Collective preferences in Dutch revealed by a covered-box experiment

de Koster, Anna; Spenader, Jennifer; Hendriks, Petra

Published in:
Linguistics in the Netherlands

DOI:
10.1075/avt.00037.kos

Collective Preferences in Dutch Revealed by a Covered-Box Experiment

Anna de Koster, Jennifer Spenader and Petra Hendriks

Abstract

Sentences with plural expressions are compatible with distributive and collective interpretations. Adults generally prefer collective interpretations, whereas children do not. Dotlačil (2010) argues that the adult collective preference arises via an implicature. Adults can reason about alternative utterances with the distributive marker *each*, thereby ruling out distributive interpretations in favor of collective interpretations. Experiment 1 used the covered-box paradigm to investigate whether adults and children make the comparisons predicted by Dotlačil's implicature account. Adults' responses suggest that they made comparisons with internally generated alternatives, supporting the implicature account. Moreover, children seem to do so from around 11 years old onwards, after they have learned the distributive character of *each*. Experiment 2 excluded the possibility that our results in Experiment 1 were influenced by participants' exposure to both collective and distributive pictures, making the collective interpretation more salient. Both experiments thus point towards an implicature underlying the adult collective preference.

Keywords: Language Acquisition, Quantification, Distributivity, Conversational Implicature, Covered-Box Experiment

1. Introduction

A fundamental problem in studying sentence interpretation is determining what information is semantic, and thus coded as part of the lexical meanings of words, and what information is derived pragmatically, through some type of pragmatic reasoning. Plural expressions present a clear example where drawing this distinction is a challenge, especially in characterizing interpretations in both adults and children.

(1) The girls are carrying a box.
(2) Each girl is carrying a box.

A plural event, such as that described by (1), is semantically compatible with more than one interpretation (e.g., Champollion, 2017; Landman, 2000). Sentence (1) could be interpreted as if the girls are individually carrying their own box, a distributive interpretation. But it could also be interpreted as if the girls are carrying one box together, a collective interpretation. The syntactic structure and the choice of determiners for the subject and the object do not disambiguate (1). However, without other cues, mature language users prefer the collective interpretation (Frazier et al., 1999; de Koster et al., 2017; Padilla-Reyes, 2018; Pagliarini et al., 2012). For (2), the situation is different. Here, the distributive interpretation is strongly preferred, with the collective interpretation being almost impossible (Brooks and Braine, 1996; de Koster et al., 2017; Padilla-

Reyes, 2018; Pagliarini et al., 2012). This is because the universal quantifier *each* is semantically distributive and therefore disambiguates (2).

However, for these same sentences, children have been shown to allow quite different interpretations in experimental studies (Brooks and Braine, 1996; de Koster et al., 2017; Pagliarini et al., 2012; Syrett and Musolino, 2013):

(i) Whereas adults prefer a collective interpretation for plurals like (1), children do not. They accept both distributive and collective interpretations at an equal rate, showing no clear preference.

(ii) Whereas adults find a collective interpretation of the distributive marker *each* (2) marginal, children until age 8 do not; only from age 8 they show a preference for the distributive interpretation. This difference has often been attributed to children having an incomplete semantic representation of *each,* lacking the information that it has distributive meaning (Dotlačil, 2010; de Koster et al., 2017; Padilla-Reyes, 2018; Pagliarini et al., 2012). This lack of knowledge results in the acceptance of *each* in collective contexts. How children arrive at the adult preferences remains unclear.

One account of the development of these different preferences (Dotlačil, 2010) considers the distributive preference in (2) to be due to the lexical semantics of *each*, but attributes the collective preference in (1) to a pragmatic process: a conversational implicature (cf. Grice, 1975). The reasoning is as follows: language users prefer to interpret plural events without distributive marking, such as (1), as collective because if speakers intended distributivity they would have used (2) with the more informative distributive marker *each*. The absence of distributive marking means that the weaker, collective interpretation is intended.

This account assumes that the implicature calculation is driven by the lexical-semantic development of *each.* Children have to learn that *each* is distributive before they can use its absence to draw an implicature, leading to a collective preference for unmarked plurals. Children who have not yet learned the lexical semantics of *each*, will give similar interpretations to unmarked plurals and plurals modified with *each*. Only when they realize that *each* is semantically distributive, can they reason that its absence is meaningful. This leads to the following hypothesis: only after children have a complete understanding of *each* will they be able to calculate the predicted implicature that underlies the adult collective preference with definite plurals. Evidence for this hypothesis has been found in several studies testing different languages (Dutch: de Koster et al., 2017, Spanish: Padilla-Reyes, 2018, Italian: Pagliarini et al., 2012).

One obstacle to interpret these earlier results, however, lies in the methodologies used: previous work used picture verification tasks, truth-value judgment tasks and picture selection tasks, but these tasks may not reveal all aspects of implicature calculation, especially in children (Katsos and Bishop, 2011). In this paper we therefore use a covered-box task (Huang et al., 2013), which indicates whether participants believe there is a better meaning for the sentence than the meaning they are presented with.

