

The exclusive interpretation of plural nominals in quantificational environments

Pranav Anand, Caroline Andrews,
Donka Farkas, and Matthew Wagers
University of California, Santa Cruz

Abstract In some contexts, plural nominals have inclusive interpretation, allowing atoms in their reference domain; in others, they are exclusive, allowing only sums. Selecting between the two interpretations has been shown to be sensitive to both world-knowledge pressures (Farkas & de Swart 2010) and contextual relevance (Grimm To Appear). The principal semantic factor claimed to be involved is monotonicity direction (Sauerland, Anderssen & Yatsushiro 2005; Spector 2007; Zweig 2009; Farkas & de Swart 2010): upward monotone environments tend to select exclusive readings; downward monotone ones, inclusive readings. In four image verification experiments, we tested this claim and found support for the generalization. The effect of monotonicity direction, however, is small. Moreover we find that varying whether a plural is in the scope of a quantified description has a much larger effect on the prevalence of the exclusive interpretation. This suggests that monotonicity, though involved, is not a decisive factor in plural interpretation.

Keywords: plurality, number, pragmatics, experimental

1 Introduction

It has long been noted that nominals marked by plural morphology in many languages can be interpreted either *exclusively*, requiring reference to two or more entities, and therefore excluding atoms from their domain, or *inclusively*, allowing both atom and sum witnesses. Thus, in (1),

(1) EXCLUSIVE REFERENCE: The baby is playing with *teddy bears*.

the plural nominal is said to have *exclusive* reference because the sentence is normally interpreted as asserting that the baby is playing with *two or more teddy bears*. In (2), on the other hand,

(2) INCLUSIVE REFERENCE: The baby is not playing with *teddy bears*.

the plural nominal is said to have *inclusive* reference because the sentence is normally interpreted as asserting that the baby is not playing with *one or more teddy bears*.

The question that we investigate here is what determines the choice of one reading over the other. Many accounts begin from the observation that this choice is largely determined by the monotonicity direction of the environment in which a plural occurs. Let us refer to this as the Monotonicity Generalization (3). (See Sauerland et al. 2005, Zweig 2009, Farkas & de Swart 2010.)

- (3) MONOTONICITY GENERALIZATION
- a. Plural nominals in upward monotone environments tend to receive exclusive interpretations.
 - b. Plural nominals in downward monotone environments tend to receive inclusive interpretations.

Our main aim here is to experimentally test the two sub-claims of (3) by contrasting the interpretation of plural nominals in the restrictor versus the nuclear scope of *each*. According to (3), plurals occurring in the restrictor of universal quantifiers will tend to receive an inclusive interpretation because that environment is downward monotone. Conversely, the nuclear scope of universals, which are upward monotone, will favor exclusive interpretations.

We tested these predictions in a series of four image verification experiments. These were designed to measure how often nuclear scope and restrictor plurals receive exclusive interpretations. We refer to this quantity as the exclusivization rate of the environment. There are two main findings. First, the exclusivization rate of the nuclear scope is higher than the restrictor, an asymmetry consistent with (3). Second, the exclusivization rate of the nuclear scope of *each* is substantially lower than that of an unquantified upward monotone environment. This fact suggests that monotonicity is not the sole factor deciding between an inclusive and an exclusive interpretation of a plural nominal (as argued in Farkas & de Swart 2010 and Spector 2007).

The rest of this paper is organized as follows. In section 2, we survey the theoretical landscape on plural interpretation, as well as the import of recent experimental investigations. Section 3 provides the details of our four experiments, and section 4 situates these results with respect to the theoretical positions staked out in section 2. In section 5 we conclude with an eye to future work in this area.

2 Background

In this section we review some prominent contemporary theories of number interpretation, concentrating on the role monotonicity entailments play in resolving the interpretation of plurals. In subsections 2.1 and 2.2 we discuss a family of analyses that rely on pragmatic competition but which differ with respect to what the

competitors are. For all of these accounts, the Strongest Meaning Hypothesis (SMH, Dalrymple et al. 1998) serves to adjudicate this competition, and therefore they all predict MG (3). In 2.3, we consider Grimm’s (To Appear) proposal, according to which inclusive plurality is the result of kind-level reference. Finally, in subsection 2.4, we survey how existing experimental work on this topic has helped advance our theoretical understanding of number interpretation.

2.1 Atoms only Competitors: Sauerland et al., Spector, and Zweig

The approaches of Sauerland et al. (2005), Spector (2007), and Zweig (2009) all assume SG restricts reference to atoms, while PL morphology is semantically vacuous. The literal meaning of bare plurals is thus inclusive. Exclusive readings arise via Gricean inference as follows. It is assumed that the inclusive plural competes with the singular form, which is informationally richer. Therefore, if a speaker uses the weaker plural form, her interlocutor may assume that the singular form could not have been used; in many scenarios, an exclusive reading of the plural is the only possibility compatible with what the speaker said.

The various differences between these proposals ultimately reduce to how they acquire the ‘atoms only’ competitor for the plural. For Sauerland et al. (2005), the two competitors are PL and SG where the latter presupposes atomicity of its argument:

- (4) a. $\llbracket \text{PL} \rrbracket = \lambda X.X$
 b. $\llbracket \text{SG} \rrbracket = \lambda X : \text{ATOM}(X).X$

The relevant notion of strength invoked is MAXIMIZE PRESUPPOSITIONS (Heim 1991), and thus exclusivity arises as an anti-presupposition (Percus 2006). The preferences for inclusive and exclusive plurality are derived by assuming that the application of Maximize Presuppositions is restricted as in (5):

- (5) Maximize Presuppositions is enforced in the scope of an existential if this strengthens the entire sentence.

To see how (5) works, consider first the simple declarative in (6), repeated from (1), whose competitor is the singular form in (7).

(6) The baby is playing with teddy bears.

(7) The baby is playing with a teddy bear.

