Differential Patterns of Children's Knowledge of Quantifier Meaning Revealed Under Different Tasks

Einat Shetreet [1,2*†] and Rama Novogrodsky [3†]

[1] Department of Linguistics, Tel Aviv University, Tel Aviv, Israel, [2] Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel, [3] Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel

This study examines children's comprehension of quantifiers in Hebrew using several tasks. We focused on a linguistic ambiguity related to universal quantifiers that express a distinction between collectivity and distributivity: *all* can be assigned with both a collective reading and a distributive reading ("a flower for all fairies" can be interpreted as an event with one flower or an event with multiple flowers), whereas *each* has a distributive reading only ("a flower for each fairy" is an event with multiple flowers). Unlike English, Hebrew has a single universal quantifier and thus, it expresses the collectivity/distributivity distinction using two morphosyntactic forms: one form (*kol+ definite plural noun*) is equivalent to *all* and has the two readings, and the other form (*kol+ indefinite singular noun*) is equivalent to *each* and has only one reading. We examined how Hebrew-speaking preschoolers (4–6 years) understand sentences in the two forms, and how they resolve the ambiguity of the ambiguous form, while focusing on the type and presence of contrast in three preference tasks. Experiment 1 used a conventional picture-matching task where the collective and distributive meanings were contrasted using two pictures (meaning contrast); Experiment 2 used a sentence-matching task where the two morphosyntactic forms were contrasted using two sentences (linguistic contrast); and Experiment 3 used a novel drawing task including instructions in one form (no contrast). In all tasks, adults showed a consistent response pattern, matching the ambiguous form (equivalent to *all*) to the collective reading and the distributive form (equivalent to *each*) to the distributive reading. Children, on the other hand, were affected by the task, showing adult-like performance pattern in the picture-matching task, but not in the other tasks. This suggests that children can distinguish between the two morphosyntactic forms, but they do not fully attain adults' preference pattern. The differences between the tasks can be attributed to the salience of the contrast, task experience, or working memory. The results highlight the need for a careful selection of language tasks, both in basic research and in clinical assessment.

Keywords: language learning, ambiguity, semantics, universal quantifiers, language acquisition

INTRODUCTION

What speakers know about language is sometimes not fully expressed in how they behave in certain linguistic tasks. This is especially evident when testing children, who often show a competence-performance distinction (e.g., Crain and Fodor, 1989). Indeed, the performance of children (and adults) depends on their linguistic competence, but often other factors such as the

materials, the instructions, or the background information, influence performance. It is therefore crucial to understand the effect of different language tasks in order to properly assess children's linguistic knowledge, even when testing one specific aspect (e.g., Crain and Thornton, 2000; Schmitt and Miller, 2010).

The current study investigated children's knowledge of sentence-level semantics with emphasis on quantifiers, which are words that indicates quantities (e.g., "all" or "some"). We focused on universal quantifiers, such as "all" and "each," which express the idea that a certain property applies to the entire group. At the center of our study is the collective/distributive distinction (as detailed below), this involves linguistic ambiguity which has been shown to have a consistent resolution in adults (e.g., Brooks and Braine, 1996; Brooks et al., 2001). In this study, we used three different comprehension (preference) tasks to assess children's knowledge of the collective/distributive distinction and examine the effects of task on their performance. The findings could contribute to our understanding of children's comprehension of quantifiers, and possibly inform clinical assessment.

Quantifiers and the Collective/Distributive Distinction

The meaning of quantifiers, such as "all" or "some," depends on the context in which they are presented, and they are often ambiguous (e.g., Brooks and Braine, 1996; Chierchia and McConnel-Ginet, 2000; Lidz and Musolino, 2002). As such, they impose a challenge to children acquiring language. A large, cross-linguistic study showed that at the age of 5;5, children have not yet mastered quantifier meaning (Katsos et al., 2016).

This study focuses on one phenomenon, the collective/distributive distinction, and its relation to the comprehension of universal quantifiers (e.g., "all"/"each"). This subtle distinction concerns different types of groups—those that refer to a unified or *collective* group of individuals or events, and those that refer to individuated or *distributive* events. An example of a collective event would be one in which a teacher takes one picture of her 20 students, whereas a distributive event would be one where she takes 20 separate pictures, one picture per student. Languages provide the means to make this distinction, but not without ambiguity. English has one way to describe collective events: using the universal quantifier "all" (Example 1a), but two ways to describe distributive events (Examples 1a and 1b): using either "all" or "each."

