017) shows, when a predicate like (20) appears in an episodic sentence such as (21), it gives rise to the following trivalent truth-conditions: it is true when both (22a) and (22b) are true, false when both are false, and undefined otherwise. This gives us the intuitively correct reading of the sentence in (21), namely that it's true if and only if Emily fed more than one giraffe.

(21) Emily fed giraffes.

(22) a. Emily fed one or more giraffes.
b. Emily fed multiple giraffes.

When (21) is negated, as in (23), the undefinedness is unaffected by negation, so that the conditions for (23) are as follows: (23) is true when both (24a) and (24b) are true, false when both are false, and undefined otherwise. These appear to be the correct conditions for (23) and in particular they capture the intuition that the sentence is true if and only if Emily didn't feed any giraffes.[5]

(23) Emily didn't feed giraffes.

(24) a. Emily didn't feed one or more giraffes.
b. Emily didn't feed multiple giraffes.

In sum, the homogeneity approach can account for the alternation between positive and negative cases.

For the cases of cancellation that we have seen above, Križ (2017) appeals to a pragmatic principle for dealing with undefinedness, which allows the use of a sentence even if undefined in the context, as long as the actual situation that makes the sentence undefined is equivalent to a situation that would make the sentence true. This principle is summarized in (25).

[5]Note that the negation assumed here is standard negation, making a true sentence false, a false sentence true, and keeping the undefinedness unaffected.

(25) An undefined sentence can be used when the situation described in the context is, for current purposes, equivalent to the situation in which the sentence is true. [Križ 2017, p. 22]

Returning to the example in (26), the prediction is that it can be used felicitously to the extent that we can accommodate in the context that the distinction between Emily feeding just one giraffe and her feeding no giraffes is irrelevant (i.e. it only matters whether she fed more than one giraffe or not).

(26) Emily didn't feed giraffes, because she fed only one!

In sum, all three approaches can account for the different readings associated with plural morphology in English and in particular for the observed sensitivity to monotonicity. In the next subsection, we turn to the different predictions they make for children and adults, which will motivate the experiments in Sections 3, 4, and 5.

1.3 Predictions

All three approaches are designed to capture the observation that multiplicity inferences arise in upward-entailing environments but generally not in downward-entailing environments. Under all three approaches, then, we should expect to observe an effect of monotonicity on the rate of multiplicity inference computation. The three approaches diverge, however, along three dimensions: the status of positive versus negative plural sentences in singular contexts, the predictions they make in relation to the acquisition of the readings of the plural, and the relationship between multiplicity inferences and implicatures more generally. We turn to each of these next.

1.3.1 The status of positive vs. negative sentences

As Križ (2017) points out, the status of positive vs. negative plural sentences is an important point of divergence in predictions among the alternative approaches to the plural. Consider the sentences in (27a) and (27b) in a context in which Emily fed only one giraffe.

(27) a. Emily fed giraffes.
b. Emily didn't feed giraffes.

The homogeneity and ambiguity approaches make symmetric predictions here. The former predicts both to be undefined. The latter predicts both to be false (or both to be true, if the Strongest Meaning Principle is overridden). The implicature approach, on the other hand, is compatible with — and in fact even predicts — a difference between the two polarities. This is because the positive case is literally true in the given context, but gives rise to a false implicature, while the negative case is plainly false.[6] The status of (27a) versus (27b) is therefore an important data point to test the predictions of the three different approaches.

[6] As we have noted in Section 1.2.2, the implicature approach does make possible a marked reading of the negative sentence along the lines of example (18); in such cases, the plural morpheme is typically focused, and the reading is captured by positing local computation of the implicature under negation. While the reading is marked to begin with (as direct implicatures typically do not occur under negation), we nevertheless ensured that the negative test sentences in our experiments were produced as neutrally as possible, without added stress on the plural morpheme.

1.3.2 The acquisition of the plural

The three approaches also differ in the predictions they make for the acquisition of the plural and how children might, at a certain stage of development, differ from adults in this area.

On the implicature approach, the expectation is that children should behave roughly as they do with implicatures more generally. If, as has been reported in much previous developmental literature, 4–6-year-old children differ from adults in their computation of scalar implicatures, then everything being equal, we should expect a similar difference between the two groups when it comes to multiplicity inferences as well. We will discuss the predictions about the relation between scalar implicatures and multiplicity inferences in the next subsection.

On the ambiguity approach, the predictions for children's acquisition of the plural depend on what assumptions are made about the acquisition of the proposed meanings for the plural and the Strongest Meaning Principle. On this approach, children need to have acquired the two proposed meanings of the plural and be able to rely on a Strongest Meaning Principle in order to behave in an adult-like manner. Given these prerequisites, there are three main scenarios in which children might not be adult-like. They might go through a developmental stage where they have only acquired one of the two meanings of the plural (either the weak or the strong). Or, they might go through a stage where they have acquired both meanings for the plural, but are not yet able to deploy the Strongest Meaning Principle in an adult-like way. These three possible scenarios will lead to different predictions for how children will respond to plural sentences, compared to adults. For instance, if children have only acquired the strong meaning of the plural, they should appear adult-like on the plural in upward-entailing linguistic environments but not in downward-entailing environments; if they have only acquired the weak meaning of the plural, they should only appear adult-like in downward-entailing environments. On the other hand, if they have acquired both meanings of the plural but cannot yet make use of the Strongest Meaning Principle in selecting a reading, they might not be guided by the relative strength of the two meanings of the plural in the same way that adults are. They might instead, for example, always charitably go with the interpretation that is made true in the context.

Finally, in the case of the homogeneity approach, adult-like behavior is dependent on children having acquired the homogeneity principle and the pragmatic principle for dealing with undefinedness. If they have acquired both, they should perform like adults; if they are missing either ingredient, they will not perform like adults. Importantly, there isn't an obvious way to distinguish between upward- and downward-entailing contexts in this respect; either children will be adult-like in both upward- and downward-entailing contexts, or they will be non-adult-like in both.

1.3.3 Multiplicity inferences vs. implicatures

The three approaches also make different predictions regarding how multiplicity inferences compare to standard scalar implicatures. Consider first the implicature approach. If multiplicity inferences are merely scalar implicatures, one might expect to observe the same kind of behavioral profile for multiplicity inferences as we observe for standard cases of scalar implicature. The nature of this *uniformity prediction* is somewhat complicated, however, by the observation that adults (as a group) vary quite widely in how much they compute different kinds of scalar implicatures (van Tiel, van Miltenburg, Zevakhina & Geurts (2014) observe that implicature rates vary

quite widely across lexical scales, e.g., "some"/"all", "try"/"succeed", "cool"/"cold", etc.). One way to tackle this complication is to compare inference rates *across populations* that are reported to differ in their performance on implicatures. In particular, children have been reported to compute fewer standard implicatures than adults (barring special experimental manipulations and/or the use of certain scales). If multiplicity inferences are merely scalar implicatures, we should expect to see similar *between group* differences when we compare the two scales. For example, we should expect to see fewer of both kinds of inferences from children compared to adults, even if adults may compute the two inferences at different rates. Let us formulate this uniformity prediction of the implicature approach as in (28).

(28) **Uniformity prediction of the implicature approach:** If multiplicity inferences and scalar implicatures are of the same nature, we expect to observe the same pattern of between-group differences (or between-group similarities) across the two kinds of implicatures.

In this respect, a comparison between 4–5-year-old children and adults is particularly useful, as it has been widely reported in the developmental literature that these two groups differ considerably in how often they compute implicatures. While there has been considerable variation in the reported rates of children's success with scalar implicatures, a consistent finding is that children compute fewer standard scalar implicatures than adults do, unless special experimental manipulations are implemented (Noveck 2001; Gualmini, Crain, Meroni, Chierchia & Guasti 2001; Chierchia, Crain, Guasti & Thornton 2001; Papafragou & Musolino 2003; Barner, Brooks & Bale 2011; Katsos & Bishop 2011; Tieu, Romoli, Zhou & Crain 2016, among many others). For instance, children typically accept sentences such as (29a) in situations where the stronger alternative (29b) is also true. This observation has usually been taken to indicate that children fail to compute the implicature in (29c).

(29) a. Emily fed some of the giraffes.
b. Emily fed all of the giraffes.
c. Emily didn't feed all of the giraffes.

If multiplicity inferences are scalar implicatures, then we expect children to likewise compute fewer multiplicity inferences from plural morphology, compared to adults.[7]

The ambiguity and homogeneity approaches, on the other hand, make no particular predictions with respect to the relationship between multiplicity inferences and implicatures, since they do not relate the two phenomena. A convincing case could therefore be made against the implicature approach, if non-uniformity of the two phenomena were observed across populations. Any other result would be compatible with all three approaches.

In sum, while all three approaches to the plural make the same prediction regarding an effect of monotonicity (namely, more multiplicity inferences in upward- than downward-entailing linguistic environments), there are other areas where their predictions diverge: the status of pos-

[7]We will focus on the 'not all' implicature of "some", as it has been extensively studied in the developmental literature. Crucially, previous acquisition studies have reported that children consistently compute fewer 'not all' implicatures from "some" than adults do; moreover, among the numerous scales tested by van Tiel et al. (2014), adults generated the 'not all' inference more often than most other implicatures, so it provides a good testing ground for children's ability to compute implicatures.

itive versus negative plural sentences in singular contexts, the relationship between multiplicity inferences and standard implicatures, and the relative performance of children and adults on the two kinds of inferences. Before moving on to our experiments, we will briefly describe one previous study that set out to test the implicature theory in children and adults, which will allow us to raise some potential issues that we then address in our own experiments.

2 Previous studies

There has been greater experimental investigation of multiplicity inferences in recent years. In particular, there have been studies investigating the suspension of multiplicity inferences compared to that of standard implicatures (Pearson et al. 2011), the sensitivity of the multiplicity inference to contextual manipulations (Grimm 2013), the different readings associated with plural sentences (Patson, George & Warren 2014; Patson 2016; Sauerland, Andersen & Yatsushiro 2005; Dieuleveut, Chemla & Spector 2019) (see Tieu & Romoli 2018 for an overview of the recent experimental literature on plurality). We will focus here on two previous studies, one involving a comparison of the multiplicity inference in children versus adults, and the other investigating the cancellability of the multiplicity inference. As we will discuss, the studies present somewhat mixed evidence for the implicature approach.

2.1 Sauerland, Andersen & Yatsushiro (2005)

Sauerland, Andersen & Yatsushiro (2005) compared adults' and children's performance on multiplicity inferences, with the aim of testing the hypothesis that the latter is a form of implicature. If multiplicity inferences are derived as implicatures, one might expect adults and children to treat the two kinds of inferences in a parallel fashion. As mentioned, while there is considerable variability in the developmental results concerning children's ability to compute scalar implicatures, a fairly consistent finding has been that children compute fewer classical scalar implicatures than adults. There has since been research indicating children can succeed on certain kinds of implicatures, in certain experimental contexts (e.g., Papafragou & Musolino 2003; Miller et al. 2005; Barner et al. 2011; Katsos & Bishop 2011; Stiller et al. 2015; Tieu et al. 2016; Singh et al. 2016). We will set aside for now a discussion of what particular features facilitate children's computation of implicatures, and for the moment assume the basic reported pattern of children's implicatures: without special experimental manipulations or the use of certain scales, children compute fewer implicatures than adults do. If this is so, and if we further assume that multiplicity inferences are implicatures, we are led to the prediction that children should also exhibit difficulty computing multiplicity inferences. In particular, they should compute fewer multiplicity inferences than adults, in the same kinds of contexts where they have been shown to compute fewer scalar implicatures than adults.

Sauerland et al. (2005) designed an experiment that involved a puppet playing the role of an alien asking questions about life on earth. In particular, the puppet would ask participants polar questions such as (30). The expectation was that a reading involving the multiplicity inference (e.g., *Does a dog have more than one tail?*) would lead to a negative answer, given common knowledge, while a literal interpretation of the question (e.g., *Does a dog have one or more tails?*) would lead to an affirmative response.

(30) Does a dog have tails?