We discuss two experiments. Experiment 1 used the covered-box paradigm to investigate the origin of the adult collective preference. Because Experiment 1 included both distributive and

collective pictures, it is conceivable that the presence of alternatives between items facilitated implicature calculation. In Experiment 2 we therefore used a between-subjects design to rule out this explanation.

2. Methodological Considerations

Picture verification tasks (PVTs) and truth-value judgement tasks (TVJTs) have been widely used to test distributivity. Investigating distributivity, both de Koster et al. (2017) and Pagliarini et al. (2012) used PVTs where participants received sentence-picture pairs without a narrative context. Padilla-Reyes (2018) used a TVJT by showing participants stop motion videos in a narrative context.

However, since sentences with and without *each* (e.g., (2) and (1)) are both semantically compatible with collectivity and distributivity, PVTs and TVJTs do not reveal much about why participants accept or reject certain items. Some studies using a TVJT ask participants to justify their responses, but these justifications are seldom analyzed and the main focus remains on participants' responses. Both these tasks ask participants to verify whether a sentence is an accurate description of a picture. This will lead to high acceptance rates, even if participants dislike a certain description, since both descriptions are semantically true. These tasks may therefore not be sensitive enough to investigate implicatures. The information they provide is whether participants accept a certain interpretation or not, but they do not show whether participants reason about alternatives. Since reasoning about alternative options is a crucial element in implicature calculation, a task requiring the comparison of alternatives could be more informative.

In contrast to PVTs and TVJTs, picture selection tasks (PSTs) directly require comparing alternatives by asking participants to choose among two or more pictures to indicate which of these pictures matches the accompanying sentence best. However, this task also has several shortcomings: if both interpretations are possible but one is preferred, picture selection will often lead to extreme responses for the preferred option, masking the fact that the less preferred option is also possible. The main shortcoming related to our study is that it can only test implicature calculation indirectly. Because alternative interpretations are explicitly presented as non-target pictures, participants do not have to come up with alternative interpretations themselves. For this reason, it can only provide indirect evidence for implicature calculation. It shows which preferences participants have, but it does not provide evidence that participants rely on a comparison process when no explicit alternatives are presented.

In this study we instead use the covered-box paradigm developed by Huang et al. (2013). In this paradigm, participants are asked to choose between a visible and a covered picture. They are instructed to choose the covered picture if they believe there is a better option than the visible picture. Thus, by using this paradigm, participants are still encouraged to consider alternative options, but unlike in a PST, they have to generate these options themselves, comparing them to the provided picture. This method offers a way to tap into participants' internal representation of sentence forms and their potential meanings, making the covered-box paradigm a particularly suitable method to test Dotlačil's implicature account. Our aim is to investigate children's development of the collective preference after they have acquired knowledge of the distributive character of *each*. Since previous studies found that children start to show a collective preference

for plural expressions around age 11 (Pagliarini et al., 2012; de Koster et al., 2017), our study focuses on children of approximately 11 years old.

3. Experiment 1

3.1 Participants

24 Dutch adults (mean age 23, range 19-35) and 22 Dutch children (mean age 10;11, range 10;2-12;2) participated in the experiment. The children were recruited from a primary school in the province of Groningen, the Netherlands, and performed the experiment in a quiet room in their school. The adults, mainly students from the University of Groningen, performed the experiment online through a website, without the experimenter being present, and received €5 for their participation.

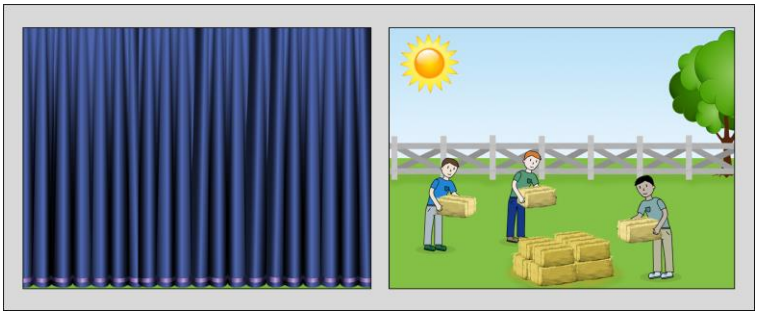

Figure 1. Distributive picture

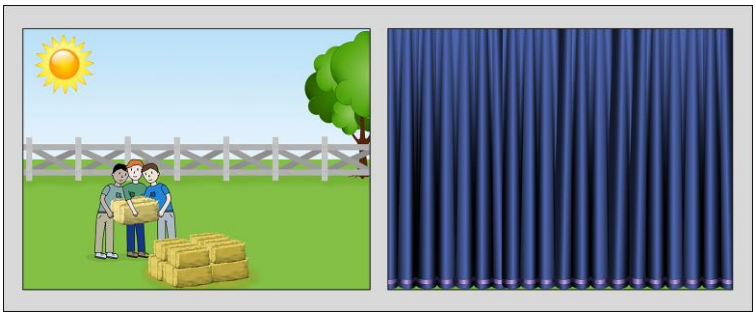

Figure 2. Collective picture

3.2 Design and Procedure

Participants received sentence-picture pairs in a covered-box design, with a visible picture and a 'covered' picture behind curtains. The experiment had a 2x2 design with the factors PICTURE and SENTENCE. The visible pictures depicted either distributive (Figure 1) or collective (Figure 2) situations and the sentences were all of the form Subject-Verb-Indefinite Object, beginning with the definite plural *de* 'the' (3) or the quantifier *elke* 'each' (4). We used four transitive verbs: *dragen, tillen, vasthouden* and *trekken* (in English: 'carry', 'lift', 'hold' and 'pull'). All direct objects were indefinite noun phrases starting with the article *een* 'a'. Note that the Dutch indefinite article *een* 'a' is pronounced differently from the Dutch numeral *één* 'one'. Therefore, it is unlikely that participants will confuse the indefinite article with the numeral.