The only difference between these is that (7) has a restriction to atomic interpretation contributed by the SG feature of *a teddy bear* while in (6), PL is semantically vacuous.

Since (6) is stronger than (7), Maximize Presuppositions is in effect. Thus, when (6) is uttered, it must be because saying (7) would lead to presupposition failure. But then *teddy bears* must be exclusive for (6) to still be true. When the plural is in a downward monotone environment, as in (8),

(8) The baby isn't playing with teddy bears.

Maximize Presuppositions is not in effect because the singular version is weaker than the plural one. As a result, no anti-presupposition is computed and the plural remains inclusive.

This analysis predicts MG (3) because of the constraint on Maximize Presuppositions in (5), which encapsulates a form of the SMH. The account then predicts a contrast between the interpretation of bare plurals in the restrictor vs. the nuclear scope of a universal: the former should be inclusive and the latter exclusive.

Whereas Sauerland et al. have PL compete with a presuppositional SG N, Spector (2007) and Zweig (2009) rely on an *exactly 1* inference generated by the singular form. For Spector, this inference arises from the scalar implicatures of the SG sentence directly: (7) implicates that the baby is playing with exactly 1 teddy bear. As this implicature is logically stronger than the interpretation of the PL competitor in (6), not using the SG competitor yields the scalar implicature that the baby is playing with more than 1 teddy bear. Given that the PL competes with a pragmatically enriched version of the SG N, the inclusive plural reading arises in environments in which the implicature of the singular does not arise, such as in downward-monotone contexts.

This account then predicts that the restrictor of *each* should favor inclusive plurality. Its predictions with respect to the nuclear scope depend on details concerning the pragmatic enrichment of the singular competitors in such contexts. Specifically, Spector assumes alongside Chierchia, Fox & Spector (To appear) that (9) can give rise to two distinct implicated meanings, one where the implicature of the singular indefinite is locally calculated (10a) and one corresponding to the global, utterance-level calculation argued for by Grice (10b).

(9) Each baby in yellow is playing with a teddy bear.

(10) a. Each baby in yellow is playing with exactly one teddy bear.

b. Each baby in yellow is playing with at least one teddy bear and at least one baby in yellow is playing with exactly one teddy bear.

Competition with (10a) produces the same strictly exclusive plural interpretation that Sauerland et al. predicts, while competition with (10b), whose truth conditions are weaker, results in an exclusive plural interpretation for one verifying witness:

- (11) Each baby in yellow is playing with teddy bears $\wedge \neg(10a) \leftrightarrow$
Each baby in yellow is playing with more than one teddy bear.
- (12) Each baby in yellow is playing with teddy bears $\wedge \neg(10b) \leftrightarrow$
Each baby in yellow is playing with at least one teddy bear, and at least one
is playing with more than one teddy bear.

To conclude, Spector 2007 allows for more flexibility in interpretation than Sauerland et al. 2005 but the details that regulate the crucial choice between the possible interpretations of (9) are left unclear.

Like Spector (2007), Zweig (2009) offers a route to exclusivity based on scalar inference and a SG competitor. Zweig argues, however, that the competition occurs at the point in interpretation preceding event quantification. The relevant partial logical form for (6) and its singular competitor are shown in (13):

- (13) The baby is playing with teddy bears.
a. $\lambda e.\exists X[*teddy'(X) \wedge *play'(e, \iota_y baby'(y), X)]$
b. SG form: $\lambda e.\exists X[*teddy'(X) \wedge \text{ATOM}(X) \wedge *play'(e, \iota_y baby'(y), X)]$

As the singular form asserts atomicity, it is stronger than the plural form. Thus, exclusive plurality will result from scalar inference. Zweig assumes, following Chierchia (2004), that all possible interpretations of a sentence are computed in this compositional process, and that the SMH chooses the strongest overall meaning. Thus, while the exclusive reading will be stronger locally, the SMH will rule against a local strengthening if (13) is embedded in a downward monotone environment. Hence, in the restrictor of *each*, inclusive readings will be preferred. Conversely, the SMH predicts that in the scope of *each*, exclusivity will be preferred, as in other upward monotone environments.

2.2 Polysemous Competitors: Farkas & de Swart

The three accounts surveyed above achieve exclusivity as an implicature from comparison with the SG form. In contrast, Farkas & de Swart (2010) assume that the PL morpheme is polysemous between inclusive and exclusive interpretations, as schematized below:

- (14) a. $\llbracket \text{INCL-PL} \rrbracket = \lambda X.X$
b. $\llbracket \text{EXCL-PL} \rrbracket = \lambda X : \text{SUM}(X).X$

In their analysis, the competition between the semantically neutral SG form and the polysemous PL one is dictated by constraints that require marking possible reference to sums by using a PL form.

The choice between the two possible interpretations of the PL form in this account falls under the rubric of word sense disambiguation. One crucial factor that plays a role here, [Farkas & de Swart](#) suggest, is the SMH, which favors that interpretation of the PL which results in a stronger overall statement compatible with the context. Like the preceding accounts, this captures the observations about bare plurals in simple declarative and negated environments and predicts that, other things being equal, bare plurals occurring in the restrictor of a universal such as *each* will tend to receive an inclusive interpretation while those occurring in the nuclear scope will tend to receive an exclusive interpretation. The discussion leaves open, however, the possibility that the SMH, which is a pragmatic principle, is overridden by contextual factors such as speaker ignorance with respect to the atom/sum nature of relevant witnesses or cases where the speaker knows that there are both atom and sum relevant witnesses. Such factors may lead to inclusive interpretations of bare plurals in upward entailing environments.