The authors compared (30) to control questions like (31), in which the literal meaning of the plural is incompatible with world knowledge, and with (32), in which both the literal meaning of the plural and the multiplicity inference are compatible with world knowledge.

(31) Does a boy have beaks?

(32) Does a cat have feet?

The authors tested 14 children aged 3-5 years, and found the following pattern. Children were adult-like on the control questions, responding negatively to (31) and affirmatively to (32). However, they produced more affirmative responses to target questions like (30) than adults did, answering 'yes' to such questions 96% of the time. This would indicate that children accessed the literal meaning of the question in (30), i.e. *Does a dog have one or more tails?*

The observed difference between children and adults closely mirrors the differences that have been reported in studies of standard implicatures, with children tending to accept implicature targets more than adults. Sauerland et al.'s (2005) results therefore provide a first piece of evidence in favor of the implicature approach to multiplicity.

As Sauerland et al. (2005) point out, however (see also Pearson et al. 2011; Tieu et al. 2014), their study has some potential limitations. First, multiplicity inferences are not always computed in polar questions, even by adults. Sauerland et al. (2005) discuss examples like (33), drawn from Krifka (1989), in which a response assuming a multiplicity inference from (33a) seems to be infelicitous.

(33) a. Does your office have windows?
b. Yes, (only one though).
c. #No, only one.

Sauerland et al. (2005) attribute the difference between a question like (33a) and one like their experimental (30) to whether the person asking the question already knows the answer to the question, with the relevant distinction being that between *true information-seeking questions* and *exam-type questions*. According to the authors, only *exam-type questions* systematically give rise to multiplicity inferences. As Pearson et al. (2011) point out, however, the experimental setting in Sauerland et al. (2005) was precisely one in which the puppet was asking true information-seeking questions; the puppet was ignorant about life on earth and wanted to acquire information by asking the participants questions about life on earth. It is therefore unclear that this distinction is relevant for explaining how the participants behaved in Sauerland et al.'s (2005) study.

Moreover, Sauerland et al.'s (2005) experimental stimuli involved generic interpretations, which could have been misinterpreted by participants as containing dependent plurals. For example, children might have interpreted (30) along the lines of, *Do dogs have tails?*, in which case *yes*-responses would be entirely reasonable, independently of the children's ability to compute multiplicity inferences.

In sum, while the findings of Sauerland et al.'s (2005) study can be taken to provide some initial suggestive evidence in favor of an implicature approach, the results are not conclusive for the reasons just discussed.

2.2 Pearson, Khan & Snedeker (2011)

Pearson, Khan & Snedeker (2011) tackle the hypothesis that multiplicity inferences are implicatures from a different angle. They report an experiment aiming to investigate whether multiplicity inferences are cancellable in upward-entailing contexts, the way that standard scalar implicatures typically are. The authors used a Covered Box task, modelled on a design by Huang, Spelke & Snedeker (2013), who investigated exact interpretations of numerals in comparison to the *not all* inference of "some". In the Huang et al. study, the authors presented participants with three boxes, only two of which had visible contents; the third box was 'covered' so that participants could not see what was inside. On each trial, participants listened to a description and had to choose the box that matched the description. As an example, participants might see a covered box alongside two visible boxes containing one and three cookies, respectively. They would then be directed by the experimenter to "Give me the box with two cookies." If the exactly-*n* meaning of the numeral was a cancellable implicature, participants should select the box containing three cookies; but if the exact meaning was instead part of the semantics of the numeral, participants should instead opt for the covered box. The main finding was that adults chose the covered box for the numeral targets, while for "some" they were more willing to cancel the implicature by selecting the box where Cookie Monster had *all* of the cookies.

On the hypothesis that the multiplicity inference is an implicature, much like the *not all* implicature of "some", one should expect participants to be able to cancel the inference and select one of the visible boxes. Pearson et al. (2011) presented participants with three cards at a time, one of them face down and two of them visible to the participants. For example, alongside the face down card, participants might see a visible card in which Big Bird had no kite and one in which Big Bird had just one kite. Participants would then heard the instruction: *Point to the card where Big Bird has kites*. If participants could cancel the multiplicity inference (*Point to the card where Big Bird has one or more kites*), they would be able to select the visible picture where Big Bird had only one kite; but if participants accessed a multiplicity inference (i.e. *Point to the card where Big Bird has more than one kite*), they would instead be expected to select the covered picture.

In Pearson et al.'s (2011) first experiment, conducted with 16 native speakers of English, participants displayed a strong preference for the face down card on target trials, in parallel with Huang et al.'s (2013) results for the numeral targets but in contrast to their results for the *some* targets. Such a finding would appear to run against the implicature approach, which would predict that the multiplicity inference and *not all* implicature should be similarly cancellable in the given contexts.[8]

In a follow-up experiment, Pearson et al. (2011) introduced the term "only" into the test sentences, e.g., *Point to the card where Big Bird only has kites*, and observed more selections of the visible picture in which Big Bird had exactly one kite. The authors speculate that this could be due to the fact that on a standard semantics of "only," the multiplicity inference is merely presupposed, and not asserted. Yet this still leaves unexplained the difference between the plural and "some": in the same experimental set-up, the latter implicature but not the former is easily cancellable, even in the absence of "only".

In sum, Pearson et al.'s (2011) results do not argue conclusively in favor of the implicature

[8]Note that Pearson et al. did not set out to test the ambiguity account. The ambiguity approach would predict that the multiplicity inference could be cancelled in certain circumstances, in particular when the context makes the multiplicity reading implausible.

account, which cannot explain why the multiplicity inference should be harder to cancel than the scalar implicature of "some".

In the remainder of this paper, we will present three experiments investigating adults' and children's interpretations of multiplicity inferences in upward- and downward-entailing linguistic environments. As we will see, the results of the three experiments are consistent with the predictions of the implicature approach but pose a challenge for the ambiguity and homogeneity approaches. More specifically, the challenge for these latter approaches is to account for the finding that children are relatively more adult-like in downward-entailing than in upward-entailing contexts, and for the finding that adults judge positive plural sentences differently from negative plural sentences in the same context, a pattern that is straightforwardly predicted by the implicature approach.

3 Experiment 1

We designed a Truth Value Judgment Task (Crain & Thornton 1998) to assess the interpretations that adults and children assign to declarative sentences with singular and plural descriptions, in both upward and downward-entailing environments.[9] Testing the interpretations of declarative sentences allowed us to avoid the potential problem that was associated with the disappearance of multiplicity inferences in polar questions in Sauerland et al. (2005).

3.1 Methods

3.1.1 Participants

We tested 28 English-speaking children ($4;01 - 5;09, M = 4;11$) and 43 adult native speakers of English at Macquarie University in Sydney, Australia. Two additional children and one adult were tested but were excluded from the data analysis, as their accuracy rate on control trials was below 75%.

3.1.2 Procedure

Participants watched a series of short stories told by an experimenter, presented through pictures on a laptop computer. The pictures were presented in a Powerpoint presentation. Participants were introduced to a puppet who they were told would interact with them via webcam. In fact, videos of the puppet's lines were pre-recorded, and were played at various points during the stories to create the ruse that the puppet was participating live via webcam. Following each story, the puppet appeared on the screen and was asked a question about the story. The participant's task was to judge the puppet's answers to the questions. Child participants were given a score sheet with two columns, one headed by a "happy face" and one headed by a "sad face"; the children were asked to place a stamp under the "happy face" if the puppet had answered the question correctly, and to place a stamp under the "sad face" if the puppet's answer was wrong. Once they indicated their responses, children were asked why they thought the puppet's sentence was right/wrong. These justifications were audiorecorded and later transcribed. Adult participants

[9] Experiment 1 is also reported in Tieu, Bill, Romoli & Crain (2014).

Figure 1: Final image accompanying either the plural test sentence *Emily fed pigs* or the singular test sentence *Emily fed a pig*.

had similar score sheets but were asked to place a checkmark in the appropriate column; adults also had a third column in which to write short justifications for their responses. The task generally took about 20 minutes to complete for children and between 10-15 minutes for adults.

3.1.3 Materials

We adopted a 2x2x2 design with three factors: group (adults vs. children), number (singular vs. plural, between subjects), and monotonicity (upward- vs. downward-entailing, within subjects). Participants were randomly assigned to either the singular or plural condition. In total, there were six test stories and eight control stories, which were presented in pseudo-randomized order. Three of the critical test stories were associated with a positive (plural or singular) sentence, and three with a negative (plural or singular) sentence.

Consider first the upward-entailing condition. On these three trials, the main character in the story executed an action on only one object from a set of objects; in the plural condition, participants heard a test sentence with a bare plural, while participants heard a singular noun phrase in the singular condition. See (34) for an example test item and Figure 1 for the accompanying final image.

(34) Story: Emily is visiting the pig farm today. It's lunchtime for the pigs. Emily has an apple, and that's just enough to feed the first pig! Oh no! What about the other pigs? The farmer says, "That's okay, Emily! I'll feed the others later!" So in the end, Emily only fed this pig!
EXPERIMENTER: Hey Ellie, what happened in the story?

a. PUPPET: Emily fed pigs! *Plural condition*
b. PUPPET: Emily fed a pig! *Singular condition*

As seen in (34) and in Figure 1, emphasis was placed on the single pig that was fed. To make it very clear that only that particular pig was fed, the prompt contained the focus particle *only* and a definite determiner (i.e. *Emily only fed this pig!*); moreover, a red arrow was added to the picture and the experimenter pointed to the pig on the screen when referring to it.[10]

[10]This was also meant to highlight the episodic nature of the description, i.e. that there was a single event involving the feeding of one pig in particular, and to avoid a possible activity reading of the plural test sentence, e.g., *Emily was*

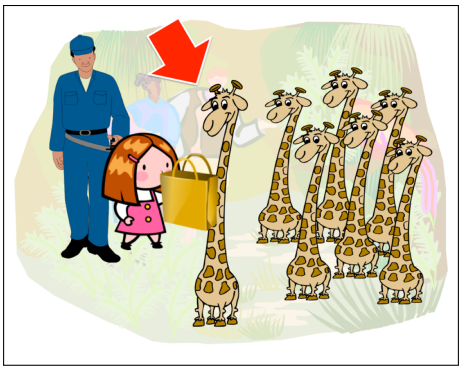


Figure 2: Final image accompanying either the plural test sentence *Emily didn't feed giraffes* or the singular test sentence *Emily didn't feed a giraffe*.

In the plural condition, participants who computed the multiplicity inference in the upward-entailing condition were expected to reject the sentence *Emily fed pigs*, since it was false that Emily fed more than one pig. The singular test sentence was true in the context, however. Thus we expected different responses to singular and plural sentences in the upward-entailing condition.

The stories for the downward-entailing condition were the same in structure, but the test sentences included negation. Here we expected participants to reject the test sentences in both the singular and plural conditions. In other words, the difference between singular and plural should be neutralized under negation. We did, however, expect some proportion of participants to accept the negative test sentences in such contexts if they were able to access a meaning like (8), *Emily didn't feed giraffes, because she fed only one!* See (35) for an example test item and Figure 2 for the accompanying final image.

(35) Story: Emily is visiting the zoo today. It's lunchtime for the animals. Emily has just enough food to feed this very tall giraffe! Oh no! What about the other giraffes? The zookeeper says, "That's okay, Emily! I'll feed the others later!" So in the end, Emily only fed this giraffe!

EXPERIMENTER: Hey Ellie, what happened in the story?

a. PUPPET: Emily didn't feed giraffes! *Plural target*

b. PUPPET: Emily didn't feed a giraffe![11] *Singular target*

Both the plural and the singular test conditions also included two positive and two negative indefinite control items. These control items were designed to elicit the opposite responses to

pig-feeding. Note, however, that even if the activity reading were to arise, it would nevertheless involve the literal meaning of the plural. The test sentence in (34a) would then have a meaning along the lines of *There was a feeding event involving Mary as the agent and* ***one or more giraffes*** *as the patient.* Crucially, whether or not participants access the activity reading (that we have actively tried to suppress), they should accept the sentence only if they access the literal 'one or more' interpretation of the plural. If they access the 'more than one' plurality inference, they should reject the sentence, regardless of whether or not they access an activity reading.