(1) a. Sarah photographed all the students.
    b. Sarah photographed each student.

These examples illustrate that "all" is ambiguous between a collective and a distributive reading and can apply to both types of events, but "each" is obligatorily distributive and cannot refer to collective events (Vendler, 1967; Ioup, 1975). This observation has been verified experimentally. For example, English-speaking adults accepted both readings when presented with "all," but only the distributive meaning when presented with "each" (Brooks and Braine, 1996).

Critically, the ambiguity is not part of the real world (i.e., events are either collective or distributive), but is present in

the linguistic representation (as shown in Example 2a). This is reflected in the cross-linguistic variation of the resolution of the ambiguous form. Whereas, both English and Dutch have equivalent universal quantifiers ("all," "each," and "every" in English; "alle," "elk," and "ieder" in Dutch; van der Ziel, 2008), research suggests that their preference pattern differs. When tested in a picture-matching task, English-speaking adults preferred the collective interpretation of the quantifier "all" (Brooks and Braine, 1996; similar results are seen in Portuguese and Mandarin speakers in Brooks et al., 2001). However, Dutch-speaking adults, who were also tested in a picture-matching task, alternated between the two readings of the Dutch quantifier *alle* with no clear preference (Rouweler and Hollebrandse, 2015; van Koert et al., submitted). Thus, it appears that languages differ with regard to the mapping between these quantifiers and their preferred meanings.

Languages also differ regarding the representation of the collective/distributive distinction. Specifically, Hebrew, the language tested in this study, uses a morphosyntactic distinction rather than a lexical one. This is because Hebrew has only one universal quantifier, *kol*. To convey the collective/distributive distinction, this quantifier is coupled with a definite noun in the plural form (kol + definite N_{pl}) or with an indefinite noun in the singular form (kol + indefinite N_{sing}) (Gil, 1995; Francez and Goldring, 2012), as in Examples (2a) and (2b):

(2) a. Sarah cilma et **kol** ha-talmidim $_{HEB}$
       "Sarah photographed all the students"
    b. Sarah cilma **kol** talmid $_{HEB}$
       "Sarah photographed each student"

Note that in these examples, we translated *kol* to two different words in English. This translation is based on several linguistic observations, demonstrating similarities between the two morphosyntactic forms in Hebrew and the lexical items in English. The form in (2a) is ambiguous, and like "all," it has both collective and distributive readings, but the form in (2b), like "each," has a distributive reading only (For English, Vendler, 1967; Ioup, 1975; Beghelli and Stowell, 1997; Tunstall, 1998; For Hebrew, Gil, 1995; Francez and Goldring, 2012). Therefore, Example (2a) can be interpreted such that the end-result is one picture of 20 students (a collective event), or 20 separate pictures with one student per picture (a distributive event). Example (2b) can only mean the latter. Another observation is that the structure in (2a) permits a cumulative interpretation, like "all," whereas the structure in (2b), like "each," does not (For English, Brisson, 2003; For Hebrew, Korat, 2016). Finally, the structure in (2a) can appear with collective predicates (e.g., "all the soldiers *gathered*"), like "all," but the structure in (2b), like "each," cannot (For English, Champollion, 2014; For Hebrew, Gil, 1995; Korat, 2016).

The Collective/Distributive Distinction in Development

The capacity to recognize distributivity in the real world is present early in typical development (age 4; Avrutin and Thornton, 1994). The ability to distinguish between the two linguistic forms in English occurs early on; there is some evidence

that even 2- and 3-year-olds can distinguish between "all" and "each" in an act-out task (Ferenz and Prasada, 2002), and that 5-year-olds do so in a forced-choice, picture-matching task (Brooks and Braine, 1996; Drozd, 2001; for Portuguese and Mandarin, Brooks et al., 2001). Dutch-speaking children start making this distinction around the age of 6. Importantly, they pattern with Dutch-speaking adults, who, as stated above, do not have a strong tendency in terms of assigning a collective meaning to the ambiguous quantifier (Rouweler and Hollebrandse, 2015; van Koert et al., submitted). This cross-linguistic variation in development has some important implications relevant to the current study. First, it is important to characterize the varying behaviors of speakers of different languages, to enhance our understanding of universality of language development (Schaeffer, 2003). To the best of our knowledge, a language such as Hebrew, where the collective/distributive distinction of the universal quantifier is marked morphosyntactically, has not yet been tested. In addition, children's performance and knowledge should be referenced to those of adults speaking the same language, as different patterns present in different languages.