(3) ***De*** *jongens droegen een hooibaal.*
The boys carried a hay bale.
(4) ***Elke*** *jongen droeg een hooibaal.*
Each boy carried a hay bale.

The sentences were presented in a narrative context that went as follows. Three boys and three girls helped a farmer by doing chores on his farm. For the farmer it was necessary to know who performed which chore and he therefore asked one of the girls to take pictures of every chore and describe the pictures to him afterwards. The participants were asked to check whether or not the description that the girl gave matched the picture completely, or whether the description only partially matched the picture. If participants felt it was not a complete match, they were instructed to indicate this by choosing the 'covered' picture. In this way we encouraged them to consider alternatives without explicitly providing those alternatives, as in a picture selection task.

In previous covered-box experiments (e.g. Huang et al., 2013; Schwarz et al., 2016), a black square was used to depict the covered picture. In our experiment we chose to depict the covered picture using curtains. We did this to clearly present the scenario of the choice between two pictures, one visible and one covered. A black square, contrary to curtains, does not suggest that something is hidden behind it, making it harder to explain the task to children.

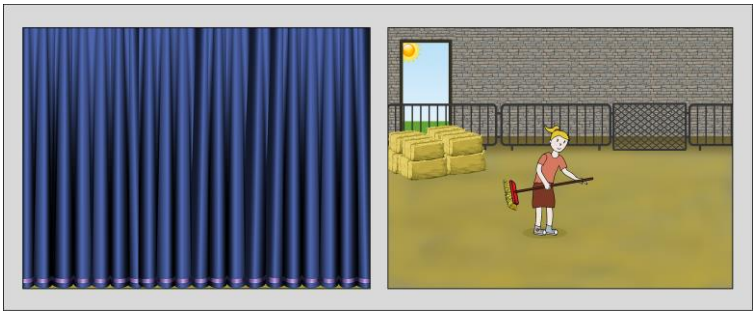

Figure 3. Pictures of practice item with curtains closed

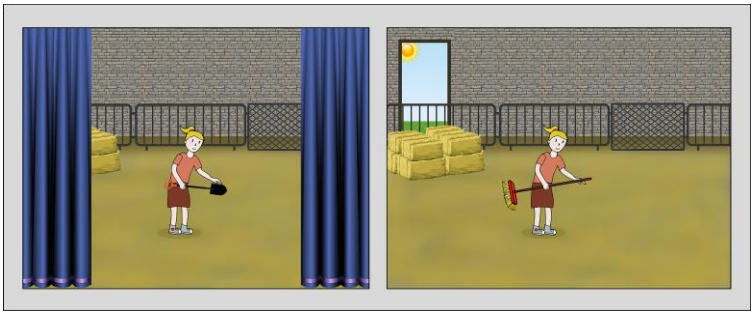

Figure 4. Pictures of practice item with curtains opened

Because a covered-box task is a fairly abstract task (especially for children), participants received four practice items to familiarize them with the method. These four practice trials involved straightforward correct or incorrect situations. After the participant made a choice, the curtains opened (regardless of the participant's choice) to show the picture behind them. This revealed whether the covered picture indeed was a better match. Figures 3 and 4 show the pictures for the

practice sentence ‘The girl carried a shovel’. Figure 4 visualizes that it was best to choose the covered picture. None of the children had problems with the practice trials, indicating that they all grasped the idea behind the task. The curtains only opened during the practice trials and remained closed during the actual task.

The experimental design resulted in four conditions: Quant-Distributive, Quant-Collective, Def-Distributive and Def-Collective. In total, participants received 56 items: 4 practice items, 24 target items (6 per condition), 20 filler items and 6 control items. Items were counterbalanced, distributed over four lists and presented randomly. The side on which the visible and covered pictures appeared was randomized within participants. Participants’ responses were labeled as either a choice for the covered picture or a choice for the visible picture. The results are analyzed by looking at the proportion of times participants chose the visible picture, indicating that they accepted this picture as the meaning of the sentence.

3.3 Results

Descriptive statistics are presented in Figure 5. Both participant groups show an overwhelming preference for the visible picture for conditions Quant-Distributive and Def-Collective, following previous findings (de Koster et al., 2017; Padilla-Reyes, 2018; Pagliarini et al., 2012). Furthermore, the adults show a preference for the visible picture in 38% of cases for condition Def-Distributive and in 19% for condition Quant-Collective. The results of the children show a different pattern. The children prefer the visible picture in 65% of cases for condition Def-Distributive, which is much higher than the 38% of the adults. This contrasts with condition Quant-Collective, where the preference rates are more similar for the children (11%) and the adults (19%).