[11]Notice that this sentence potentially has an interpretation on which the indefinite takes scope over negation: *there is a giraffe that Emily didn't feed.* In this case, the sentence would be true in the given context. To avoid this specific/wide-scope reading as much as possible, the relevant pictures, for example that in Figure 2, contained multiple identical animals that were not fed; this way, none of them would stand out from the others. As we will see below, adults and children generally did not interpret the sentence in this way.

those of the critical test trials. In addition to helping to make sure participants stayed on task, the addition of these controls allowed us to ensure that the participants could give both *yes*- and *no*-responses to the test sentences under appropriate circumstances. In the plural condition, the controls corresponded to two positive plural sentences in contexts that satisfied the multiplicity inference, e.g., (36), and two true negative plural sentences, e.g., (37):

(36) Plural positive control (*yes*-target):
Context: Sammy painted two birds.
PUPPET: Sammy painted birds.

(37) Plural negative control (*yes*-target):
Context: Sammy drew one cat.
PUPPET: Sammy didn't draw dogs.

In the singular condition, two control stories made the positive singular test sentences false, e.g.,(38), and two control stories made the negative test sentences true, e.g., (39).

(38) Singular positive control (*no*-target):
Context: Sammy painted one bird.
PUPPET: Sammy painted a tree.

(39) Singular negative control (*yes*-target):
Context: Sammy drew one cat.
PUPPET: Sammy didn't draw a dog.

All participants also received four negation controls, which involved negative sentences containing a definite noun phrase rather than a bare plural or an indefinite *a*-noun phrase, e.g., (40). These trials ensured that child participants had no difficulty interpreting negation independently of the plural or *a*-NP. The negation control trials could be associated with either a *yes*- or a *no*-target; the targets were selected based on participants' responses to the critical test items, enabling us to balance the overall number of *yes*- and *no*-responses.

(40) Negation control:
Context: Sally had a choice between the chocolate and the apple, and in the end ate the apple.
   a. PUPPET: Sally didn't eat the chocolate. (*yes*-target)
   b. PUPPET: Sally didn't eat the apple. (*no*-target)

To summarize, each participant received 6 critical test trials: 3 corresponding to either plural or singular sentences in upward-entailing environments, and 3 corresponding to either plural or singular sentences in downward-entailing environments. In addition, each participant received 8 control trials: 2 positive and 2 negative sentences that elicited the opposite truth values from those of the corresponding critical trials, and 4 negation controls. The 14 trials were presented in pseudo-randomized order.

3.2 Results

3.2.1 Singular targets

Adults and children accepted the positive singular targets 100% of the time, justifying their responses as in (41). They also rejected the negative singular targets (adults 100% of the time and children 98% of the time), justifying their responses as in (42).

(41) Justifications for accepting positive singular target
(Test sentence: *Emily fed a pig.*)
a. "Because she gave the apple to one pig."
b. "Because Emily fed the pig."

(42) Justification for rejecting negative singular target
(Test sentence: *Emily didn't feed a giraffe.*)
"Because she did feed a giraffe."

3.2.2 Plural targets

In the upward-entailing condition, *no*-responses were taken to be indicative of multiplicity inferences, while in the downward-entailing condition, *yes*-responses corresponded to multiplicity inferences. Recall that the latter would correspond to the strong reading in the ambiguity approach and to the locally computed implicature reading in the implicature approach. The *yes*- and *no*-responses to the critical test items were recoded as either 1 (an inference-based response) or 0 (a non-inference response). Figure 3 provides a scatterplot over a barplot, indicating the mean proportions of responses consistent with the computation of the multiplicity inference.

We fitted a mixed effects logistic regression model to the recoded data (using the `lme4` package in R, R Core Team 2016; Bates et al. 2015), with Polarity (Positive vs. Negative), Group (Adult vs. Child), and their interaction as fixed effects, and random by-participant slopes for Polarity. We then used χ^2 statistics with one degree of freedom to compare models with and without the given fixed effects, which revealed significant effects of Polarity ($\chi^2(1) = 15.4, p < .001$) and Group ($\chi^2(1) = 10.6, p < .01$), and no interaction between Polarity and Group ($\chi^2(1) = 1.64, p = .20$).

The results indicate that both groups computed more multiplicity inferences in the upward-entailing condition than in the downward-entailing condition, and that adults computed more inferences than children in both conditions. That participants computed multiplicity inferences at all in the downward-entailing condition suggests that participants sometimes accessed the interpretation in (43), repeated from above. On the ambiguity approach, such a reading would arise from participants ignoring the Strongest Meaning hypothesis, while on the implicature approach, this would correspond to local computation of the multiplicity inference in the scope of negation.

(43) Emily didn't feed giraffes, she fed only one!

The follow-up justifications produced by participants were used to confirm that they were indeed accessing one of the possible target interpretations. For example, some justifications from children who did not compute the multiplicity inference in the positive condition (i.e. children who responded 'yes') made reference to a singularity:

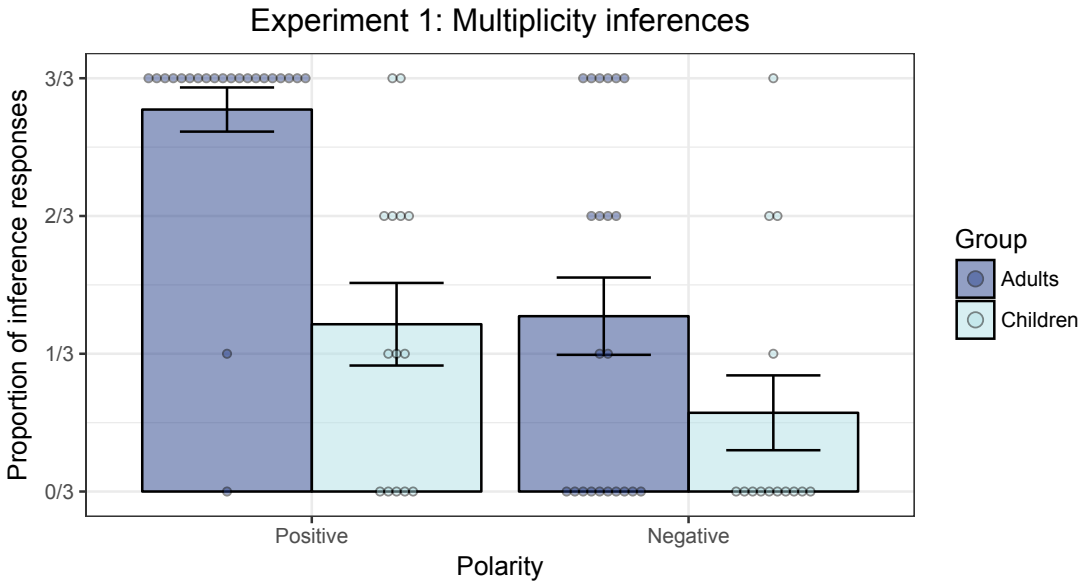


Figure 3: Mean proportion of inference-based responses to the positive and negative plural targets. Dots represent individual participants' rates of inference-based responses.

(44) Sample justifications for *yes*-responses in upward-entailing condition
(Test sentence: *Emily fed pigs.*)
a. "Because she feed a pig."
b. "Because she said the pig has been feeded, and that happened."

Justifications from children who computed the multiplicity inference in the positive condition (i.e. children who responded 'no') referred to the fact that Emily had not fed a plurality of pigs:

(45) Sample justifications for *no*-responses in upward-entailing condition
(Test sentence: *Emily fed pigs.*)
a. "Because she didn't feed all of them."
b. "Because she didn't feed pigs, she only fed a pig."
c. "Because she was only going to feed that big fat pig."

The following is a justification from a child who responded 'no' in the negative condition:

(46) Sample justification for *no*-responses in downward-entailing condition
(Test sentence: *Emily didn't feed giraffes.*)
"Because she said Emily didn't feed the giraffes, and she did."

Finally, justifications from children who appear to have computed the multiplicity inference locally in the scope of negation made reference to a singularity:

(47) Sample justifications for *yes*-responses in downward-entailing condition
(Test sentence: *Emily didn't color diamonds.*)
a. "Because she only did one diamond."

b. "Because she colored that one, but not the other ones."

3.3 Summary

Experiment 1 overcame some possible limitations of the previous study by Sauerland et al. (2005): first, we examined upward-entailing contexts, which allowed us to find a clear difference between children and adults, and second, there is no clear way in which our stimuli could have been misinterpreted as involving dependent plurals. The experiment revealed that the child participants computed fewer multiplicity inferences than adults did. Eight out of the 14 children in the plural condition accepted the test sentences on at least two of the three upward-entailing trials, apparently failing to access the 'more than one' meaning of the plural. This finding is consistent with previous findings from the developmental literature on children's computation of scalar implicatures, and is consistent with the scalar implicature approach to multiplicity inferences. If multiplicity inferences and scalar implicatures are one and the same, then it is not surprising that children exhibit the same developmental pattern across the two phenomena. On the other hand, as we will discuss further in Section 6, the finding that children distinguished the two polarities at all requires further explanation under the ambiguity and homogeneity approaches.

4 Experiment 2

While the finding that children computed fewer multiplicity inferences than adults in Experiment 1 is consistent with a general implicature approach, the reported results in the existing literature reveal variation across both scales and studies. We reasoned that a within-subject design would enable us to better compare children's performance on multiplicity inferences and classical scalar implicatures, and conducted a second experiment to further target the relationship between the two.

4.1 Methods

4.1.1 Participants

We tested 17 English-speaking children ($4;01 - 5;05, M = 4;07$) and 27 adult native speakers of English, at Macquarie University in Sydney, Australia. All participants correctly answered at least six of 8 controls and were therefore included in the analysis.

4.1.2 Procedure

The procedure was the same as in Experiment 1. Participants watched a series of short stories presented by an experimenter, told through sequences of cartoon images on a laptop computer. Participants were asked to judge a puppet's descriptions of the short stories. Child participants recorded their answers with stamps on a score sheet and provided verbal justifications, while adults filled out similar score sheets and provided written justifications. As in the first experiment, the task took about 20 minutes to complete for children, and between 10-15 minutes for adults.

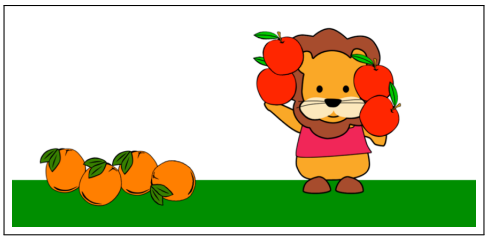


Figure 4: Final image accompanying the scalar implicature test sentence *Lion carried some of the apples.*

4.1.3 Materials

We manipulated three factors: group (children vs. adults), inference type (multiplicity inference vs. standard implicature, within subjects), and within the plural condition, polarity (positive vs. negative, within subjects). The multiplicity inference and scalar implicature conditions were presented in blocks, the order of which was counterbalanced across the participants.

Consider first the scalar implicature condition. On these four trials, the main character in the story executed an action on one complete set of objects; participants then heard a description of the event as involving *some* rather than *all* of the objects. See (48) for an example test item and see Figure 4 for the accompanying final image.

(48) Story: Lion loves to help his mom with the groceries. Look at these apples and oranges! Lion wants to carry the fruit, but they're very heavy! Lion carries these four apples over here. Then his arms are full, so he leaves the oranges on the ground. So remember, Lion only carried these four apples here! Now let's see if Ellie's paying attention.
EXPERIMENTER: Okay, Ellie, so Lion didn't carry any oranges. What about the apples?
PUPPET: Lion carried some of the apples! *Scalar implicature target*

Consider next the multiplicity inference condition. Crucially, we aimed to make the scalar implicature and plural conditions as parallel as possible, so as to be able to directly compare the children's performance in the two conditions. Just as in the scalar implicature condition, the main character in the plural stories was faced with two sets of objects. Instead of carrying out an action with one of the full sets, however, the character executed an action on only one object from the first set of objects. See (49) for an example test item and see Figure 5 for the accompanying final image. Participants who computed the multiplicity inference were expected to reject the critical positive plural target.