Additionally, it has been shown that English-speaking 5-year-olds, who demonstrated adult-like patterns in a forced-choice task, differed from English-speaking adults in a judgment task. In the judgment task, children accepted the distributive lexical item ("each") as appropriate for both distributive events and collective events, whereas adults accepted "each" only for the former (Brooks and Braine, 1996). This finding shows that various tasks can reveal different aspects of children's linguistic knowledge, which is what motivated the inclusion of different preference tasks in the current study. Furthermore, this finding indicates that the acquisition of this distinction is not complete at the age of 5, which prompts the decision to include children ages 4–6 years.

The Current Study

Based on the observations reported above, we asked whether the capacity to make the collective/distributive distinction is observed in Hebrew-speaking preschoolers, as they rely on morphosyntactic cues alone. Because this distinction is more subtle in Hebrew than in other tested languages (that have both lexical and morphosyntactic cues), Hebrew-speaking children might struggle to learn it. However, in the age range tested in the current study (4–6 years), children have already acquired the morphosyntactic markers required for the distinction (i.e., the definite marker and the plural markers). Experimental work shows that children master definiteness toward the end of the third year (Meir et al., 2017). Based on standardized tests in Hebrew, the plural markers are expected to be successfully produced and comprehended toward the age of 3 (*Goralnik Screening Test for Hebrew*, Goralnik, 1995; *PLS-4 Hebrew version*, Zimmerman et al., 2010). We, therefore, predict that Hebrew-speaking preschoolers (ages 4–6) will be able to make the collective/distributive distinction, and demonstrate differing behaviors with the ambiguous form and the distributive form.

Our second main goal was to examine the effects of task on children's preferences for the collective vs. the distributive

readings of the ambiguous form. We asked what the preference pattern in Hebrew is, and whether it is stable across different tasks. To this end, we conducted three experiments, each using a different preference task. The first experiment tested whether there is a preference for one reading of the ambiguous form in Hebrew-speaking adults and children. To determine the preference pattern in Hebrew, we used a conventional picture-matching task, like that used in previous studies of this phenomenon. One possibility is that Hebrew speakers have a strong preference for the collective reading with the ambiguous form, as shown in English. Alternatively, they may not show a clear preference for one of the readings, as shown in Dutch. Under both possibilities, we predict that children will have adult-like preference patterns, possibly to a lesser degree, as was seen in previous studies that tested children speaking languages that have both lexical and morphosyntactic cues.

The second and third experiments used two other tasks, manipulating the type and presence of contrast. Our second experiment tested whether similar preference patterns will be observed across tasks that contrast different aspects of the distinction. The ambiguity of the collective/distributive distinction is in the linguistic domain rather than in the real world. Therefore, we attempted to conduct a more direct comparison in the linguistic domain by contrasting the two morphosyntactic forms in the same trial (asking to match one picture to one of two sentences). Previous studies of quantifier comprehension have shown children's performance is facilitated in a sentence-matching task (Chierchia et al., 2001). They tested quantifier meaning that was derived through scalar implicatures. Scalar implicatures occur when the logical meaning of certain scalar expressions (e.g., *some* which logically means "some and possibly all") is pragmatically-interpreted as excluding the strongest expression of the same scale (e.g., *some* is enriched to mean "some but not all"). It has been suggested that the preference for collective reading with the ambiguous quantifier also involves an implicature (for a detailed explanation, see Dotlačil, 2010; Pagliarini et al., 2012). Aravind and de Villiers (2015) argued that the sentence-matching task might be easy for children, because they were explicitly given the possible descriptions of a picture, and therefore were not required to reason about the best possible alternative descriptions. In short, presenting the ambiguous quantifier (or ambiguous form in Hebrew) would indicate to the comprehender that the more informative, distributive quantifier (which has only one reading) does not hold. Therefore, we predict that when the two linguistic alternatives (e.g., sentences with "all"/the ambiguous form and sentences with "each"/the distributive form) are present, as in the sentence-matching task (Experiment 2), children should show a strong preference for the collective reading over the distributive reading with the ambiguous form.