2.3 Kind/Concept-Level Interpretations: Grimm

[Grimm \(To Appear\)](#) presents an account of the interpretation of plurals in which monotonicity direction plays no role. This account, building on [Krifka 1995](#), assumes that nominal denotation can involve *objects* and *concepts/kinds*. Object-level reference is quantized and when an object-referring nominal is marked as plural it receives exclusive plural reference. Concept-level reference is not quantized (i.e., it is number neutral) and the plural marker on a nominal that has concept-level reference is not interpreted, resulting in an interpretation equivalent to an inclusive plural. Whether a bare plural is interpreted exclusively turns on what the question under discussion (QUD) is. If the QUD raises an issue about how many entities are involved, then the bare plural will be interpreted exclusively. If, however, the QUD raises the issue of what type or kind of entities are involved, then the bare plural will be interpreted inclusively. Different types of linguistic and extra-linguistic factors are involved in determining what type of QUD a particular utterance addresses. Because this account is silent on the role of monotonicity in resolving the QUD, it does not predict the exclusivization rate of the restrictor to differ from the exclusivization rate of the nuclear scope.

2.4 Experimental Background

Experimental work on the interpretation of plurality has focused on the claim that exclusive interpretations are generated via implicature. [Sauerland et al. \(2005\)](#) asked both adults and children between 3 years 4 months and 5 years 9 months polar questions about animal anatomy, such as those in (15).

- (15) a. Does a girl have noses?
 b. Does a cat have feet?

In the target condition, (15a), the bare plural involved features where members strictly possess only one of the item in question. In such cases, answering ‘Yes’ would only be compatible with an inclusive plural interpretation, while a ‘No’ would likely indicate rejection of an exclusive plural. Only 4% of children answered ‘No’ (in contrast to 99% of adults), suggesting that children interpret the bare plural inclusively. Given the documented differences in computing scalar implicature between adults and young children (see Gualmini et al. 2001, Noveck 2001, Papafragou & Musolino 2003), Sauerland et al. conclude that this similar difference is suggestive of a common link.

Pearson, Khan & Snedeker (2010) examine adult plural interpretation on the Covered Box Task (Huang, Snedeker & Spelke 2004), a paradigm that reduces participants’ likelihood of computing an implicature. In this task, participants are asked to: a) choose the one object out of an array that best meets a description, and b) if no object matches the description, choose an occluded object (the “covered box”). Huang et al. show that when participants are given a description containing *some*, but where the object array contains only objects satisfying *all* or *no*, participants choose an *all* object over choosing the covered box, apparently accessing the literal interpretation of *some*. Following this logic, Pearson et al. performed two Covered Box Task experiments. In the first, participants were given prompts such as "*Point to the card where Big Bird has kites*" and arrays containing a picture of Big Bird with nothing and a picture of Big Bird with one kite. Choice of the covered box in such a setup would indicate an exclusive plural interpretation. Participants chose the covered box 96% of the time. In the second experiment, prompts were manipulated to be of the form *Point to the card where Big Bird has only kites.*; the covered box was chosen 35% of the time, significantly less than in the first experiment. Pearson et al. suggest that the low rate of covered box selection in the second experiment indicates that plural exclusivization is an implicature, but that the high rate in the first experiment indicates that this implicature is relatively hard to cancel.

Finally, Grimm (To Appear) investigates the importance of normative contexts in preferences for plural exclusivity. Participants on Amazon’s Mechanical Turk service were shown image-polar question pairs where the polar question contained a bare plural (e.g., *Is the woman holding mugs?*) and where the image depicted either atom or sum referents for the PL description. Without additional context, participants assented to sum images 92% of the time and to atom images only 32% of the time. However, when participants were told that the question was posed in a corporate setting and asked to respond as they thought an employee would, they assented to atom images 78% of the time (and to sum images 99% of the time). As Grimm

assumes that normative contexts induce kind-favoring QUDs, this contrast is taken to demonstrate the importance of QUD types for plural interpretation.

These three investigations help to chart the contextual factors influencing plural interpretation. All three indicate that whatever the source of exclusivity is, it is a persistent effect in adult grammars: overwhelmingly, adults interpreted the bare plurals in unelaborated scenarios as exclusive. However, the prevalence of this interpretation is diminished in scenarios where the relevance of the singular/plural distinction is obviated, as in those Grimm employs, as well as in grammatical environments such as the prejacet of *only*. It is less clear how they bear on the relation between exclusivization and implicature calculation. Grimm’s scenarios test for the influence of relevance, a persistent factor across several accounts surveyed. Similarly, Sauerland et al. compare adult performance to that of children of an age when fairly radical changes in pragmatic competence are occurring. Finally, we suspect that the difference Pearson et al. observe is due more to the presence of *only* than the implicature-suspending properties of the Covered Box Task; we take their results to provide another environment in which inclusivity is strongly preferred.

3 Experimental Investigations

None of the experiments above directly examine the Monotonicity Generalization. We conducted four experiments to investigate the degree to which it holds. All experiments were image verification studies, in which we presented participants with image-description pairs and asked them to judge whether the text accurately described the image. For all items, the descriptions were sentences containing the strong distributor *each* in subject position. In this way, the distinction between upward and downward monotonicity could be tested by comparing the interpretation of bare plurals in the restrictors and nuclear scopes. For Experiments 1-3, the main difference between items concerns the number and nature of exemplars in the image that would satisfy exclusive PL interpretations, as illustrated in Figure 1. Experiment 4 then contrasts these with unquantified descriptions and images containing only one exemplar. In subsection 3.1 we outline those design features that our individual experiments share, and then discuss each experiment in turn.