(49) Story: Zebra is visiting his favourite garden today. Look at these oranges and bananas! Zebra wants to pick the fruit, but he only has a very small basket. Zebra picks this banana over here. Now he has no more room in the basket, so he leaves the rest of the fruit in the tree. So remember, Zebra only picked this banana here![12] Now let's see if Ellie's paying attention.
EXPERIMENTER: Okay, Ellie, so Zebra didn't pick any oranges. What about bananas?

[12] As in Experiment 1, we tried to highlight the episodic nature of the event by making it very explicit that only one object was acted upon. The experimenter would gesture to the relevant object on the screen, making the use of the deictic pronoun felicitous.

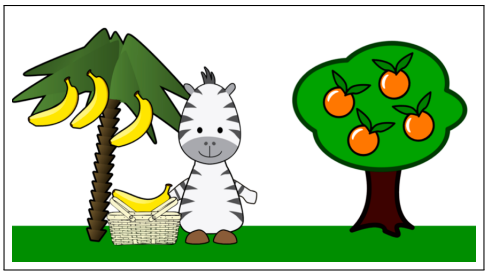


Figure 5: Final image accompanying the plural test sentence *Zebra picked bananas.*

a. PUPPET: Zebra picked bananas! *Positive plural target*
b. PUPPET: Zebra didn't pick bananas! *Negative plural target*

As in Experiment 1, participants also received two positive and two negative control items designed to elicit the opposite responses to those of the critical test trials. The positive plural controls involved contexts that satisfied the multiplicity inference, e.g., (50), and the negative plural controls made reference to the second set of objects that were not acted upon, e.g., (51):

(50) Plural positive control (*yes*-target):
Context: Zebra carried four watermelons and zero pineapples.
Test sentence: Zebra carried watermelons.

(51) Plural negative control (*yes*-target):
Context: Sheep carried four potatoes and zero carrots.
Test sentence: Sheep didn't carry carrots.

Also as in Experiment 1, all participants received four negation control sentences containing a definite noun phrase rather than a bare plural, e.g., (52). As before, the negation control trials could be associated with either a *yes*- or a *no*-target, enabling us to balance the overall number of *yes*- and *no*-responses.

(52) Negation control:
Context: Kangaroo carried four boxes and zero houses.
a. Kangaroo didn't carry the houses. (*yes*-target)
b. Kangaroo didn't carry the boxes. (*no*-target)

To summarize, each participant received 2 training items, 10 test trials, and 8 control trials. The 10 test trials corresponded to 3 plural positive sentences, 3 plural negative sentences, and 4 scalar implicature trials. The 8 control trials corresponded to 2 plural positive sentences, 2 plural negative sentences, and 4 negation trials. The multiplicity inference and scalar implicature conditions were presented in counterbalanced blocks; within the multiplicity inference block, the test and control trials were pseudorandomized to avoid overly long sequences of same-target items.

4.2 Results

Figure 6 provides a scatterplot over a barplot indicating the mean proportions of inference-based responses in the target plural conditions. As in Experiment 1, *no*-responses were taken to be

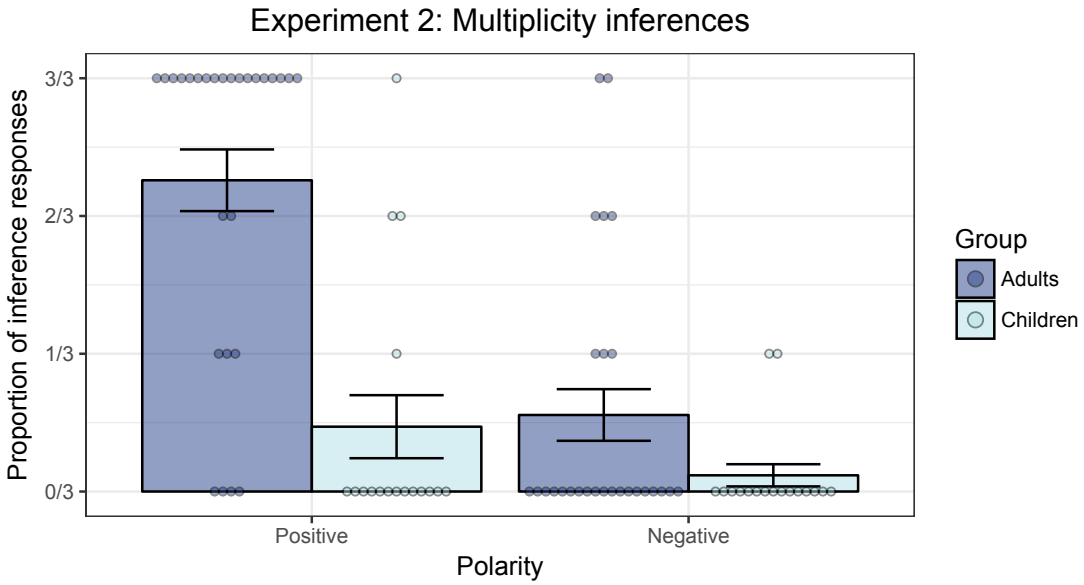


Figure 6: Mean proportion of inference-based responses to the positive and negative plural targets. Dots represent individual participants.

indicative of multiplicity inferences in the upward-entailing condition, whereas *yes*-responses corresponded to multiplicity inferences in the downward-entailing condition (the latter corresponding to the weak reading of the sentence on the ambiguity approach and to locally computed implicatures on the implicature approach). The *yes*- and *no*-responses were recoded as either 1 (an inference-based response) or 0 (a non-inference response).

We fitted a mixed effects logistic regression model on the recoded responses to the plural targets, with Group, Polarity, and their interaction as fixed effects and random by-participant slopes for Polarity; we then used χ^2 statistics with one degree of freedom to compare models with and without the given fixed effects. This revealed significant effects of Group ($\chi^2(1) = 17.3, p < .001$) and Polarity ($\chi^2(1) = 6.74, p < .01$), and a significant interaction between Group and Polarity ($\chi^2(1) = 7.20, p < .01$). In short, participants computed more inferences in the positive condition than in the negative condition, adults computed more multiplicity inferences than children did, and adults and children differed more in the positive condition than in the negative condition.

The results from the plural conditions replicate some of the findings from Experiment 1, namely that more multiplicity inferences are computed in upward-entailing environments than in downward-entailing environments, and that children compute fewer multiplicity inferences than adults do. Also as in Experiment 1, adults computed more multiplicity inferences in the negative condition than children did, appearing to either ignore the Strongest Meaning hypothesis or to compute the multiplicity inference locally in the scope of negation.

Consider next the comparison between multiplicity inferences and scalar implicatures. Figure 7 provides a scatterplot over a barplot indicating the mean proportions of responses (to the plural and scalar implicature targets) that were consistent with the computation of the target inferences. For both the multiplicity inference and the scalar implicature, *no*-responses corresponded to inference-based responses and *yes*-responses corresponded to non-implicature responses. We

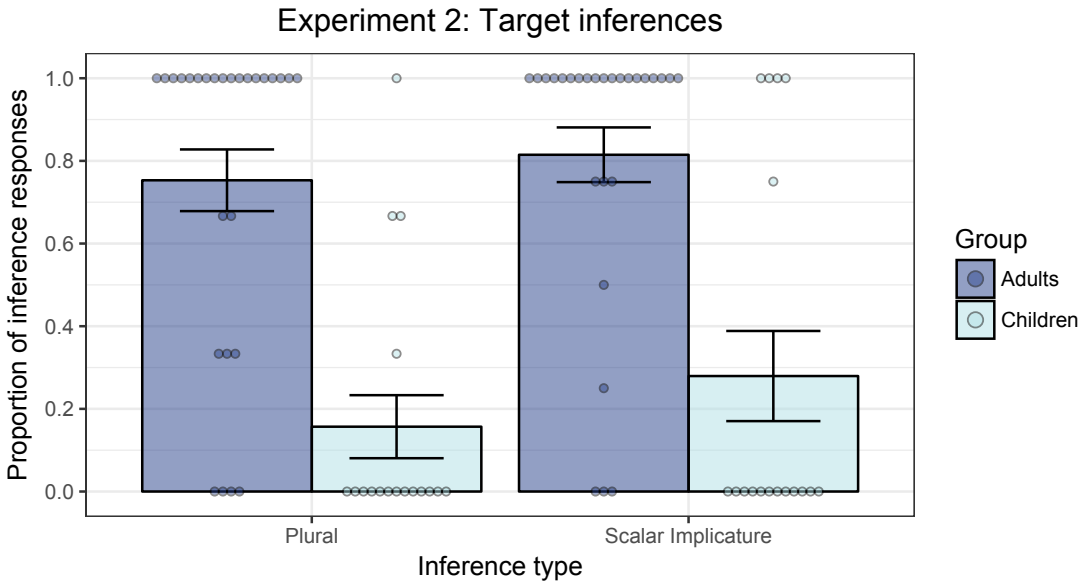


Figure 7: Mean proportion of inference-based responses to the plural and scalar implicature targets. Dots represent individual participants' rates of inference-based responses.

fitted a mixed effects logistic regression model to the recoded data, with Inference Type (Plural vs. Scalar implicature) and Group (Adult vs. Child) as fixed effects, and random by-participant intercepts. We then used χ^2 statistics with one degree of freedom to compare models with and without the fixed effects, which revealed significant effects of Inference Type ($\chi^2(1) = 8.09, p < .01$) and Group ($\chi^2(1) = 28.0, p < .001$) and no interaction between Inference Type and Group ($\chi^2(1) = 1.94, p = .16$). On the whole, then, participants computed fewer multiplicity inferences than scalar implicatures,[13] and children computed fewer of both kinds of inferences than adults did.

In order to evaluate individual performance on multiplicity inferences and scalar implicatures, we ran a correlation analysis in R (`cor.test`), which revealed that children's acceptance of the plural targets was significantly correlated with their acceptance of the scalar implicature targets ($r(15) = .63, p < .01$); no such correlation was found for the adults.

In summary, children computed fewer multiplicity inferences and scalar implicatures than adults did, and their responses in the two conditions were significantly correlated. On the whole, these data are consistent with an implicature approach to plurality inferences, whereby multiplicity inferences are computed using the same mechanism as standard scalar implicatures. On the other hand, the finding that children distinguished the two polarities at all, and in particular differed more from adults in the positive condition than in the negative condition, poses a challenge for the ambiguity and homogeneity approaches. Before discussing this in more detail, we will next turn to a third experiment investigating adults' treatment of the positive and negative sentences in a ternary judgment task, providing another perspective on the differing predictions of the three theories of multiplicity inferences.

[13]The difference between the two inference types does not seem very marked, but could be related to van Tiel et al.'s (2014) finding of variation in the strength of inferences across different scales.

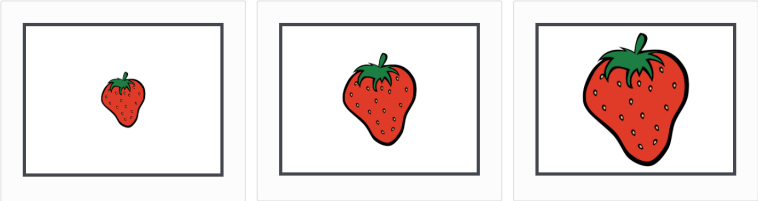


Figure 8: Response buttons for the ternary judgment task.

5 Experiment 3

Experiment 3 used a ternary judgment task to test the different predictions of the three approaches regarding the status of positive versus negative plural sentences in singular contexts. As we have discussed, the ambiguity and homogeneity approaches do not predict a difference in the status of such positive and negative sentences; they are either both undefined or both false. The implicature approach, on the other hand, is compatible with a difference between the two; it in fact predicts a different status for the positive and negative sentences. This is because the positive sentence is literally true but carries a false implicature, while the negative sentence is plainly false. The goal of Experiment 3 was to tap into this potential difference in status in order to tease apart the different approaches to the multiplicity inference. To do so, we compared multiplicity inferences to the exclusivity ('not both') implicature of disjunction.