Our third experiment concerned the role of contrast in children's performance (following Aravind and de Villiers, 2015). In a forced-choice task, a contrast is present (either in pictorial or linguistic form) and may facilitate children's choice of the correct (adult-like) response (e.g., Katsos and Bishop, 2011, Experiment

3). To eliminate this potential influence, we designed a novel drawing task where we asked participants to add drawings to the picture based on a single sentence. This is also a preference task, as participants could draw only one interpretation of the sentence. However, this task did not include a direct meaning or a direct linguistic form contrast within any single trial. Under the assumption that children's comprehension of the collective/distributive meaning is not completed around the age of 5 (Brooks and Braine, 1996), and given that this task requires a selection between two (or possibly more) implicit alternatives, we predicted that children would show a weaker preference pattern in comparison to Experiments 1 and 2 that included a contrast.

EXPERIMENT 1: PICTURE-MATCHING TASK

To establish the preference pattern of Hebrew-speaking adults and children, we first conducted a conventional sentence-picture task of the type that has been used in studying the collective/distributive distinction (e.g., Brooks and Braine, 1996). In this task, the experimenter presents a sentence together with a set of pictures (or objects or stories) which include the target interpretation and one or more distractors. The participant's task is to choose which stimuli best represent or matches the sentence. This setting can facilitate children's performance as the interpretation of the sentence is presented to them, often together with a contrastive, yet related interpretation (e.g., Katsos and Bishop, 2011, Experiment 3).

In the case of the collective/distributive distinction, the set of pictures included a collective event and a distributive event. Thus, the two possible readings of the ambiguous form were available from the input and could have been contrasted by the participants while interpreting the target sentence. We hypothesized that the explicit contrast between the alternatives, as presented in the pictures, would facilitate children's performance. While we expected that the distributive form would prompt distributive choices, we could not predict whether Hebrew speakers would

tend to choose collective events following the ambiguous form (like English speakers, e.g., Brooks and Braine, 1996), or whether they would not show a clear preference (like Dutch speakers, e.g., Rouweler and Hollebrandse, 2015; van Koert et al., submitted).

Method
Participants
Fifteen children (7 females; mean age = 4;11; range = 4;0–5;10) and 14 adults (9 females; mean age = 25; range = 22–30) participated in the experiment. All were native Hebrew speakers. All the children attended kindergartens and schools in major cities in Israel. No other information was collected. Written informed consent was obtained from the adult participants and from the parents of the child participants.

Materials
Twenty picture pairs were used. In each pair, one picture presented a distributive event with 4–6 characters each holding/using a single object, and the other presented a matching collective event with the same characters holding the same type of a single object (**Figure 1**). The pictures were presented vertically, counterbalancing the location of the collective and distributive events. Sentences in either the ambiguous form or in the distributive form were read to the participants (**Figure 1**). Each condition (i.e., sentence form: distributive/collective) appeared 10 times, such that each participant saw each picture pair only once.

Procedure
Participants were tested individually in their schools, universities or homes. The sentence was presented by the experimenter and the participant was asked to choose between the two pictures. Two familiarization trials at the beginning of the session were used to introduce the task and test children's knowledge of the singular/plural form (without the quantifier). Halfway through the experiment, the children were given a break, if needed. Each half had the same number of collective and distributive items. No feedback was given during the task.
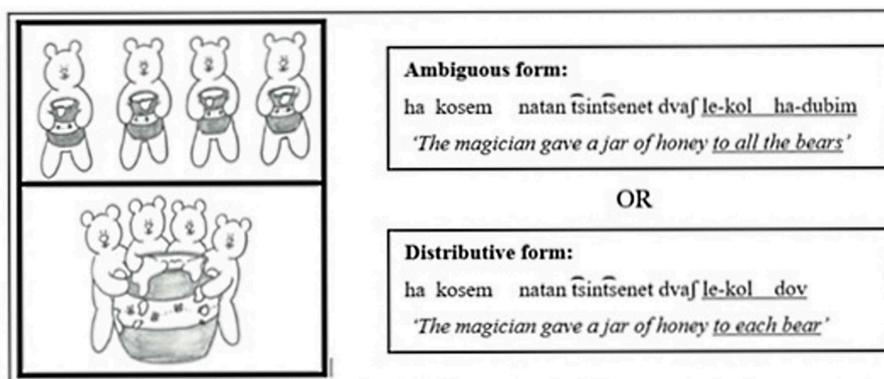
FIGURE 1 | An example of one picture pair in the picture-matching task. Participants heard either the ambiguous form **or** the distributive form in each trial.