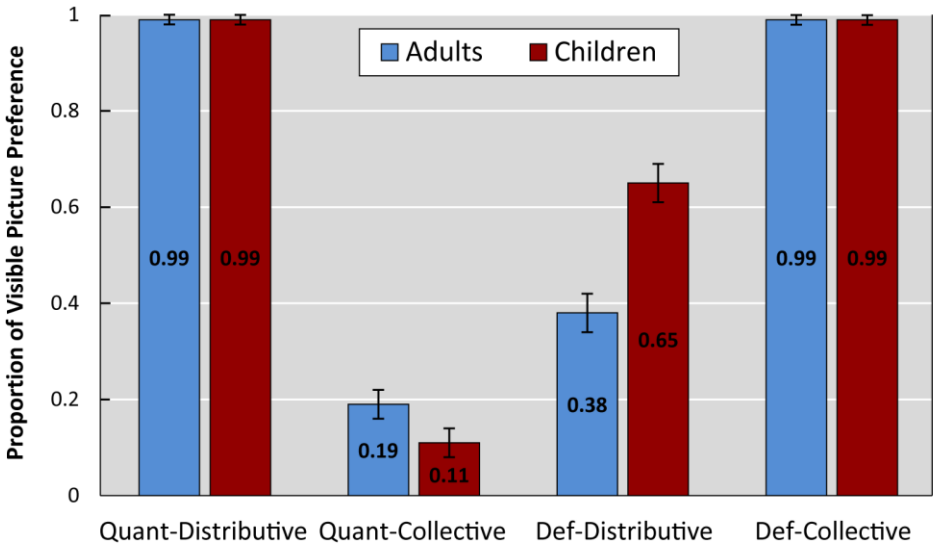

Figure 5. Proportion of visible picture preferences per condition for the adult and child participants. Error bars show standard errors.

The data were analyzed using generalized mixed-effects logistic modelling. We determined the final model based on model comparisons (cf. Baayen et al., 2008) using the Akaike Information Criterion (AIC). An AIC decrease of more than two indicates that the goodness of fit of the model improves significantly (Akaike, 1974). The final model contained the factors CONDITION(Quant-

Distributive, Quant-Collective, Def-Distributive, Def-Collective), AGE-GROUP(adults, children) and ITEM-ORDER (treatment coded: reference levels were 'Def-Distributive' for CONDITION and 'adults' for AGE-GROUP). The factor ITEM-ORDER was a continuous factor taking into account participants' responses over time. The dependent variable was the response (0 for a choice for the covered picture, 1 for a choice for the visible picture). The maximal random-effects structure licensed by the data included a random intercept for participants and by-participant random slopes for CONDITION.

We found a main effect of CONDITION indicating that condition Def-Distributive differed significantly from all three conditions for both age-groups, with a higher preference for the visible picture compared to condition Quant-Collective (adults: β=-2.769; z=-2.352; p<.05, children: β=-6.317; z=-4.770; p<.001), and a lower preference for the visible picture compared to conditions Def-Collective (adults: β=6.586; z=4.346; p<.001, children: β=3.622; z=2.166; p<.05) and Quant-Distributive (adults: β=5.357; z=5.257; p<.001, children: β=3.984; z=2.765; p<.01). A relevelled model indicated no difference between conditions Def-Distributive and Def-Collective (adults: β=-0.611; z= -0.496; p=.620, children: β=0.958; z=0.738; p=.460).

Furthermore, we found an interaction between CONDITION and AGE-GROUP (β=-4.032; z=-2.898; p<.01), which showed that there is only a significant difference between age groups in condition Def-Distributive (β=2.215; z=2.271; p<.05) and not in the other three conditions (Quant-Collective: β=-1.600; z=-1.356; p=.175, Quant-Distributive: β=0.760; z=0.607; p=.544, Def-Collective: β=-1.007; z=-0.799; p=.425). Children and adults showed similar preference rates for conditions Quant-Collective, Quant-Distributive and Def-Collective, contrasting with condition Def-Distributive. Lastly, we also found a main effect of ITEM-ORDER (β=-0.03; z=-3.824; p<.001), indicating lower preference rates for the visible picture over time.