5.1 Methods

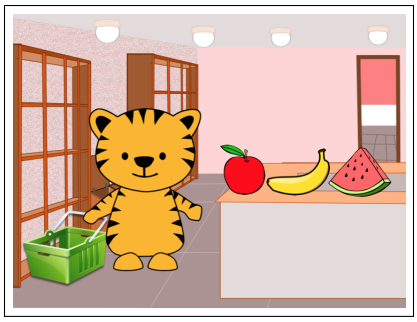
5.1.1 Participants

We tested 24 adult native speakers of English, recruited through Amazon Mechanical Turk. All participants reported English as a native language.

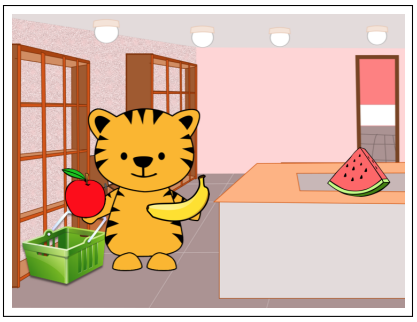
5.1.2 Procedure

Participants were directed to an online version of the ternary judgment task originally used to test scalar implicatures by Katsos & Bishop (2011).[14] The task was implemented and hosted on the Qualtrics platform. Participants saw pictures of cartoon characters deciding what to buy at the store, and heard a puppet's guesses of what would or would not happen. Participants then saw the outcome and had to decide whether to reward the puppet with a small strawberry, a medium-sized strawberry, or a large strawberry. As reported in Katsos & Bishop (2011), participants generally offer the smallest strawberry for false sentences, the intermediate strawberry for literally true sentences whose implicatures are false, and the largest strawberry for true and felicitous sentences. Participants indicated their responses by clicking on the desired strawberry button (see Figure 8).

[14]See also Križ & Chemla (2015), who used a ternary response task with adult participants to investigate homogeneity violations; these authors presented participants with the response options: *completely false*, *completely true*, and *neither*.



(a) Context picture.



(b) Outcome picture.

Figure 9: Context and outcome pictures for the disjunction test sentences *Tiger will buy the apple or the banana* and *Tiger will not buy the apple or the banana.*

5.1.3 Materials

We manipulated two within-subject factors: inference type (multiplicity inference vs. exclusivity implicature) and polarity (positive vs. negative).

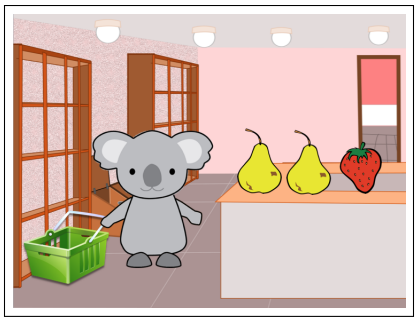
Consider first the disjunction trials. Participants received four positive and four negative target trials in which both disjuncts turned out to be true. On the four positive target trials, for example, the main character in the story ended up buying both of the objects mentioned in the puppet's guess, leaving behind the third object. See (53) for an example test item and Figure 9 for the accompanying images. If participants computed the exclusive *not-both* inference, they were expected to give the intermediate reward of two strawberries.

(53) Story: Now it's Tiger's turn to shop! There's an apple, a banana, and a slice of watermelon! Listen to Baba's guess!
PUPPET: Hmm... Tiger will buy the apple or the banana. *Positive disjunction target*

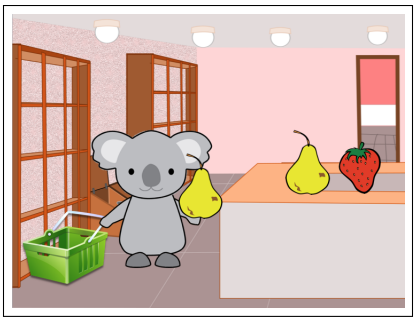
On the four negative target trials, again both disjuncts turned out to be true; here, the puppet's negative guess (e.g., *Tiger will not buy the apple or the banana*) would be expected to elicit a minimal reward if it were interpreted as plainly false.

In addition to the eight target disjunction trials, participants also received 12 control trials: 4 positive and 4 negative disjunction controls in which only one of the disjunctions turned out to be true, and 2 positive and 2 negative disjunction controls in which neither disjunct turned out to be true (see the Appendix for the full list of experimental items).

Consider next the plural condition. Here too, participants received four positive plural targets and four negative plural targets in which the main character bought only a single one of the (plural) objects mentioned in the puppet's guess. See (54) for an example of a positive plural target and Figure 10 for the accompanying images. Here the predictions of the different approaches diverge. Under the implicature approach, participants who computed the multiplicity inference were expected to give the intermediate reward for the positive plural targets. On the four negative plural targets, the puppet's negative statement (e.g., *Koala will not buy pears*) was expected to elicit a minimal reward if the sentence was interpreted as plainly false. Under the homogeneity and ambiguity approaches, on the other hand, participants were expected to give the same response (either minimal or intermediate) to both the positive and negative test sentences.



(a) Context picture.



(b) Outcome picture.

Figure 10: Context and outcome pictures for the plural test sentences *Koala will buy pears* and *Koala will not buy pears*.

(54) Story: Now it's Koala's turn to shop! There's a pear, another pear, and a strawberry! Listen to Baba's guess!
PUPPET: Hmm... Koala will buy pears. *Positive plural target*

In addition to the four positive and four negative plural targets, participants also received two positive and two negative plural controls in which the character did buy the plurality of objects (e.g., both pears), and two positive and two negative plural controls in which the character did not buy the mentioned objects (e.g., Koala did not buy either pear).

To summarize, each participant received 3 training items, followed by a total of 36 test trials: 16 targets (8 disjunction, 8 plural) and 20 controls (12 disjunction, 8 plural). The order of presentation of the 36 trials was randomized across participants.

5.2 Results

Twenty-three of the 24 participants scored at least 75% (9/12) accuracy on the control trials and were included in the analysis. Figure 11 presents the results for the plural and disjunction targets.

We fitted an ordinal regression model on the responses with Inference Type, Monotonicity, and their interaction as fixed effects; we also included by-participant random slopes for Monotonicity and Inference Type and their interaction. Comparisons of the full model versus models without the factors of interest revealed a significant effect of Inference Type ($\chi^2(1) = 15, p < .001$), Monotonicity ($\chi^2(1) = 28, p < .001$), and a significant interaction between Inference Type and Monotonicity ($\chi^2(1) = 14, p < .001$). Overall, adults gave lower rewards for the disjunction targets, lower rewards for the negative sentences, and displayed a greater difference between positive and negative targets in the disjunction condition compared to the plural condition.

On the whole, the results seem more in line with the implicature approach than with the ambiguity and homogeneity approaches. Neither of the latter approaches predicts a difference in the status of the positive and negative sentences. On the implicature approach, on the other hand, the positive sentences are literally true sentences with false implicatures, while the negative sentences are plainly false. The finding that adults gave more minimal rewards for the negative targets, while favoring the intermediate reward for the positive targets, is rather consistent with the predictions of the implicature approach.

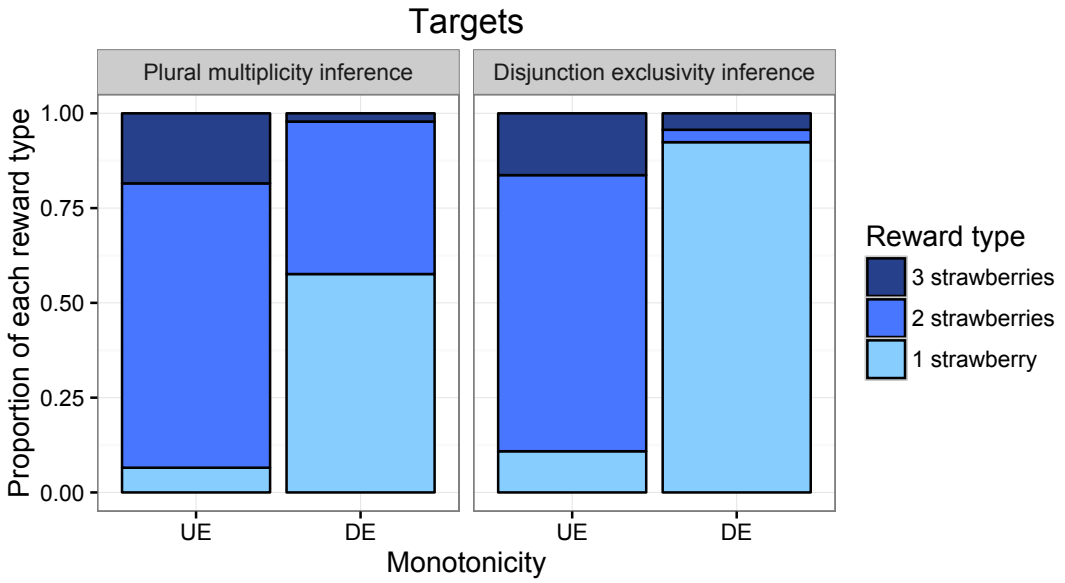


Figure 11: Proportion of inference-based responses to the plural and disjunction targets.

As an anonymous reviewer points out, the proportion of intermediate rewards observed in the negative plural condition is striking, especially as participants gave very few such intermediate responses in the negative disjunction condition. The significant interaction between Inference Type and Monotonicity indeed suggests that the plural is not interpreted exactly along the same lines as disjunction. On this point, the implicature account seems better positioned to accommodate the results than the alternative accounts. As the reviewer points out, (under the implicature account) perhaps participants were unsure whether to compute an embedded implicature, and therefore opted for the intermediate reward; perhaps embedded implicatures are easier to access with the plural than with the disjunction, hence the greater proportion of intermediate rewards for the negative plural than for the negative disjunction. On the alternative accounts, on the other hand, it's unclear how to explain the observed difference between the two inference types.

6 General discussion

The findings from our experiments shed light on the predictions discussed in Section 1.3. In the following, we discuss how our results relate to the different theoretical approaches.

All three experiments revealed a significant effect of monotonicity, with (both adult and child) participants computing more multiplicity inferences in response to positive plural sentences like (55a) than in response to negative plural sentences like (55b). This effect of monotonicity is expected on all three approaches to the plural, and confirms the intuitions that have been reported in the literature.

(55) a. Koala bought pears.
b. Koala didn't buy pears.

Let us now discuss the results against the predictions that we identified earlier, in particular those

where the theoretical approaches diverge.

6.1 On the status of positive vs. negative plural sentences

Recall that in the singular contexts we tested, the homogeneity and ambiguity approaches predict the same status for the positive and negative plural test sentences. In a context in which Koala bought only one pear, (55a) and (55b) are predicted to be either both undefined or both false. That is, no difference in judgment is predicted between the two. Yet in Experiment 3, we observed a clear difference between the two polarities: participants tended to give the intermediate reward for (55a) when Koala only bought one pear, whereas they tended to give the minimal reward for the corresponding negative sentence (55b) in the same context. This difference between the two polarities was also observed for disjunctive test sentences such as (56a) and (56b), in a context in which Tiger bought both the apple and the banana.

(56) a. Tiger will buy the apple or the banana.
b. Tiger will not buy the apple or the banana.

In sum, we find a clear difference in status between positive and negative sentences. This finding is challenging for the homogeneity and ambiguity approaches, which make symmetric predictions across the two polarities, but is in line with the predictions of the implicature approach.

6.2 Children vs. adults

With respect to the comparison between children and adults, the main findings of the first two experiments can be summarized as follows. First, children computed fewer multiplicity inferences overall than adults did. Second, like adults, children were generally sensitive to polarity/monotonicity; however, Experiment 2 revealed that they diverged *more* from adults in the positive condition than in the negative condition; that is, they gave far fewer inference-based responses than adults in the positive condition, but like the adults, they tended to give non-inference-based responses in the negative condition.[15] The finding of fewer inferences overall from children than adults is in line with children's behavior on standard implicatures, and is consistent with an implicature approach to multiplicity inferences. On the other hand, the findings are challenging for the ambiguity and homogeneity approaches, as follows.