3.1 Design of Experiments 1-3

3.1.1 Materials

In the first three experiments, the target trials contained descriptions with the universal quantifier *each* in subject position. We chose *each* to ensure a distributive interpretation and thus rule out the possibility of cumulative readings, which exist

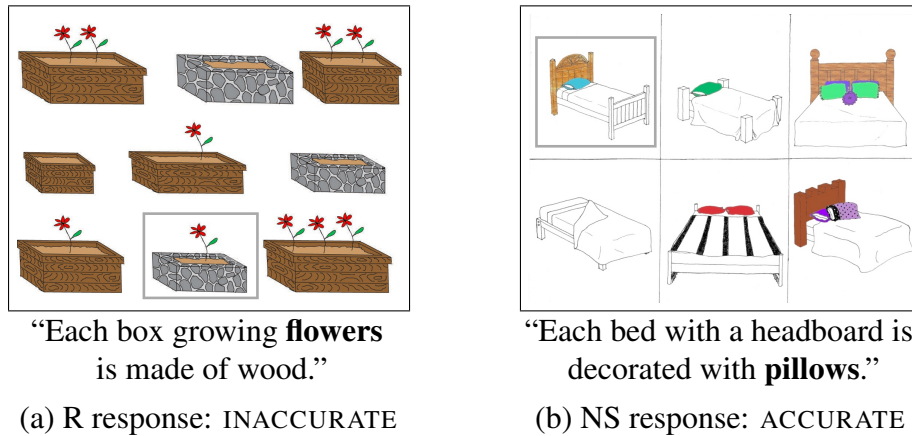


Figure 1 Image-description pairs and expected exclusive PL response for sample Restrictor item (a) and Nuclear Scope item (b). Controversial witnesses are highlighted.

for *all* and, to a lesser degree, *every*. In *Restrictor* (R) conditions, such as Figure 1a, a bare plural noun occurred in the quantifier’s restrictor (“Each box growing **flowers** is made of wood.”). In *Nuclear Scope* (NS) conditions (Figure 1b), a bare plural noun occurred in the quantifier’s nuclear scope (“Each bed with a headboard is decorated with **pillows.**”).

The images were scenes composed of between four and ten individual exemplars. In the NS condition, individual exemplars were either *witnesses*, which satisfied the restrictor, or *non-witnesses*. In the NS example in Figure 1, only exemplars containing headboards would be witnesses. The truth of the description turns on the perception of a *controversial witness*, one which contains only an atomic depiction of the bare plural. In Figure 1, this witness is highlighted in grey: it is a bed with a headboard and a single pillow. Non-witnesses, which do not affect the truth of the sentence, consist both of depictions that incidentally satisfy the nuclear scope and ones which did not.

In the R condition, individual exemplars were likewise witnesses if they satisfied the restrictor. As in the NS condition, witnesses were controversial witnesses if they contained an atomic depiction of a bare plural noun. Such controversial witnesses always satisfied the nuclear scope. In the R example in Figure 1, only boxes in which flowers grow are potential witnesses. The controversial witness, highlighted in grey, is a box which contains a single flower.

The stimuli were designed such that the weaker plural interpretation was always true and the stronger interpretation always false. Thus, for R trials, an inclusive reading would be false and the exclusive reading true. For NS trials, on the other

hand, the inclusive reading would be true and the exclusive reading false.

Filler items were constructed such that truth of the descriptions relative to the image did not depend on the interpretation of any bare plural noun. These included descriptions containing *pluralia tantum*, positional expressions with implicit arguments (e.g., *on the left/right*), and general properties (e.g., color, size). Fillers spanned a range of difficulty and perceived ambiguity.

In Experiment 1, there were 24 target trials (12 NS, 12 R) and 20 filler trials. In Experiment 2, there were 16 target trials (8 NS, 8 R) and 20 fillers. Finally, in Experiment 3, there were 8 target trials (4 NS, 4 R) and 20 fillers. We successively decreased the proportion of target trials due to a concern that participants were detecting the key trials and deploying a strategic response pattern.

A full list of materials and detailed analyses for all experiments discussed in this paper (including an eyetracking version of Experiment 1), can be found at <http://linglabs.org/plurals/>.

3.1.2 Procedure

Participants took the experiment on a web-based platform built for this task. The instructions were as follows: *The following experiment will ask you to read a sentence and look at a picture describing a situation. You will then be asked to determine if the sentence is an accurate description of the picture or not. Take as much time as you need, but do not feel pressured to find the “correct” answer. These are not trick questions, but some of the sentences may be more complicated than others.* After completing 4 practice examples with feedback, participants proceeded through the target and filler trials. These were randomized and interleaved such that no target stimuli were adjacent. Interested readers may take the experiment for themselves, which can be accessed on the authors’ web site.

3.1.3 Analysis

In our data we report the *exclusive PL response rate*. For R trials, this corresponds to the percentage of image-description pairs *accepted*, while for NS trials, it corresponds to the percentage of image-description pairs *rejected*. For statistical analysis, these by-participants rates were submitted to a logistic mixed-effects model.

We also report an *asymmetry index*: this corresponds to a d' score obtained by assuming that exclusive responses on R trials are incorrect (“misses”) and exclusive responses on NS trials are correct (“correct rejections”) (MacMillan & Creelman 2005). This value is calculated by taking the difference of the normal quantile function when it is applied to the two exclusive PL response rates. The d' score, whose origin is in signal detection theory, has useful interpretations in psychology,

but for our purposes we use it simply as a convenient distance metric. When the response rates are far from one another (a large asymmetry), then d' is large. When the response rates are close to one another (a small asymmetry), then d' is small. When they are equal, d' is zero.

No existing theory provides a basis for predicting the exact magnitude of the asymmetry. However, consider three scenarios as points of reference. If the R-NS exclusivization rates reflect an absolute asymmetry – 0%-100% – then the index would be approximately **3.5**. If the asymmetry were softer – 25%-75% R-NS exclusivization – then the index would be **1.4**. Finally, suppose R enforces inclusive interpretations, whereas contextual factors determine the NS rate. In this scenario, if R-NS exclusivization is 0%-50%, then the index would be **1.6**. We thus expect a positive asymmetry index and our intuitive scenarios suggest it will be above 1.