Results and Discussion

Adults showed a ceiling effect in both conditions: they chose collective events for sentences in the ambiguous form (mean of collective responses: 98%) and distributive events for sentences in the distributive form (mean of distributive responses: 99%). Thus, the preference pattern in Hebrew is similar to English, rather than Dutch. As predicted, children showed a similar pattern to that observed in adults, with slightly weaker preference (mean of collective responses: 75%; mean of distributive responses: 83%) (**Figure 4**) below.

We tested chance-performance in the children's group using a binomial test. As a group, the children reliably performed above chance in both conditions ($p < 0.001$). Additionally, most individual children performed above chance in both conditions (**Table 1**).

Further statistical analysis is given in the section named "comparisons between the three experiments," where we compared the performance across tasks and sentence types. We did not conduct any within-experiment statistics to avoid multiple comparisons. Based on the chance-level analysis, it appears that in a picture-matching task, Hebrew speakers, like English speakers, have a strong preference to assign the collective reading to the ambiguous form. This occurs in both adults and children (to a lesser extent). This finding further suggests that children can make the distinction between the two forms in Hebrew.

TABLE 1 | Numbers of children with above-chance/chance-level performance in the picture-matching task.

	Distributive form	Ambiguous form
Above-chance performance (*choosing the collective event*)	0	9
Chance performance	3	6
Above-chance performance (*choosing the distributive event*)	12	0

EXPERIMENT 2: SENTENCE-MATCHING TASK

Based on the binomial test, we found that both adults and children show a clear tendency to interpret the ambiguous form as collective in a context where the two interpretations are contrasted (further analysis is given in the section named "comparisons between the three experiments"). In Experiment 2, we used a variation of the forced-choice task, where we reversed the form of presentation, such that one picture was presented with two sentences, one for each morphosyntactic form (following Chierchia et al., 2001). In this task, only one interpretation was depicted, and the contrast was given in the linguistic input.

In this experiment, we expected that trials depicting collective events would prompt ambiguous form choices because this is the only form that can be used with such events. It is to be determined whether adults and children prefer the distributive form to describe depicted distributive events or whether participants will alternate between the two forms, as both are valid.

Method

Participants

Eighteen children (10 females; mean age = 5;0; range = 4;1–6;0) and 20 adults (11 females; mean age = 31; range = 22–40), who did not participate in the previous experiment, were included in this experiment. We initially tested 22 children: three were excluded because they consistently chose the same puppet/side (see procedure below), suggesting their performance was biased by a factor that was irrelevant to the task; another participant was excluded due to not understanding of the task. All participants were native Hebrew speakers. All the children attended kindergartens and schools in major cities in Israel. No other information was collected. Written informed consent was obtained from the adult participants and the parents of the children.

Materials

We used a variation of a sentence-picture-matching task where 2 possible sentences were presented with a single picture. Using the same pictures from Experiment 1, we divided the 20 picture
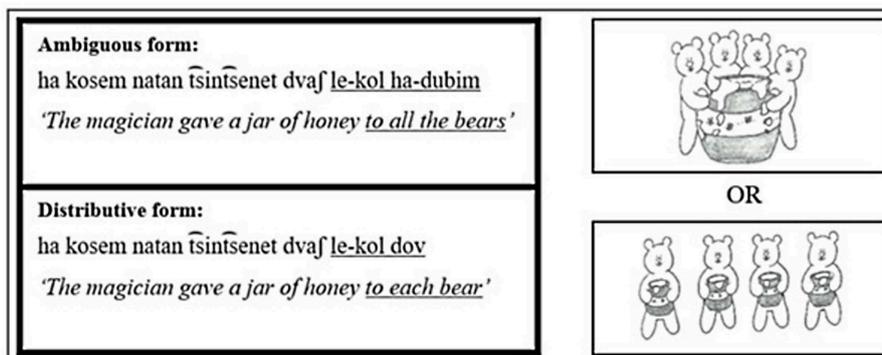
FIGURE 2 | An example of one sentence pair in the sentence-matching task. Participants saw either a picture with the collective event **or** with the distributive event.

pairs into 2 different lists, such that each participant saw only one picture of each pair (**Figure 2**). Thus, each condition (i.e., picture type) was presented 10 times. Each picture appeared with two sentences, one in the ambiguous form and the other in the distributive form.