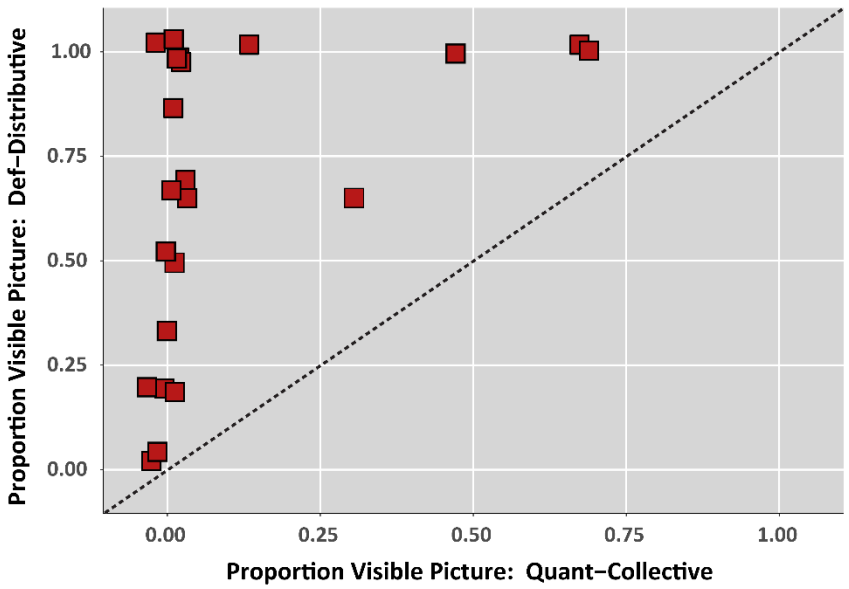

Figure 6. Correlation plot showing the proportions of visible picture preferences for conditions Quant-Collective and Def-Distributive for the children. Squares present individual children. A proportion of 1 indicates a full preference for the visible picture, a proportion of 0 indicates a full preference for the covered picture.

We also looked at the hypothesis that children have to learn *elke* 'each' before they will show a collective preference for unmarked sentences. A correlation plot is presented in Figure 6. This figure shows the proportions of visible-picture preferences for conditions Quant-Collective and Def-Distributive per child. It reveals that the children who prefer the covered picture for condition Def-Distributive (calculating the implicature) are children who also prefer the covered picture for *elke* in a collective context. Data points in the bottom right-hand corner would have been evidence against our hypothesis and would show that a full understanding of the distributive character of *elke* is not a prerequisite for being able to calculate the predicted implicature.
Crucially, there are no data points in the bottom right-hand corner. This shows that children only start choosing the covered picture for condition Def-Distributive once they understand the distributive character of *elke*.

3.4 Discussion
The results from conditions Def-Collective and Quant-Distributive were fully accepted by both participant groups and followed previous findings (Padilla-Reyes, 2018; de Koster et al., 2017; Pagliarini et al., 2012), so we will not discuss them further.

The results for condition Quant-Collective also matched our predictions. The adults showed a preference for the covered picture, thus implicitly rejecting the visible collective picture. The children showed a similar pattern, showing a full understanding of the distributive marker *elke* 'each'.

Condition Def-Distributive is the condition where we predicted an implicature to be calculated, leading to a preference for the covered picture. Here adults preferred the visible picture in 38% of cases. This might seem high, but these results are lower than the acceptance rates found in previous studies using a different methodology, e.g., 52% in de Koster et al. (2017) and 50% in Pagliarini et al. (2012). This also fits well with the observation that implicatures can show varying acceptance rates (cf., van Tiel et al., 2016). The children showed a higher preference for the visible picture than the adults (65% vs. 38%), suggesting that they have not reached the adult preference even at age 11. However, they seem to be moving towards this preference, since they do not consistently prefer the visible picture. In addition, our results are consistent with the prediction of the implicature account that children have to learn the distributive character of *elke* 'each' before they start showing a preference for the collective interpretation. There were no children who preferred the covered picture for condition Def-Distributive and at the same time preferred the visible picture for condition Quant-Collective.

To conclude, the results of the more sensitive covered-box task suggest that an implicature underlies the collective preference for unmarked plurals. Participants showed that they considered alternatives by choosing the covered picture. Crucially, we did not provide these alternatives, unlike in a picture selection task.

One potential problem, however, may be that participants were exposed to alternatives indirectly via the other items, which included both collective and distributive situations. Being exposed to these explicit alternatives, making the collective interpretation more salient, might have facilitated implicature calculation. Because participants will only gradually be exposed to alternatives as they

progress through the experiment, they may respond differently to earlier than to later items. The significant main effect of item order showing lower preferences for the visible picture over time might indicate such an exposure effect. To further examine this effect we performed a second experiment, in which we removed the exposure to alternatives by testing participants on only one picture type. By doing this we can distinguish a possible exposure effect from a more general effect of task experience.

4. Experiment 2

In Experiment 2 we examined whether implicature calculation in Experiment 1 was facilitated by an exposure to both picture types. If we show participants only one picture type (only distributive or only collective pictures), do we see similar results or will the collective preference found in Experiment 1 disappear? Because our aim was to investigate a possible alternative explanation for the findings in Experiment 1, we only carried out Experiment 2 with adults.

4.1 Participants

48 Dutch adults participated in the experiment. They performed the experiment online through a website, without the experimenter being present, and received €5 for their participation. We tested participants in a between-subjects design: 24 participants (mean age 25, range 19-53) received items with distributive pictures only, and 24 participants (mean age 23, range 18-34) received items with collective pictures only.

4.2 Design and Procedure

The covered-box experiment of Experiment 2 had a 2x2 design with the factors PICTURE and SENTENCE, similar to Experiment 1. Experiment 2 however, tested participants in a between-subjects design, resulting from the between-subjects factor PICTURE. This entailed that participants in Experiment 2 received only one picture type; they saw either distributive or collective pictures. Sentence type remained a within-subjects factor, resulting in two conditions per participant group (either Def-Distributive and Quant-Distributive or Def-Collective and Quant-Collective). Participants again received 56 items in total: 4 practice items, 24 target items, 20 filler items and 6 control items. The procedure of Experiment 2 was the same as the procedure of Experiment 1 (see Section 3.2).