3.2 Experiment 1: Mixed sum and atom referents

3.2.1 Introduction and design

In Experiment 1, the images against which the truth of the description had to be judged contained a mixture of controversial and uncontroversial witnesses. That is, there were exemplars containing both atom and sum depictions. Only when sums are present is the exclusive use of the plural felicitous. Thus only when there is a mixture of sums and atoms is it predicted that there will be competition between a 2+ interpretation of the plural and a 1+ interpretation. It is under these circumstances that most theories predict an asymmetry between upward and downward monotone environments. An example of a ‘mixed sum and atoms’ scene is given in Figure 2 (top panel, left). If the monotonicity generalization (3) is correct, the rate at which participants adopt an exclusive PL response should be much higher in NS trials than in R trials, and therefore the asymmetry index should be positive.

For the average NS trial, there were 6 exemplars. Of 3-4 total witnesses, 1-2 were controversial witnesses that contained only an atomic depiction of the plural. For the average R trial, there were 7 exemplars and 4 witnesses. 3 witnesses uncontroversially satisfied both restrictor and nuclear scope and 1 was more controversial, containing only an atomic depiction of the plural. 27 persons participated in this experiment for course credit in an introductory linguistics course.

3.2.2 Results and Discussion

By-participants exclusive PL response rates are plotted in Figure 2 (bottom panel, left plot), with NS rates on the x -axis and R rates on the y -axis. The average R rate was 24% and the average NS rate was 34%, yielding an asymmetry index of 0.29.

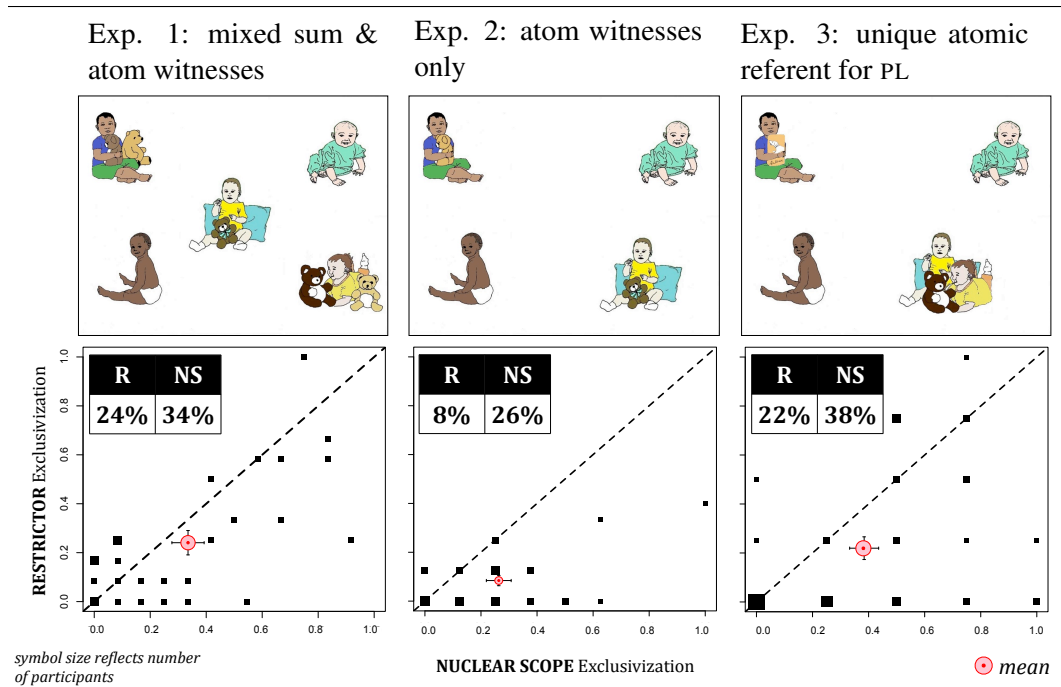


Figure 2 Sample items and participant PL exclusivization rates for Experiments 1-3. *Top Panel.* Images corresponding to the description “Each baby wearing yellow is playing with teddy bears.” *Bottom Panel.* Participant PL response rates (NS rates on the x-axis, R rates on y-axis). Symbol size corresponds to number of participants. Dashed line indicates where points would lie if NS and R rates were equal. Mean rates are inset in the top left corner and indicated by a red bulls-eye.

The effect of the R v. NS environment was significant ($p < .005$).

That there is an R-NS asymmetry in the predicted direction can be appreciated by observing that most plotting symbols lie below the line of equality. It is striking however that the overall rate at which participants adopt an exclusive plural interpretation is rather low: around 30% of the time. This can be appreciated by noticing that most of the points on the plot crowd around the origin.

Thus the basic prediction was upheld: participants were significantly more likely to adopt an exclusive plural interpretation in the NS conditions, when the plural was in an upward monotone environment, than in the R conditions, when the plural was in a downward monotone environment.

It was surprising, however, that the overall rate of exclusive interpretation was so low. The exclusive PL response rate in NS conditions, which was predicted to

be maximal, was less than 50%. Thus being in the nuclear scope of the universal quantifier *each* is not enough to guarantee that an exclusive PL interpretation will prevail. This finding is consistent with the suggestion that relevance and other contextual parameters play a strong role in addition to monotonicity (Farkas & de Swart 2010; Grimm To Appear).

In Experiments 2 & 3 we tested the hypothesis that the rate of exclusivization in upward monotone environments would increase when the singular form was a salient competitor. This is a prediction of the competition based accounts introduced earlier: if the context contains only relevant individuated atom witnesses, the singular form should be preferred and the plural form should be assigned the stronger exclusive interpretation. In Experiment 2, we convert all sum witnesses from Experiment 1 into atom witnesses. In Experiment 3, we remove all but one atom witness to guard against a cumulative interpretation.

3.3 Experiment 2: Atoms only witnesses

3.3.1 Introduction and design

The goal of Experiment 2 was to test the hypothesis that exclusive and inclusive interpretations of the PL compete, such that the exclusive plural arises when it is the strongest interpretation available. We modified the scenes from Experiment 1 such that all witnesses were atomic witnesses. If the scenes contain only atomic witnesses, the exclusive use of the plural is not compatible with the context. Moreover, if the singular form is in the relevant set of competitors, it should outcompete any plural form. Consequently we expect the exclusivization rate of the nuclear scope to increase.