Procedure

Participants were tested individually in their schools, universities, or homes. The participants were introduced to two puppets and were told that each puppet would give a description of the picture, one puppet used the ambiguous form and the other used the same sentence with the distributive form (as shown in **Figure 2**). Participants were instructed to choose which puppet said it better. While presenting the picture, the experimenter read the two descriptions, one after the other, in a neutral tone with no emphasis. Two familiarization trials without quantifiers were included at the beginning. Half way through the experiment, the children were given a break, if needed. Each half had the same number of collective and distributive items. No feedback was given during the task.

Results and Discussion

Adults chose sentences in the ambiguous form when presented with collective events (mean of ambiguous-form responses: 100%), and sentences in the distributive form when presented with distributive events (mean of distributive-form responses: 95%). Similar to the picture-matching task, adults showed a strong preference for the collective reading of the ambiguous form in the sentence-matching task. In contrast to adults, children's rates of adult-like responses in this task were relatively low, with 64% ambiguous-form responses for collective events and 42% distributive responses for distributive events (**Figure 4**) below.

As in Experiment 1, chance-performance was tested using a binomial test. As a group, the children reliably performed above-chance only with collective events ($p < 0.001$), choosing the ambiguous form more often than the distributive form. However, a few individual children (5/18) had above-chance performance in this condition (**Table 2**). In the distributive condition, no child consistently chose the distributive form, but 4 children reliably chose the ambiguous form.

Further statistical analysis is reported in the section named "comparisons between the three experiments," as part of the comparison between the three tasks, to avoid multiple comparisons. The results from the binomial test, as well as the higher exclusion rates, suggest that children had more difficulties

TABLE 2 | Numbers of children with above-chance/ chance-level performance in the sentence-matching task.

	Distributive event	Collective event
Above-chance performance (*choosing the ambiguous form*)	4	5
Chance performance	14	13
Above-chance performance (*choosing the distributive form*)	0	0

with this task. Despite the similarities between the sentence-pictures task and the picture-sentences task, the number of children that performed at chance-level was higher in the sentence-matching task than in the picture-matching task in both conditions (14 vs. 3 in the distributive condition, 13 vs. 6 in the collective condition). Thus, the binomial test results suggest that children had greater difficulty with this task, indicating inconsistent performance across tasks (this was verified by a logistic regression reported in the section named "comparisons between the three experiments").

EXPERIMENT 3: DRAWING

The third task was a novel drawing task, in which participants were presented with a single picture that included a series of similar characters. They were instructed with one of the two morphosyntactic forms to add drawings to the pictures (see example 3 below). No contrast (linguistic or meaning) was provided in each trial. However, the contrast was present throughout the experiment.

As in Experiment 1, we expected distributive responses following drawing instructions in the distributive form but could not predict whether instructions in the ambiguous form would elicit the distributive or the collective interpretation. In this task, the participants had to construct the meaning of the sentences by themselves, with no cues to possible interpretation from the input. Given this, and the lack of contrast between the two potential meanings, it was possible that we would observe a different preference pattern than that seen in Experiment 1.

Method
Participants

Fifteen children (12 females; mean age = 5;7 years; range = 4;5–6;5 years) and 15 adults (10 females; mean age = 26.5 years; range = 18–37 years), who did not participate in the previous experiments, participated in this experiment. All participants were native Hebrew speakers. All the children attended kindergartens and schools in major cities in Israel. No other information was collected. Written informed consent was obtained from the adult participants and the parents of the child participants.

Materials

In this novel drawing task, participants were given a picture with 4–6 objects of the same type (e.g., **Figure 3A**) and were asked to draw certain items on the picture. The instructions were given in the ambiguous (Example 3a) or the distributive form (Example 3b). An additional form of instruction was included but will not be discussed in this paper. Each condition appeared five times with different pictures, such that each participant saw each picture only once, and the combination of pictures and instructions was counterbalanced.