4.3 Results

Descriptive statistics are presented in Figure 7. In accordance with our predictions, conditions Quant-Distributive and Def-Collective show ceiling performance and reveal no differences between experiments, so we will not discuss them any further. Condition Quant-Collective shows a slightly lower preference for the visible picture in Experiment 2 compared to Experiment 1 (10% vs. 19%). This contrasts with condition Def-Distributive, which shows a higher preference for the visible picture in Experiment 2 (38% vs. 47%). The descriptive results of condition Def-Distributive seem to indicate that participants were more likely to prefer the covered picture in Experiment 1 than in Experiment 2. It is important to find out whether this difference is due to participants' exposure to both distributive and collective pictures in Experiment 1, making the collective interpretation more salient, or a more general effect of task experience such as familiarity with the task. We therefore need to compare the results of Experiment 2 to the results

of Experiment 1, since the exposure effect was ruled out in Experiment 2 due to its between-subjects design.[1]

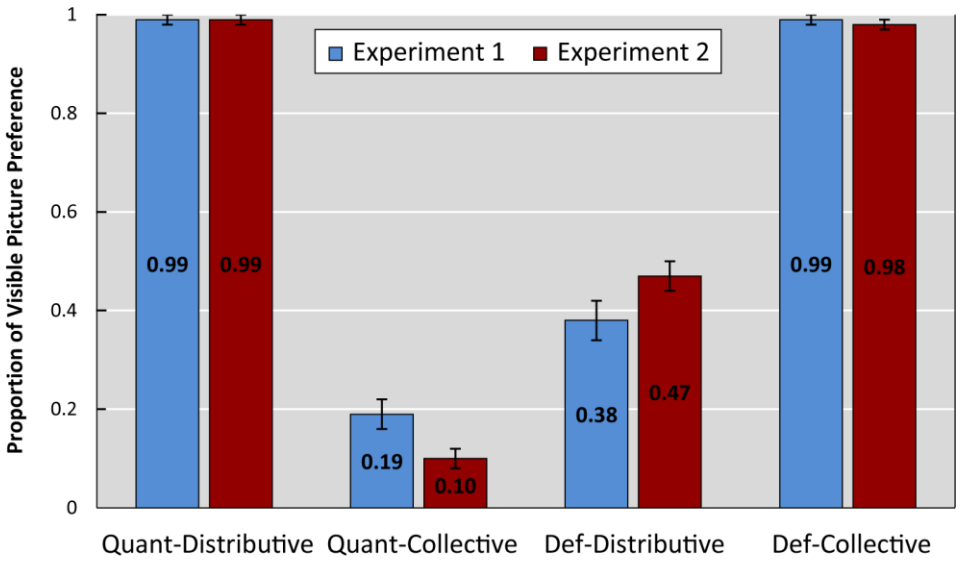

Figure 7. Proportion of visible picture preferences per condition for the adult participants in Experiment 1 and Experiment 2. Error bars show standard errors.

We analyzed the data of conditions Quant-Collective and Def-Distributive from Experiments 1 and 2 using generalized mixed-effects logistic modelling. Conditions Quant-Distributive and Def-Collective were left out, since they revealed an overwhelming preference for the visible picture and did not differ between experiments.

Based on model comparisons (cf. Baayen et al., 2008), we determined whether the following fixed-effect factors improved the goodness of fit of the model using the Akaike Information Criterion (AIC): CONDITION(Quant-Collective, Def-Distributive), EXPERIMENT(1, 2) and ITEM-ORDER. The dependent variable was the response (0 for a choice for the covered picture, 1 for a choice for the visible picture).

The final model included a significant three-way interaction between the factors CONDITION, EXPERIMENT and ITEM-ORDER ($\chi^2(3)=16.805$, $p<0.001$; AIC difference:-10.8) (treatment coded: reference levels were 'Def-Distributive' for CONDITION and 'experiment 2' for EXPERIMENT). The maximal random-effects structure licensed by the data included a random intercept for participants and by-participant random slopes for CONDITION. We will unravel the three-way interaction in the next paragraphs. The interaction effect is plotted in Figure 8.

[1] A model analyzing only the data of Experiment 2 showed similar results as presented in Section 3.3 for the adult data of Experiment 1, with condition Def-Distributive differing significantly from all other conditions and no significant difference between conditions Def-Collective and Quant-Distributive.