An example scene with only atomic witnesses can be found in Figure 2 (top panel, middle). The stimuli in Experiment 1 were modified such that depictions of more than one referent in any given exemplar were changed into depictions of exactly one referent. The descriptions remained identical. 28 persons participated in this experiment.

3.3.2 Results and Discussion

By-participant R and NS exclusive PL response rates are plotted in Figure 2 (bottom, middle). The average R rate was 8% and the average NS rate was 26%, corresponding to an asymmetry index of 0.97. The effect of environment was significant ($p < .001$).

Participants in this experiment were even more liable to adopt an inclusive interpretation of the plural than participants in Experiment 1. Moreover, the asymmetry between upward and downward monotone environments is preserved. Indeed, the

magnitude of the asymmetry increased, with the R conditions seeming to promote inclusive readings to a greater extent than in Experiment 1. However, the conditions for judging a response ‘Accurate’ are stricter for Experiment 2 R conditions. In Experiment 1, if a participant considered only the sum witnesses, which were uniformly verifiers, the scene was accurate. As these have been eliminated in Experiment 2, for a participant to mark ‘Accurate’, she must not only exclusivize the plural, but also consider a sentence trivially true when *each* quantifies over a null domain. Given this latter constraint, we may be under-reporting the restrictor exclusivization rate (but see footnote 2 for an attempt to measure this effect in Experiment 3).

Exclusive plural readings were expected to increase in the NS conditions, but instead they decreased. One possibility is that, because participants are confronted with image-description pairs in which plural nouns were used whose depicted exemplars were uniformly atomic, they assumed the sum/atom distinction was simply irrelevant. Consequently, there was no competition between available forms to realize sums versus atoms.

3.4 Experiment 3: Unique atom referent for PL

3.4.1 Introduction and design

One potential concern about Experiment 2 is that participants were coercing a cumulative interpretation of the quantifier; atomic exemplars could license the plural if participants cumulated those atoms into one sum referent. To address this concern, in Experiment 3 we modified the scenes such that there was now a unique referent for the PL which was atomic.

An example scene with only one witness, which was atomic, can be found in Figure 2 (top panel, right). The stimuli in Experiment 2 were modified such that only one depiction of an atomic PL referent was preserved. To maintain the appropriateness of the universal quantifier, we further modified the images such that there were multiple depictions of the head noun of the quantified phrase interacting with the single, atomic referent. In the example in Figure 2, this resulted in two babies playing with the same teddy bear. If participants accepted plural descriptions in Experiment 2 because they could access a cumulative representation of the multiple atomic exemplars, then they should disfavor the plural descriptions when that cumulative representation is removed. Thus exclusive plural interpretations should rise across the board.

The second issue we addressed concerns the composition of the stimulus materials. In Experiments 1 and 2, the scenes and descriptions comprising the NS and R trial sets respectively were different. The use of different sets of lexical materials and different depicted scenes could be associated with different rates of exclusiviza-

tion. This is consistent with assumptions made by, e.g., Farkas & de Swart (2010), that real-world knowledge plays a crucial role in balancing the appropriateness of plural interpretation. We may, therefore, have inadvertently selected plural nouns in our R conditions that were more disposed to inclusivization per se or, consistent with Grimm, more disposed to kind-level interpretation. To remove this potential confound, we revised our descriptions and scenes such that for any given scene, we could form an NS description and an R description. For example, the scene presented in Figure 2 has both an NS description, “Every baby wearing yellow is playing with teddy bears”, and an R description, “Every baby playing with teddy bears is wearing yellow.” Thus the R and NS conditions in Experiment 3 were balanced for lexical items. In some cases it was necessary to compose entirely new scenes.¹ 41 persons participated in this experiment.

3.4.2 Results and Discussion

By-participant R and NS exclusive PL response rates are plotted in Figure 2 (bottom, right). The average R rate was 22% and the average NS rate was 38%, corresponding to an asymmetry index of 0.46. The effect of environment was significant ($p < .05$).

In both R and NS conditions, the rate of exclusive PL responses increased relative to Experiment 2. Crucially, however, the rates are not significantly different from those observed in Experiment 1. Therefore it seems unlikely that a cumulative interpretation was responsible for the failure to see an increase in nuclear scope exclusive PL response rates between Experiments 1 and 2. Moreover, the continued appearance of an R/NS asymmetry suggests that mere lexical differences were not generating that contrast in prior experiments.²

3.5 Experiment 4: Unquantified descriptions

3.5.1 Introduction and Design

In Experiment 4, we sought to assess why rates of exclusive PL interpretation were so low in Experiments 1-3. Our intuition was that the rate of exclusive plural interpretations should be reasonably high in the NS conditions – at a minimum, greater than 50%. There are no theories on the market that directly predict the

¹ Several stimuli from Experiments 1 and 2 involved relations that could not simultaneously hold of one PL referent and several witnesses. These were replaced with novel descriptions.

² Experiment 3 also included a three item followup to the experiment to test participants’ response patterns and justifications for null quantificational domains (e.g., “Each castle surrounded by a moat is gray” coupled with an image where no castle has a moat). Participants who indicated that such items violated a presupposition were significantly more likely to have exclusivized plurals in both the restrictor and nuclear scope. See the web resources for further details.

baseline rate of exclusivity, but this is true in general when adapting semantic theories to the operational notions that experiments require, like ‘rates of interpretation.’ However, it seems reasonable to conclude that if monotonicity is a dominant factor in determining the appropriateness of the exclusive plural, the exclusive plural should predominate in upward monotone environments. In Experiments 1-3, we observed that a weaker version of this prediction was upheld: exclusive interpretations are only relatively more abundant in upward monotone environments.