While the ambiguity approach is generally compatible with children behaving in a similar way on multiplicity inferences and scalar implicatures, it cannot readily account for the pattern observed across the positive and negative sentences. To illustrate, recall that on the ambiguity approach, the sentence in (57a) is ambiguous between the readings in (57b) and (57c), while the sentence in (58a) is ambiguous between the readings in (58b) and (58c). The Strongest Meaning hypothesis is expected to favor the strong reading in (57b) for (57a) and the strong reading in (58c) for (58a).

[15] A different way of looking at this is that children actually showed less of a distinction between the two polarities than adults did. Figures 3 and 6 reveal a small handful of individual children who tended to give adult-like inference-based responses in the positive condition — it is possible that whatever effect of polarity was observed for children was driven by these individual children. We should caution, however, against placing too much weight on a small handful of individual children, especially as our study was designed to look at group-level effects.

(57) a. Emily fed giraffes.
b. Emily fed more than one giraffe. STRONG
c. Emily fed one or more giraffes. WEAK

(58) a. Emily didn't feed giraffes.
b. Emily didn't feed more than one giraffe. WEAK
c. Emily didn't feed one or more giraffes. STRONG

The challenge that our findings pose for the ambiguity approach is the following. Unlike adults, children seem to predominantly interpret positive sentences like (57a) on the weak reading; that is, they tended to accept the underinformative positive test sentences, whereas adults tended to reject them. Yet for negative sentences like (58a), children appeared *more* adult-like in preferring the strong interpretation (58c); they, like adults, tended to reject the negative test sentences. It is unclear how the ambiguity approach might capture this observed behavioral pattern across the polarities. In greater detail, consider that the ambiguity approach is based on there being two meanings of the plural (and consequently two readings of plural sentences), as well as a principle for resolving the ambiguity. In order to account for an observed difference between children and adults, one might hypothesize that children at 4–6 years of age either fail to access one of the two readings, or fail to apply the Strongest Meaning hypothesis in an adult-like fashion. If we consider these options systematically, however, we will see that neither can entirely account for our results.

For instance, the ambiguity approach might account for the main findings reported above by positing that children have indeed acquired both readings of plural sentences, but engage with the Strongest Meaning principle differently from adults. More precisely, they obey the principle only in downward-entailing contexts. Such a scenario appears implausible without further elaboration and/or assumptions. While it is entirely possible that children at a certain age might differ from adults in their use or understanding of a pragmatic principle like the Strongest Meaning hypothesis, there is no reason why they should be able to apply such a principle only in certain linguistic contexts. In particular, we can see no reason why their application of the principle would systematically vary with the polarity of the context.

Alternatively, the ambiguity approach might hypothesize that children at a certain age have only acquired one of the two *meanings* of the plural morpheme. The scenario in which they have only acquired the strong *exclusive* reading of the plural morpheme would predict the opposite pattern of what we observed in our experiments. This leaves us with the scenario in which children in this age range have only mastered the weak *inclusive* meaning of plural morphology. This would indeed account for the pattern of behavior that we observe, namely that children end up with a globally weak reading of the positive sentence (57c) but a globally strong reading of the negative sentence (58c) (since the weak interpretation of the plural morpheme yields the strongest interpretation under negation). One challenge inherent to this explanation is a learnability one: having first acquired the weak meaning of the plural morpheme, how do children then acquire the strong meaning? In the absence of appropriate negative evidence, it is not clear what would trigger a shift from the weak to the strong meaning. This kind of subset problem is a much discussed topic in the acquisition literature (see, for example, Berwick 1985; Crain, Ni & Conway 1994; Gualmini & Schwarz 2009), and on the face of it would appear to pose a challenge

for this choice point of the ambiguity theory.[16]

Finally, the results are also challenging for the homogeneity approach. In particular, it is unclear how to account for the finding that children were *more* adult-like in response to the negative sentences than to the positive sentences. Under this approach, we would expect children to either have acquired the homogeneity principle and the pragmatic principle for dealing with undefinedness, and therefore to be adult-like in both the positive and negative conditions, or we would expect them not to have acquired either one and therefore to be non-adult-like in both conditions. It is not clear how to account for the observed difference across the two polarities. One could, for instance, posit that children differ from adults in how they deal with a sentence that is undefined in the context. The problem is that given that the homogeneity approach predicts undefinedness in both the negative and positive conditions, we would need to assume that children deal with undefinedness differently for undefined positive sentences and undefined negative sentences. It is not obvious how this would work.[17]

In sum, children's behavior in the first two experiments is more in line with the expectations of the implicature account than of the ambiguity or homogeneity approaches. The key evidence comes from the observation that children were *more adult-like* in their responses to negative plural sentences than to positive plural sentences. This asymmetry between the two polarities is challenging for the ambiguity approach (at least under certain assumptions about acquisition and learnability) as well as for the homogeneity approach.[18]

6.3 Multiplicity inferences vs. scalar implicatures

We set out to empirically investigate the divergent predictions of the three main approaches to multiplicity inferences. To do so, we looked at (i) the status of positive and negative plural

[16]Whether this scenario on the ambiguity approach poses a genuine learnability problem is worth further investigation. As an anonymous reviewer points out, one could posit within a Bayesian framework that a child starts out with a high prior for the hypothesis that the plural has a weak (inclusive) meaning and a low prior for the hypothesis that the plural has a strong meaning; each time they encounter evidence for the strong meaning of the plural, they would use Bayes' rule to increase the posterior probability of the hypothesis that the plural has a strong meaning. For the moment, the potential subset problem won't constitute a key piece of evidence in adjudicating between theories of the plural; but we raise it here as an issue that is worth further careful investigation, particularly if one wants to maintain the ambiguity approach.

[17]A similar pattern can be found in a study by Tieu, Križ & Chemla (2019), who investigated children's interpretation of plural definite descriptions. They report that a subset of their child participants, unlike adults, tended to accept sentences like (ia) in non-homogeneous contexts (e.g., in which only some of the trucks were blue), but *like adults*, tended to reject (ib) in the same context.

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

We refer the reader to Tieu et al. (2019) for discussion of the theoretical consequences for theories of definite plurals.

[18]As an anonymous reviewer points out, this is not to say that there isn't any acquisition challenge associated with the plural under the implicature approach. When it comes to implicatures, the child learner would have to acquire (the ingredients for) a mechanism that allows them to generate inferences on top of literal meanings. On a grammatical view of scalar implicatures, for instance, the child learner would have to acquire the grammatical mechanism of exhaustification as well as be able to access the required alternatives for each relevant inference. In the context of the implicature account of the plural, our experimental findings would suggest that the children we tested hadn't fully acquired some aspect of the generation of the multiplicity inference, for example, the enriched singular alternative to the plural. We return to the acquisition of the multiplicity inference in Section 6.4.

sentences in singular contexts, (ii) children's understanding of plural sentences, and (iii) the relationship between multiplicity inferences and scalar implicatures. As we have discussed, the experimental results, which show an asymmetry both between positive and negative sentences and between children's responses to the two polarities, seem more in line with the implicature approach than with the ambiguity and homogeneity approaches. With respect to (iii), the finding that both children and adults behaved uniformly on multiplicity inferences and scalar implicatures is also consistent with the implicature approach. Experiment 2 moreover revealed that children's performance on multiplicity inferences and scalar implicatures was correlated. That being said, the experimental results with respect to (iii) are technically also compatible with the other two formal approaches. That is, the uniformity between multiplicity inferences and scalar implicatures, while suggestive, could very well be accidental. Had we found that the difference between children and adults was different in nature across the inference types, this would have constituted convincing evidence *against* the implicature approach. As it stands, the results in relation to (iii) are in line with the implicature approach, but alone would be inconclusive in adjudicating among the three theoretical approaches.

6.4 More on the acquisition of the multiplicity inference

One of the hypotheses that the present study investigates is that multiplicity inferences are implicatures. Given some variability in the results that have been reported in the literature on scalar implicatures (both across studies of the same scalar implicature and across different scales), it is natural to ask how the present results align with the broader developmental literature on implicatures. For instance, children have been reported to succeed only on certain kinds of implicatures. Where do multiplicity inferences fit in the overall developmental landscape?

One recent proposal in relation to previous studies on implicatures is that children's reported difficulties relate to a difficulty in accessing lexical alternatives. It has been observed in several studies that children are more likely to compute scalar inferences when the required alternatives are made easily accessible to them, either by explicitly mentioning the alternatives in the discourse context or by introducing the alternatives as substrings of the test sentences themselves (Gualmini, Crain, Meroni, Chierchia & Guasti 2001; Chierchia, Crain, Guasti & Thornton 2001; Reinhart 2006; Barner, Brooks & Bale 2011; Chemla & Bott 2014; Tieu, Romoli, Zhou & Crain 2016; Singh, Wexler, Astle-Rahim, Kamawar & Fox 2016; Tieu, Yatsushiro, Cremers, Romoli, Sauerland & Chemla 2017). The finding that children are more successful at computing inferences in these circumstances has led to the proposal that children have difficulty with inferences that require the lexical replacement of alternatives (for discussion of this point, see Barner et al. 2011, Tieu et al. 2016, and Singh et al. 2016). If the retrieval of alternatives is the relevant factor, then there are at least two crucial requirements for children to be able to compute multiplicity inferences. First, the child must learn the co-scalar status of the singular and the plural. This is critical on an implicature-based analysis like the one proposed in Spector (2007), because it is the negation of the enriched singular that gives rise to the multiplicity inference. To calculate a multiplicity inference, then, the child must know that the singular and the plural are competitors.[19]

[19]In fact, Spector's (2007) account of multiplicity inferences requires the recursive application of the scalar implicature algorithm. It is worth noting in this regard that free choice inferences have also been argued to involve the recursive calculation of implicatures (Kratzer & Shimoyama 2002; Alonso Ovalle 2005; Fox 2007; Klinedinst 2007; Chemla 2009; van Rooij 2010; Franke 2011; Chierchia 2013). We take it that children's success on free choice infer-

In addition, in order to compute the multiplicity inference, the child must be able to activate the singular alternative, in order to carry out the required lexical replacement. Poor performance in computing the multiplicity inference could result from difficulties in either of these key ingredients.[20]

It seems reasonable to suppose that learning the relevant relationship between the singular and the plural may be at least as difficult as learning the co-scalar status of full lexical items such as "some" and "all". For example, it is at least conceivable that children encounter positive evidence regarding the 'not all' implicature, e.g., *I said you could have some of the cookies, not all of them!* The relevant alternatives can be highlighted simply by replacing an instance of one with an instance of the other. In contrast, the kind of evidence that would signal the difference between the (enriched) singular and the plural is less obvious. The difference between the (regular) plural and the singular in English is signaled by the presence of the plural marker *–s*, which stands in contrast to the combination of the absence of the plural marker and the presence of an indefinite determiner. In this sense the co-scalar status of the singular and the plural seems to be qualitatively more distant from relevant evidence than the contrast between "some" and "all"; there isn't an obvious way in which one of the alternatives can be contrasted with the other.[21][22]

Before closing, it may be worth considering the present results in light of previous studies of *younger* children's production and comprehension of plural morphemes, and in particular, studies that have demonstrated the sensitivity of toddlers to plurality. For instance, it has been reported that children begin producing plural morphemes around the age of 22 months (Fenson, Dale, Reznick, Bates, Thal & Pethick 1994; Barner, Thalwitz, Wood & Carey 2007). Similarly, 24-month-olds have been shown to display sensitivity to plural marking in comprehension (Wood,

ences constitutes evidence that children in this age range are indeed capable of recursive implicatures. Note that the fact that children can succeed on recursive implicatures while struggling with non-recursive implicatures can be reconciled on a view on which children have no problem with the mechanism of exhaustification *per se*, but struggle with non-recursive implicatures for an independent reason, namely difficulty in accessing lexical alternatives (see Barner et al. 2011; Tieu et al. 2016; Singh et al. 2016; Pagliarini et al. 2018 for relevant discussion). Thanks to an anonymous reviewer for raising this question.