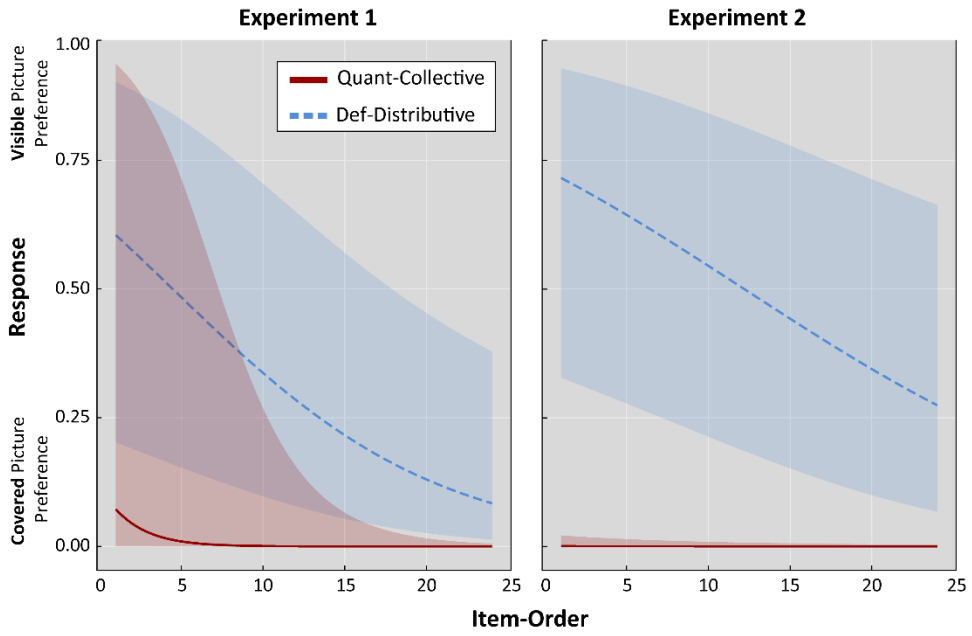

Figure 8. The three-way interaction between the factors Condition, Experiment and Item-Order. The shaded areas present the 95% confidence interval.

First, we found a significant difference between conditions Quant-Collective and Def-Distributive in Experiment 2 (β=-8.421; z=-4.074; p<.001). Condition Def-Distributive had a significantly higher preference rate for the visible picture compared to condition Quant-Collective (47% vs. 10%), similar to the findings in Experiment 1 (38% vs. 19%, see Section 3.3).

Second, we did not find a significant difference in visible picture preference for condition Def-Distributive between experiments (β=-0.458; z=-0.362; p=.717). Crucially, participants responded to condition Def-Distributive to a similar degree in both experiments. This finding serves as evidence against an exposure effect in Experiment 1, even though the descriptive statistics initially seemed to indicate such an effect (38% vs. 47%). Condition Quant-Collective, on the other hand, did differ between experiments (β=5.836; z=2.372; p<.05). Experiment 2 showed a significantly lower preference rate for the visible picture than Experiment 1 (10% vs. 19%).

Third, we also found a significant effect of ITEM-ORDER (β=-0.083; z=-2.584; p<.01) in condition Def-Distributive in Experiment 2. This shows that participants had a lower preference for the visual picture in condition Def-Distributive over time in Experiment 2 (taking into account the reference levels), which can also be seen in Figure 8. Crucially, however, we found no difference for condition Def-Distributive in the effect of ITEM-ORDER between experiments (β=-0.040; z=-0.734; p=.463). ITEM-ORDER had a similar effect on condition Def-Distributive in both Experiments 1 and 2, indicating that participants had a similar lower preference for the visual picture over time in Experiment 1. We did find a difference in effect of ITEM-ORDER between experiments for condition Quant-Collective (β=-0.379; z=-2.260; p<.05). Participants had a lower preference for the visible picture over time for condition Quant-Collective in Experiment 1, but no such drop in preference over time is seen in Experiment 2 (see Figure 8).

5. General Discussion

The results of Experiment 1 suggested that an implicature underlies the collective preference for distributively unmarked plurals and showed that participants considered alternatives while interpreting these unmarked plurals. Because the implicature calculation might have been facilitated by participants' exposure to explicit alternatives via the other conditions in the experiment, we ran a follow-up study with a between-subjects design. The implicature condition (Def-Distributive) did not differ between experiments. The fact that the collective preference remained in Experiment 2, strongly suggests that participants internally generate and compare alternative situations themselves when interpreting unmarked plurals.

We found an effect of item order on condition Def-Distributive, indicating that participants' responses changed over time. However, because this effect was present in both experiments, this general effect is most likely due to task experience, and is not caused by exposure to alternatives. This is supported by the results of condition Quant-Collective from Experiment 1. There, a similar effect of item order was found, although performance on condition Quant-Collective depends on the lexical semantics of *elke* 'each' and does not require a comparison with alternatives. The unexpected stronger preference for the visible picture in this condition that we found in Experiment 1 compared to Experiment 2 (19% vs. 10%) might be due to the within-subjects design of Experiment 1. This may have led to a larger effect of task experience because the participants in Experiment 1 saw different picture types, making the experiment less straightforward.
In short, the results of both experiments seem to point to an implicature underlying the adult collective preference, providing support for Dotlačil's (2010) account. Participants were able to compare the provided picture to internally generated alternatives in both experiments, even when these alternatives were not provided. In addition, the different interpretations of adults and children in Experiment 1 follow from the implicature account.