In Experiment 4, we measured the baseline exclusivization rates for our lexical items and images by presenting image-description pairs that consisted of unquantified descriptions and singleton exemplars. To do so, we extracted both sum and atom exemplars from the scenes used in Experiment 1-3 and paired them with descriptions in which *each* was replaced by the definite determiner *the*. Images depicting either an atom or sum NP referent were paired with sentences containing a SG or PL description. Plurals were always contained inside VP; however the noun either corresponded to R conditions or NS conditions in Experiment 1-2. 10 critical nouns were from the R conditions (R-class nouns) and 10 from the NS conditions (NS-class nouns).

If the exclusivization rates in the NS conditions of Experiments 1-3 (26% - 38%) reflect the natural variability in exclusivization rates which are determined by pure lexical biases or world knowledge, then we expect the rate of exclusive interpretation in this experiment to be in that range. In other words, the appropriateness of using plural descriptions with atom images should be accepted more often than not.

Data for this experiment were collected on Amazon Mechanical Turk. Each participant judged 4 filler trials and 1 target trial. For each of the eight target conditions, we collected 160 judgments.

3.5.2 Results and Discussion

The endorsement rates for each image-description pair are given in Table 1. The crucial rate, corresponding to the exclusive PL response rate in Experiments 1-3, is how often participants endorse the use of a plural noun when an atom is depicted (highlighted in Table 1). Overall there was an effect of noun number ($p < .001$) and an interaction between noun number and the number of objects depicted ($p < .001$). There was, however, no effect of noun class – i.e., whether the noun was drawn from the R conditions or NS conditions.

On average, participants endorsed this inclusive use of the plural 17% of the time, corresponding to an exclusive PL response rate of 73%. This rate is nearly double what was observed in Experiments 1-3. In cases where a sum was depicted but a singular noun was used, participants endorsed the image-description pair 57% of the time, on average. This is unsurprising, since the singulars were expressed as

	Exemplar Type			
	ATOM		SUM	
	R-class	NS-class	R-class	NS-class
N Number				
SG	99%	99%	60%	54%
PL	17%	17%	97%	98%

Table 1 Experiment 4 ACCURATE response rates, varying by number morphology (SG vs. PL), number of referents in the image (ATOM vs. SUM), and source of the nominal head (R-class vs. NS-class). Shaded boxes correspond to controversial witnesses in Experiments 1-3.

indefinites and a singular indefinite description is true of a sum depiction: if a person is depicted holding more than 2 gifts, that person is also depicted holding one gift.

This experiment leads to two conclusions. Firstly, the R/NS asymmetry observed in Experiments 1-3 is not an accident of the nouns/pictures chosen for each condition. Secondly, the low exclusivization PL rates in Experiments 1-3 do not characterize the nouns we selected in a general way: in this experiment, the exclusive PL rate was much higher. We conclude that it was the syntactic/semantic environment of the plural that affected its overall low exclusivization rate in the earlier experiments.

4 General Discussion

Let us take stock. We have observed a statistically significant difference between exclusivization rates of the restrictor and nuclear scope of *each*. However, the nuclear scope exclusivization rate was roughly two-fifths of that for the unquantified items in Experiment 4, and Experiments 2 & 3 document that removing sum witnesses does not appreciably raise nuclear scope exclusivization.

This set of findings is puzzling for all of the theories outlined in section 2. The competition based accounts predict a contrast in exclusivization rates correlating with monotonicity direction, a contrast that is confirmed. However, these accounts have no explanation for the fact that plurals in unquantified descriptions and upward monotone quantified environments appear not to behave alike. While these theories may have some flexibility in accounting for the mixed witnesses images in Experiment 1, they all predict that bare plural descriptions should be unacceptable for pure atom witness images, contrary to what we observed in Experiments 2 and 3. [Grimm \(To Appear\)](#) predicts no relation between exclusivization and sum/atom witnesses, and so the lack of contrast between Experiments 1-3 in nuclear scope exclusivization is not problematic. However, this account does not predict the difference between restrictor and nuclear scope exclusivization rates that our experiments confirmed.

When we compare Experiments 1-3 with Experiment 4, it appears that simply being in the scope of quantification depresses exclusivization. Why should this be? Most of the theories surveyed highlight the importance of contextual relevance in plural interpretation. In our items, the only sources of relevance cues are provided by the experiment itself: the form of the description, the organization of the scene, and the prior history of the experiment. It may be that image arrays and quantified descriptions suggest to participants that, in Grimm's terms, there is a kind-favoring QUD. Suggestively, recent experimental work on the interpretation of scalar terms in quantified environments shows that implicature calculation rates can vary drastically depending on stimulus construction and task design (see [Geurts & Pouscoulous 2009](#), [Clifton & Dube 2010](#), and [Chemla & Spector 2011](#) for details).

Another possibility concerns the nature of the image verification task itself. Little is known about the procedures participants employ in verifying a description against a scene. In an eye-tracking version of Experiment 1 (not reported here), we observed that participants concentrate their attention in nuclear scope cases on witnesses, but we do not know how witnesses were filtered into verifiers and falsifiers. One possibility is that participants do not try to verify a description, but to falsify it. In such a case, a participant will be looking for a witness that fails to satisfy the nuclear scope predicate – for example, she will be looking for a baby wearing yellow that isn't playing with teddy bears. If such a strategy is adopted, inclusivization will be the predominant interpretation of the PL. More detailed understanding of the timecourse of image verification is necessary to address this speculation.