[20] If it is the second ingredient that is relevant, one may expect that performance on multiplicity inferences could be improved by making the singular/plural contrast salient, using experimental maneuvers that have proven successful in evoking successful performance from children in response to classical scalar implicatures (Gualmini, Crain, Meroni, Chierchia & Guasti 2001; Barner, Brooks & Bale 2011; Tieu, Romoli, Zhou & Crain 2016; Gotzner, Barner & Crain 2015; Skordos & Papafragou 2016, among others). In these previous cases, however, the relevant alternatives were free morphemes; for instance, increasing the salience of the conjunction word "and" could boost the rate at which children might access the exclusive meaning of the disjunction word "or". It's not immediately obvious how one might achieve the same effect in the case of multiplicity inferences. It's not clear, for example, whether activating a lexical item corresponding to a singular meaning (for example, "a dog" as opposed to "dogs") would suffice to trigger the enriched singular feature that is required to compute the multiplicity inference.

[21] In fact, some developmental studies seem to suggest that plural meanings emerge earlier than singular meanings (Davies et al. 2019). Arias-Trejo et al. (2014) and Davies et al. (2017), for example, report that their 20–24-month-old participants displayed better than chance performance on plurals but were at chance on the singular. Such findings suggest that the plural meaning is acquired prior to that of the singular (see also Davies et al. 2019 for further evidence that performance on the plural, but not the singular, improves with age between 2–3 years of age). These studies pose a puzzle for the (entirely reasonable) hypothesis that acquiring the plural requires acquiring the distinction between the singular and plural.

[22] With respect to overt morphological marking of the plural, it may prove insightful to investigate the development of multiplicity inferences in languages that do not morphologically mark the plural on the noun; alternatively, one might investigate the development of multiplicity inferences using irregular nouns in English (thanks to Scott AnderBois for this suggestion).

Kouider & Carey 2009; Davies, Xu Rattanasone & Demuth 2017).[23] At first glance, these data from 2-year-olds appear to be at odds with our data from 4-year-olds. To better understand this divergence in the pattern of behavioural responses by children at different ages, it may be worth considering the nature of the tasks that have been used to investigate children's knowledge of plurality.

For example, Wood et al. (2009) used a manual search paradigm in which 24-month-old children were prompted by linguistic descriptions to search for objects in a box. The authors found that the verbal descriptions contained in the prompts (e.g., "There are some cars in the box" vs. "There is a car in the box") influenced the search patterns of the 24-month-olds, such that they searched longer in the plural condition. This finding suggests that the infants were expecting 'more than one' object in the plural condition.[24] By contrast, in the paradigm we adopted in the present study, 4-year-olds encountered situations in which only a single object was depicted, but they accepted plural descriptions of these situations. This doesn't necessarily mean that the child participants didn't 'expect' there to be more than one object, in the same way that the 2-year-olds did in the Wood et al. study. The 4-year-olds in the present study could have associated the plural description with a 'more than one' meaning, but they may also have been willing to accept these uses of plural morphology in contexts in which they themselves would not have produced the plural.

In fact, this acquisition scenario establishes a further parallel with the scalar implicature literature, in light of an existing proposal that children are more *pragmatically tolerant* than adults are (Katsos & Bishop 2011). Katsos & Bishop (2011) show that children can display adult-like implicature performance when they are given more than just a binary choice between *yes* and *no* (or *right* and *wrong*). Based on the findings from their study, Katsos and Bishop suggest that 4-year-olds are simply more tolerant than adults are in responding to pragmatically underinformative descriptions. In a similar vein, despite the observation that children tend to accept underinformative "or"-statements in standard binary judgment tasks, it has also been demonstrated that children clearly distinguish between statements with "or" and statements with "and" when they are presented side by side. This is the case in the 'felicity judgment task', which involves two puppets; one produces the statement with "or" and the other produces the statement with "and" (Chierchia, Crain, Guasti & Thornton 2001; Gualmini, Crain, Meroni, Chierchia & Guasti 2001). The fact that children consistently judge the puppet that used the statement with "and" to have described the situation better can be taken as evidence that children are sensitive to the relative strength of two alternatives, although they are more accepting of weak alternatives than adults are. More generally, these data also serve to highlight the importance of the nature of the tasks used by researchers to investigate children's acquisition of scalar knowledge.

[23] On the other hand, it has also been reported that children do not always use the plural in all required contexts (for relevant discussion, see Berko 1958; Cazden 1968; Brown 1973; Mervis & Johnson 1991, and Zapf & Smith 2008). Previous explanations for non-adult-like performance have appealed to morpho/phonological rules and constraints (see, for example, Marcus, Pinker, Ullman, Hollander, Rosen & Xu 1992; Marchman, Plunkett & Goodman 1997, and Winitz et al. 1981). There are far fewer studies that address the meaning that children assign to the plural, which is our primary focus here.

[24] More recent work by Davies, Xu Rattanasone & Demuth (2017) suggests that 24-month-olds can indeed demonstrate understanding of English plural morphology, but only when they are presented with the voiceless /s/ allomorph, not the voiced /z/ allomorph. These authors propose that factors such as the phonetic salience of the morpheme are relevant to the process of acquiring plural morphology and its meaning.

7 Conclusion

There are three main theoretical approaches to the 'more than one' meaning of plural morphology. The goal of the present study was to experimentally test the predictions of the ambiguity, homogeneity, and scalar implicature approaches. In particular, the three approaches diverge on three points: (i) the status of positive plural sentences versus negative plural sentences in singular contexts, (ii) the acquisition of the different readings of plural sentences, and (iii) the relationship between multiplicity inferences and scalar implicatures. We conducted three experiments, which revealed a few main findings bearing on these points.

First, participants computed more multiplicity inferences in upward-entailing linguistic environments compared to downward-entailing environments. This effect of polarity is expected on all of the theoretical approaches. Experiment 3, which used a ternary judgment task, moreover revealed that adults assigned different values to positive and negative plural sentences in the same context. As we have discussed, this is challenging for the homogeneity and ambiguity approaches, which predict that the positive and negative sentences should have the same status in the relevant contexts; the results though, are unsurprising on the scalar implicature theory.

With respect to children's acquisition of the readings of plural sentences, we observed that children were rather non-adult-like in accepting underinformative *positive* plural sentences — in line with their behavior on standard scalar implicatures — but they were *relatively* adult-like in rejecting the *negative* plural sentences in singular contexts. This finding of asymmetric behavior from the children across the two polarities is in line with the implicature approach, but presents another challenge for the other two approaches, at least under certain assumptions about acquisition and learnability.

Finally, in relation to (iii), children and adults behaved uniformly on multiplicity inferences and scalar implicatures. This latter finding, while consistent with the implicature approach, can also be made compatible with the other two approaches. When we look at the experimental findings as a whole, we find that the data pertaining to points (i) and (ii) above present more support for the scalar implicature approach to multiplicity inferences, while leaving open some challenges for the ambiguity and homogeneity approaches.

Supplementary material

The raw data and R scripts for analysis for all three experiments can be accessed at `https://osf.io/9sek5`.

Appendix: Test sentences

Experiment 1

Training items

(59) Mary took the yellow bus to school. *True target*

(60) Lucy held the rabbits. *False target*

Singular condition

(61) ***Positive targets — presented in singular contexts***

a. Emily fed a pig.
b. Sally coloured a triangle.
c. Sue picked a banana.

(62) ***Negative targets — presented in singular contexts***

a. Emily didn't feed a giraffe.
b. Sally didn't colour a diamond.
c. Sue didn't pick an apple.

(63) ***Positive controls***

a. Sammy painted a tree. *False target*
b. Mary folded a star. *False target*

(64) ***Negative controls***

a. Sammy didn't draw a dog. *True target*
b. Lucy didn't buy a cookie. *True target*

Plural condition

(65) ***Positive targets — presented in singular contexts***

a. Emily fed pigs.
b. Sally coloured triangles.
c. Sue picked bananas.

(66) ***Negative targets — presented in singular contexts***

a. Emily didn't feed giraffes.
b. Sally didn't colour diamonds.
c. Sue didn't pick apples.

(67) ***Positive controls***

a. Sammy painted birds. *True target*
b. Mary folded stars. *True target*

(68) ***Negative controls***

a. Sammy didn't draw dogs. *True target*
b. Lucy didn't buy cookies. *True target*

Negation controls (appeared in both Singular and Plural conditions)

(69)
a. Sally didn't eat the chocolate. *True target*
or: Sally didn't eat the apple. *False target*
b. Billy didn't go for a swim. *True target*
or: Billy didn't read his book. *False target*
c. Lucy didn't take a nap. *True target*
or: Lucy didn't walk her dog. *False target*
d. Mary didn't buy the ice cream. *True target*
or: Mary didn't buy the cookie. *False target*

Experiment 2

Training items

(70) Pig painted the cup. *True target*

(71) Monkey took the scooter. *False target*

Multiplicity inference condition

(72) ***Positive targets — presented in singular contexts***
a. Tiger fed pigs.
b. Zebra picked bananas.
c. Bunny painted guitars.

(73) ***Negative targets — presented in singular contexts***
a. Frog didn't feed birds.
b. Kangaroo didn't pick pears.
c. Lion didn't paint bowls.

(74) ***Positive controls***
a. Giraffe carried watermelons. *True target*
b. Chicken fed cats. *True target*

(75) ***Negative controls***
a. Sheep didn't carry carrots. *True target*
b. Tiger didn't feed mice. *True target*

(76) ***Negation controls***
a. Zebra didn't paint the bowls. *True target*
or: Zebra didn't paint the vases. *False target*
b. Bunny didn't carry the tables. *True target*
or: Bunny didn't carry the chairs. *False target*
c. Frog didn't paint the hearts. *True target*
or: Frog didn't paint the stars. *False target*
d. Kangaroo didn't carry the houses. *True target*
or: Kangaroo didn't carry the boxes. *False target*

Scalar implicature condition

(77)
a. Lion picked some of the apples.
b. Giraffe painted some of the cars.
c. Chicken fed some of the ladybugs.
d. Sheep carried some of the boats.

Experiment 3

Training items

(78)
a. Pig will buy the plant. *True target*
b. Penguin will not buy the plant. *True target*
c. Hippo will not buy the submarine. *False target*

Multiplicity inference condition

(79) ***Positive targets — presented in singular contexts***
a. Koala will buy pears.
b. Deer will buy lemons.
c. Tiger will buy plums.
d. Zebra will buy balls.

(80) ***Negative targets — presented in singular contexts***
a. Giraffe will not buy hats.
b. Chicken will not buy flowers.
c. Sheep will not buy bowls.
d. Monkey will not buy airplanes.

(81) ***Positive controls***
a. Fox will buy apples. *True target*
b. Raccoon will buy peaches. *True target*
c. Bunny will buy trucks. *False target*
d. Frog will buy boats. *False target*

(82) ***Negative controls***
a. Panda will not buy cupcakes. *True target*
b. Owl will not buy boots. *True target*
c. Kangaroo will not buy cars. *False target*
d. Lion will not buy guitars. *False target*

Disjunction condition

(83) ***Positive targets — presented in 2DT contexts***
a. Tiger will buy the apple or the banana.
b. Zebra will buy the peach or the watermelon.
c. Bunny will buy the pear or the pineapple.
d. Frog will buy the lemon or the strawberry.

(84) ***Positive targets — presented in 1DT contexts***

a. Kangaroo will buy the plum or the kiwi.
b. Lion will buy the ball or the train.
c. Giraffe will buy the truck or the doll.
d. Chicken will buy the boat or the balloon.

(85) ***Negative targets — presented in 2DT contexts***

a. Panda will not buy the hat or the scarf.
b. Owl will not buy the flower or the star.
c. Pig will not buy the bowl or the cup.
d. Penguin will not buy the airplane or the soccer ball.