A few other studies have tried to explain the collective preference and the difference in interpretations between adults and children (e.g., Frazier et al., 1999; Musolino, 2009). However, none of them are able to account for both the adult preference and children's interpretations at the same time, contrary to the implicature account. Musolino (2009), for example, argues that children's non-adult-like interpretations relate to another well-known phenomenon in quantification, namely children's so-called spreading errors. This account however does not explain how adults arrive at a collective preference, focusing only on explaining children's over-acceptance of distributive interpretations. Frazier et al. (1999), on the other hand, argue that the adult collective preference is caused by the fact that the distributive interpretation is semantically more complex. A distributive interpretation introduces multiple events, whereas a collective interpretation only introduces one event. This difference in semantic complexity results in a default collective interpretation. Frazier et al.'s (1999) explanation, however, does not match the interpretations of children. Children are overly accepting distributive interpretations, and this is unexpected if the distributive interpretation is indeed more semantically complex, as Frazier et al. (1999) argue. The implicature account instead offers a pragmatic explanation.

One issue, however, remains unexplained regarding the acquisition path of the proposed implicature. Dotlačil's account predicts that the implicature leading to the collective preference for unmarked plurals is driven by the semantically distributive *elke* 'each'. Children first need to understand that *elke* 'each' is a distributive marker, before they can use its absence to develop a

collective preference. Children show a full understanding of *elke* 'each' around age 8, but only show a collective preference for unmarked plurals around age 11. So how can we explain this age gap? After children have acquired the lexical semantics of the quantifier, they may still need experience with calculating the implicature for plural definites to be able to reliably use the meaning of the quantifier as an alternative in their interpretation of plural definites. Thus, the implicature account merely predicts that the lexical semantics of the quantifier *elke* 'each' is a prerequisite for implicature calculation. Factors related to performance may explain why the implicature for plural definites takes some more time to develop.

6. Conclusions

In sentence interpretation the division of labor between semantic and pragmatic information and processing is often hard to determine. In this study we investigated how the interpretation of plural expressions can be explained by both semantic and pragmatic processes. We presented new evidence supporting the hypothesis that the preference for a collective interpretation with plural definites is the results of an implicature. This study not only yields support for the implicature account, thus explaining adults' and children's interpretations of plurals, but also demonstrates how the covered-box task can be used to test implicature calculations.

Funding Information

The first author is supported by a grant from the Netherlands Organization for Scientific Research (322-75-010).

Acknowledgements

We thank all participants, especially the children and teachers from CBS de Regenboog. We would also like to thank Jakub Dotlačil and Dorothée Hoppe for their helpful comments and feedback.

References

Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6): 716-723.

Baayen, R. Harald, Davidson, Douglas J., and Bates, Douglas M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4): 390-412.

Brooks, Patricia J., and Braine, Martin D.S. 1996. What do children know about the universal quantifiers all and each? *Cognition.* 60(3): 235-268.

Champollion, Lucas. 2017. Parts of a whole: Distributivity as a bridge between aspect and measurement. Vol. 66, Oxford.

Dotlačil, Jakub. 2010. Anaphora and Distributivity. A study of same, different, reciprocals and others. Doctoral Dissertation, Utrecht University.

Frazier, Lyn, Patch, Jeremy M. and Rayner, Keith. 1999. Taking on semantic commitments, II: collective versus distributive readings, *Cognition*, 70(1): 87-104.

Grice, Herbert. P. 1975. Logic and conversation. In P. Cole and J. J. Morgan, eds. *Syntax and semantics*, volume 3: speech acts. New York: Academic Press.

Huang, Yi Ting, Spelke, Elizabeth & Snedeker, Jesse. 2013. What exactly do numbers mean? *Language Learning and Development.* 9(2): 105-129.

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

de Koster, Anna M.B., Dotlačil, Jakub & Spenader, Jennifer K. 2017. Children's understanding of distributivity and adjectives of comparison. In *Proceedings of the 41st Annual Boston University Conference on Language Development*. 1: 373-386.

Landman, Fred. 2012. Events and plurality: The Jerusalem lectures. Springer Science & Business Media.

Musolino, Julien. 2009. The logical syntax of number words: Theory, acquisition and processing. *Cognition*. 111(1): 24-45.

Padilla-Reyes, Ramon E. 2018. Connections among scales, plurality, and intensionality in Spanish. Doctoral dissertation, The Ohio State University.

Pagliarini, Elana, Fiorin, Gaetano & Dotlačil, Jakub. 2012. The acquisition of distributivity in pluralities. In *Proceedings of the 36st Annual Boston University Conference on Language Development*. 2: 387-399.

Schwarz, Florian, Bill, Cory & Romoli, Jacopo. 2016. Reluctant acceptance of the literal truth: Eye tracking in the covered box paradigm. In *Proceedings of Sinn und Bedeutung*. 20: 61-78.

Syrett, Kristen, and Musolino, Julien. 2013. Collectivity, distributivity, and the interpretation of plural numerical expressions in child and adult language. *Language Acquisition*. 20(4): 259-291.

Van Tiel, Bob, Van Miltenburg, Emiel, Zevakhina, Natalia & Geurts, Bart (2016). Scalar diversity. *Journal of Semantics.* 33(1): 137-175.