5 Conclusion

We have shown that the Monotonicity Generalization in (3) is confirmed: the exclusivization rate in the restrictor of *each* is lower than in the nuclear scope. The difference in rates is, however, quite modest. We found, however, that there is a greater difference between two upward monotone environments, namely the nuclear scope of *each* and upward monotone unquantified assertions: the former environment is significantly more inclusivization-friendly than the latter, even in cases where no sum witnesses were present in the scenes used for testing quantificational sentences. This is a novel result and we suggest that the quantificational environment itself is responsible for the increase in inclusivization. The origin of this effect is a matter of speculation at present. We have hypothesized that unrecognized and hitherto uncontrolled contextual factors – how quantificational environments alter relevance assumptions or trigger different verification strategies – may be at fault. If so, this indicates that contextual factors occupy an important role in plural interpretation (as emphasized by [Farkas & de Swart](#)), and that such factors should be the focus of much greater scrutiny in both experimental and theoretical work on plural interpretation.

References

- Chemla, Emmanuel & Benjamin Spector. 2011. Experimental evidence for embedded scalar implicatures. *Journal of Semantics* 28(3). 359–400. doi:10.1093/jos/ffq023.
- Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Andrea Belletti (ed.), *Structures and Beyond: The Cartography of Syntactic Structures*, 39–103. New York: Oxford UP.
- Chierchia, Gennaro, Danny Fox & Benjamin Spector. To appear. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Paul Portner, Claudia Mainborn & Klaus von Heusinger (eds.), *Handbook of Semantics*, Berlin: Mouton de Gruyter.
- Clifton, Charles & Charles Dube. 2010. Embedded implicatures observed: A comment on Geurts and Pouscoulous (2009). *Semantics and Pragmatics* 3(7). 1–13. doi:10.3765/sp.3.7.
- Dalrymple, Mary, Makoto Kanazawa, Yookyung Kim, Sam Mchombo & Stanley Peters. 1998. Reciprocal expressions and the concept of reciprocity. *Linguistics & Philosophy* 21. 159–210.
- Farkas, Donka. 2006. The unmarked determiner. In Svetlana Vogeleer & Liliane Tasmowski de Rijk (eds.), *Non-definiteness and plurality*, 81–106. Amsterdam: John Benjamins.
- Farkas, Donka & Henriette de Swart. 2010. The semantics and pragmatics of plurals. *Semantics and Pragmatics* 3(6). 1–54. doi:10.3765/sp.3.6.
- Geurts, Bart & Nausicca Pouscoulous. 2009. Embedded implicatures?!? *Semantics and Pragmatics* 2(4). 1–34. doi:10.3765/sp.2.4.
- Grimm, Scott. To Appear. Plurality is Distinct from Number Neutrality. In Yelena Fainleib, Nick La Cara & Yangsook Park (eds.), *North East Linguistic Society (NELS) 41*, Amherst, MA: GLSA.
- Gualmini, Andrea, Stephen Crain, Lisa Meroni, Gennaro Chierchia & Maria T. Guasti. 2001. At the semantics/pragmatic interface in child language. In Rachel Hastings, Brendan Jackson & Zsofia Zvolenszky (eds.), *Semantics and Linguistic Theory (SALT) 11*, 231–247. Ithaca, NY: CLC Publications, Cornell University.
- Heim, Irene. 1991. Artikel und Definitheit. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, 487–535. Berlin & New York: de Gruyter.
- Huang, Yi-ting, Jesse Snedeker & Elizabeth Spelke. 2004. What Exactly do Numbers Mean? In Kenneth Forbus, Dedre Gentner & Terry Regier (eds.), *26th annual meeting of the Cognitive Science Society*, 1570. Mahwah, NJ: Lawrence Erlbaum.
- Krifka, Manfred. 1995. Common Nouns: A Contrastive Analysis of English and Chinese. In Manfred Krifka, Greg Carlson & Jeff Pelletier (eds.), *The Generic*

- Book*, 389–411. Chicago: Chicago UP.
- MacMillan, Neil & C. Douglas Creelman. 2005. *Detection Theory: A User's Guide*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Noveck, Ira. 2001. When children are more logical than adults: experimental investigations of scalar implicature. *Cognition* 78. 165–188. doi:10.1016/S0010-0277(00)00114-1.
- Papafragou, Anna & Julian Musolino. 2003. Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition* 86. 253–282. doi:10.1016/S0010-0277(02)00179-8.
- Pearson, Hazel, Manizeh Khan & Jesse Snedeker. 2010. Even more evidence for the emptiness of plurality: An experimental investigation of plural interpretation as a species of implicature. In Nan Li & David Lutz (eds.), *Semantics and Linguistic Theory (SALT) 20*, 489–508. Ithaca, NY: CLC Publications, Cornell University.
- Percus, Orin. 2006. Antipresuppositions. In *Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science* Report of the Grant-in-Aid for Scientific Research (B), Project No. 15320052, 52–73. Japan Society for the Promotion of Science.
- Sauerland, Uli. 2003. A new semantics for number. In Robert B. Young & Yuping Zhou (eds.), *Semantics and Linguistic Theory (SALT) 13*, 258–275. Ithaca, NY: CLC Publications, Cornell University.
- Sauerland, Uli, Jann Anderssen & Kazuki Yatsushiro. 2005. The plural is semantically unmarked. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives*, 413–434. Berlin: Mouton de Gruyter.
- Spector, Benjamin. 2007. Aspects of the pragmatics of plural morphology: On higher-order implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presuppositions and implicatures in compositional semantics*, 243–281. London & New York: Palgrave Macmillan.
- Zweig, Eytan. 2009. Number-neutral bare plurals and the multiplicity implicature. *Linguistics & Philosophy* 32(4). 353–407. doi:10.1007/s10988-009-9064-3.

Anand, Andrews, Farkas, and Wagers

Pranav Anand
Stevenson College
1156 High Street
Santa Cruz, CA 95064
panand@ucsc.edu
<http://people.ucsc.edu/~panand>

Matthew Wagers
Stevenson College
1156 High Street
Santa Cruz, CA 95064
mwagers@ucsc.edu
<http://people.ucsc.edu/~mwagers>

Donka Farkas
Stevenson College
1156 High Street
Santa Cruz, CA 95064
farkas@ucsc.edu
<http://people.ucsc.edu/~farkas>