(86) ***Negative targets — presented in 1DT contexts***

a. Hippo will not buy the cupcake or the ice cream.
b. Duck will not buy the boot or the chair.
c. Bear will not buy the tennis ball or the block.
d. Cat will not buy the car or the dinosaur.

(87) ***Positive controls***

a. Sheep will buy the car or the dinosaur. *False target*
b. Monkey will buy the guitar or the piano. *False target*

(88) ***Negative controls***

a. Dog will not buy the balloon or the truck. *True target*
b. Elephant will not buy the shoe or the hat. *True target*

References

Alonso Ovalle, Luis. 2005. *Disjunction in Alternative Semantics*: University of Massachusetts at Amherst dissertation.

Arias-Trejo, Natalia, Lisa M. Cantrell, Linda B. Smith & Elda A. Alva. 2014. Early comprehension of the spanish plural. *Journal of Child Language* 41(6). 1356–1372.

Barner, D., D. Thalwitz, J. Wood & S. Carey. 2007. On the relation between the acquisition of singular-plural morpho-syntax and the conceptual distinction between one and more than one. *Developmental Science* 10. 365–373.

Barner, David, Neon Brooks & Alan Bale. 2011. Accessing the unsaid: the role of scalar alternatives in children's pragmatic inferences. *Cognition* 118(1). 84–93.

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

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Berwick, Robert C. 1985. *The Acquisition of Syntactic Knowledge*. Cambridge, Massachusetts: MIT Press.

Brown, Roger. 1973. *A First Language: The Early Stages*. London: Allen and Unwin.

Cazden, Courtney B. 1968. The acquisition of noun and verb inflections. *Child Development* 39. 433–448.

Chemla, Emmanuel. 2009. Similarity: Towards a unified account of scalar implicatures, free choice permission, and presupposition projection. Ms., LSCP.

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

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6(4). 339–405.

Chierchia, Gennaro. 2013. *Logic in Grammar: Polarity, Free Choice, and Intervention*. Oxford University Press.

Chierchia, Gennaro, Stephen Crain, Maria Teresa Guasti & Rosalind Thornton. 2001. 'Some' and 'or': A study on the emergence of logical form. In S. Catherine Howell, Sarah A. Fish & Thea Keith-Lucas (eds.), *Boston University Conference on Language Development (BUCLD) 24*, 22–44. Somerville, Massachusetts: Cascadilla Press.

Crain, Stephen, Weijia Ni & Laura Conway. 1994. Learning, parsing, and modularity. In Jr. Clifton, Charles, Lyn Frazier & Keith Rayner (eds.), *Perspectives on Sentence Processing*, 443–467. Hillsday, New Jersey: Lawrence Erlbaum.

Crain, Stephen & Rosalind Thornton. 1998. *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, Massachusetts: MIT Press.

Dalrymple, M., M. Kanazawa, Y. Kim, S. Mchombo & S. Peters. 1998. Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy* 21. 159–210.

Davies, Benjamin, Nan Xu Rattanasone & Katherine Demuth. 2017. Two-year-olds' sensitivity to inflectional plural morphology: Allomorphic effects. *Language Learning and Development* 13(1). 38–53.

Davies, Benjamin, Nan Xu Rattanasone, Tamara Schembri & Katherine Demuth. 2019. Preschoolers' developing comprehension of the plural: The effects of number and allomorphic variation. *Journal of Experimental Child Psychology* 185. 95–108.

Dieuleveut, Anouk, Emmanuel Chemla & Benjamin Spector. 2019. Distinctions between primary and secondary scalar implicatures. *Journal of Memory and Language* 106. 150–171.

Farkas, Donka & Henriette de Swart. 2010. The semantics and pragmatics of plurals. *Semantics & Pragmatics* 3(6). 1–54.

Fenson, L., P.S. Dale, J. Reznick, E. Bates, D.J. Thal & S.J. Pethick. 1994. Variability in early communicative development. *Monograph of the Society for Research in Child Development* 59(5, Serial No. 242).

Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*, 71–120. Basingstoke: Palgrave Macmillan.

Franke, Michael. 2011. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics* 4(1). 1–82. doi:10.3765/sp.4.1.

Gotzner, Nicole, David Barner & Stephen Crain. 2015. 4-5 year olds exhaustify disjunctive statements. Poster presented at Architectures & Mechanisms for Language Processing 2015.

Grice, Paul. 1975. Logic and conversation. In Donald Davidson & Gilbert H. Harman (eds.), *The Logic of Grammar*, 64–75. Encino, CA: Dickenson Publishing Company.

Grimm, Scott. 2013. Plurality is distinct from number-neutrality. In *Proceedings of the North East Linguistic Society 41*, 247–258. Amherst, MA: University of Massachussetts at Amherst.

Gualmini, Andrea, Stephen Crain, Luisa Meroni, Gennaro Chierchia & Maria Teresa Guasti. 2001. At the semantics/pragmatics interface in child language. In Rachel Hastings, Brendan Jackson

& Zsofia Zvolenszky (eds.), *Semantics and Linguistic Theory (SALT) 11*, 231–247. Cornell University, Ithaca, NY: CLC Publications.

Gualmini, Andrea & Bernhard Schwarz. 2009. Solving learnability problems in the acquisition of semantics. *Journal of Semantics* 26. 185–215.

Huang, Yi Ting, Elizabeth Spelke & Jesse Snedeker. 2013. What exactly do numbers mean? *Language Learning and Development* 9. 105–129.

Ivlieva, Natalia. 2013. *Scalar implicatures and the grammar of plurality and disjunction*: Massachusetts Institute of Technology dissertation.

Katsos, Napoleon & Dorothy V.M. Bishop. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition* 120(1). 67–81.

Klinedinst, Nathan. 2007. *Plurality and Possibility*: UCLA dissertation.

Kratzer, Angelika & Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. In Yukio Otsu (ed.), *Proceedings of the Tokyo conference on psycholinguistics*, vol. 3, 1–25. Tokyo: Hituzi Syobo.

Krifka, Manfred. 1989. Nominal reference, temporal constitution and quantification in event semantics. In Renate Bartsch (ed.), *Semantics and contextual expressions*, 75–116.

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

Križ, Manuel. 2017. Bare plurals, multiplicity, and homogeneity. Ms Institut Jean-Nicod, École Normale Supérieure Paris.

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(3). 205–248.

Lasersohn, Peter. 1995. *Plurality, Conjunction and Events*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Löbner, Sebastian. 1987. The conceptual nature of natural language quantification. In I. Rusza & Anna Szabolcsi (eds.), *Proceedings of the '87 Debrecen Symposium on Logic and Language*, Budapest: Akadémiai Kiadó.

Magri, Giorgio. 2014. An account for the homogeneity effects triggered by plural definites and conjunction based on double strengthening. In Salvatore Pistoia Reda (ed.), *Pragmatics, Semantics and the Case of Scalar Implicatures*, 99–145. Palgrave Macmillan.

Marchman, V.A., K. Plunkett & J. Goodman. 1997. Overregularization in English plural and past tense inflectional morphology: A response to Marcus (1995). *Journal of Child Language* 24(3). 767–779.

Marcus, G.F., S. Pinker, M. Ullman, M. Hollander, T.J. Rosen & F. Xu. 1992. *Overregularization in language acquisition*, vol. 57 (Monographs of the Society for Research in Child Development 4, Serial No. 228).

Mayr, Clemens. 2015. Plural definite nps presuppose multiplicity via embedded exhaustification. In *Proceedings of the 25th Semantics and Linguistic Theory Conference*, 204–224.

Mervis, Carolyn B. & Kathy E. Johnson. 1991. Acquisition of the plural morpheme: A case study. *Developmental Psychology* 27. 222–235.

Miller, K., C. Schmitt, H. Chang & A. Munn. 2005. Young children understand some implicatures. In *Proceedings of the 29th Annual Boston University Conference on Language Development*, Cascadilla Press.

Noveck, Ira. 2001. When children are more logical than adults: experimental investigations of scalar implicatures. *Cognition* 78(8). 165–188.

Pagliarini, Elena, Cory Bill, Jacopo Romoli, Lyn Tieu & Stephen Crain. 2018. On children's variable success with scalar inferences: Insights from disjunction in the scope of a universal quantifier. *Cognition* 178. 178–192. Doi:10.1016/j.cognition.2018.04.020.

Papafragou, Anna & Julien Musolino. 2003. Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition* 86(3). 253–282.

Patson, N.D., G.E. George & T. Warren. 2014. The conceptual representation of number. *Quarterly Journal of Experimental Psychology* 67. 1349–1365.

Patson, Nikole D. 2016. Evidence in support of a scalar implicature account of plurality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42(7). 1140–1153.

Pearson, Hazel, Manizeh Khan & Jesse Snedeker. 2011. Even more evidence for the emptiness of plurality: An experimental investigation of plural interpretation as a species of scalar implicature. In David Lutz & Nan Li (eds.), *Semantics and Linguistic Theory (SALT) 20*, Cornell University Ithaca, NY: CLC Publications.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reinhart, Tanya. 2006. *Interface Strategies. Optimal and Costly Derivations*. Cambridge, Mass.: MIT Press.

van Rooij, Robert. 2010. Conjunctive interpretations of disjunctions. *Semantics and Pragmatics* 3(11). 1–28. doi:10.3765/sp.3.11.

Sauerland, Uli. 2003. A new semantics for number. In Robert B. Young & Yuping Zhou (eds.), *Semantics and Linguistic Theory (SALT) 13*, 258–275. Cornell University Ithaca, NY: CLC Publications.

Sauerland, Uli, Jan Andersen & Kazuko Yatsushiro. 2005. The plural is semantically unmarked. In Stephan Kepser & Marga Reis (eds.), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, 413–434. Mouton de Gruyter.

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

Singh, Raj, Ken Wexler, Andrea Astle-Rahim, Deepthi Kamawar & Danny Fox. 2016. Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics* 24(4). 305–352.

Skordos, Dimitrios & Anna Papafragou. 2016. Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition* 153. 6–18.

Spector, Benjamin. 2007. Aspects of the pragmatics of plural morphology: On higher-order implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*, 243–281. Palgrave.

Stiller, A., N. Goodman & M. Frank. 2015. Ad-hoc implicature in preschool children. *Language Learning and Development* 11(2). 176–190.

van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina & Bart Geurts. 2014. Scalar diversity. *Journal of Semantics* 33(1). 137–175.

Tieu, Lyn, Cory Bill, Jacopo Romoli & Stephen Crain. 2014. Plurality inferences are scalar implicatures: Evidence from acquisition. In Todd Snider, Sarah D'Antonio & Mia Wiegand (eds.), *Proceedings of the 24th Semantics and Linguistic Theory Conference*, 122–136. LSA and CLC Publications.

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

Tieu, Lyn & Jacopo Romoli. 2018. Plurality. In Chris Cummins & Napoleon Katsos (eds.), *Handbook of Experimental Semantics and Pragmatics*, 208–227. Oxford University Press.

Tieu, Lyn, Jacopo Romoli, Peng Zhou & Stephen Crain. 2016. Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics* 33(2). 269–298.

Tieu, Lyn, Kazuko Yatsushiro, Alexandre Cremers, Jacopo Romoli, Uli Sauerland & Emmanuel Chemla. 2017. On the role of alternatives in the acquisition of simple and complex disjunctions in French and Japanese. *Journal of Semantics* 34(1). 127–152.

Winitz, H., R. Sanders & J. Kort. 1981. Comprehension and production of the /ez/ plural allomorph. *Journal of Psycholinguistic Research* 10(3). 259–271.

Winter, Yoad. 2001. *Flexibility Principles in Boolean Semantics*. Cambridge, Mass.: MIT Press.

Wood, Justin N., Sid Kouider & Susan Carey. 2009. Acquisition of singular-plural morphology. *Developmental Psychology* 45(1). 202–206.

Zapf, Jennifer A. & Linda B. Smith. 2008. Meaning matters in children's plural productions. *Cognition* 108(2). 466–476.

Zweig, Eytan. 2009. Number neutral bare plurals and the multiplicity agreement. *Linguistics and Philosophy* 32. 353–407.