(3.) a. ambiguous form: Cajer perax le-kol ha-dvorim
Draw a flower to-all the-bees
b. distributive form: Cajer perax le-kol dvora
Draw a flower to-each bee.
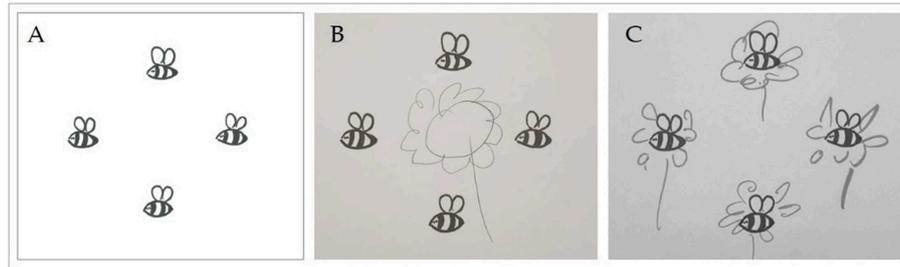
FIGURE 3 | Examples for **(A)** one drawing item in the drawing task; **(B)** collective response **(C)** distributive response.
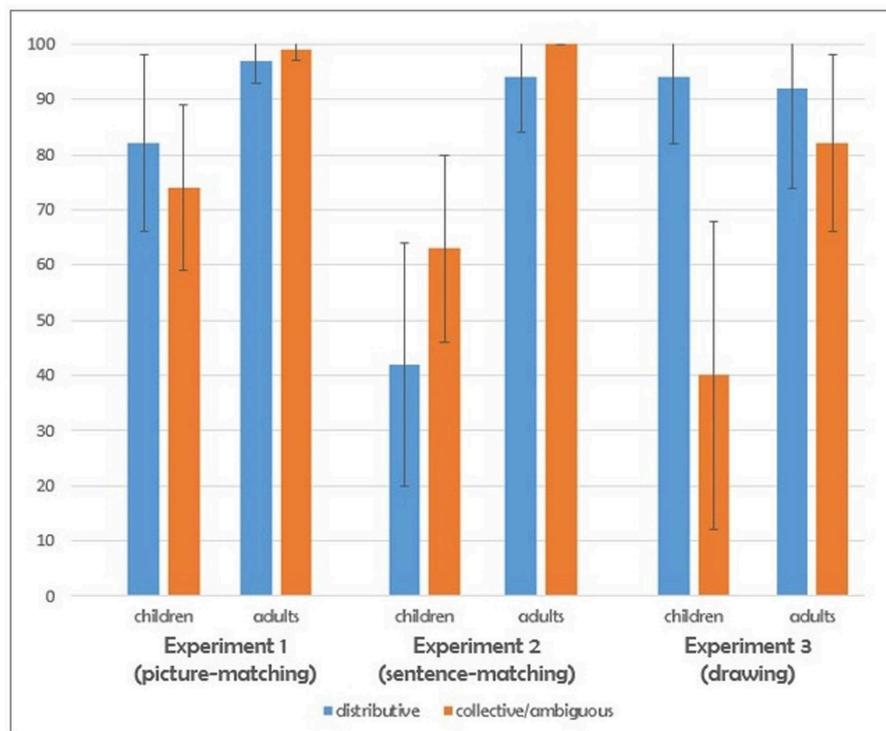
FIGURE 4 | Mean correct fit responses (matching distributive events with the distributive form and collective events with the ambiguous form) for all three experiments.

Procedure

Participants were tested individually in their schools, universities or homes. Each participant was presented with a picture and was then given the drawing instructions. If a child expressed difficulties with drawing, the experimenter suggested that the child draw a simple line. No other feedback was given during the task.

Results and Discussion

In this task, participants were free to produce several response types. Yet, almost all of the responses (446/450) were collective, consisting of a single item drawn for all the depicted objects (**Figure 3B**); or distributive (**Figure 3C**), consisting of multiple items drawn, one per depicted object (four unclear responses were excluded from the analysis).

Adults gave collective responses following instructions in the ambiguous form (mean of collective responses: 83%), and distributive responses following instructions in the distributive form (mean of distributive responses: 92%)[1]. Similar to adults, children's rates of distributive responses following distributive instructions were high (mean of distributive responses: 94%). However, they gave many distributive responses following ambiguous instructions (mean of collective responses: 40%) (**Figure 4**) below.

Here too, we tested chance-performance in children using a binomial test. As a group, children reliably performed above-chance only in the distributive condition ($p < 0.001$). In fact,

---

[1] A small number of errors is also expected in adults across all tasks, due to inattentiveness, carelessness, etc.

TABLE 3 | Numbers of children with above-chance/chance-level performance in the drawing task.

	Distributive form	Ambiguous form
Above-chance performance (*drawing collective events*)	0	0
Chance performance	3	14
Above-chance performance (*drawing distributive events*)	12	1

most individual children had above-chance performance in this condition (**Table 3**). In the ambiguous condition, one child gave above-chance distributive responses (rather than collective responses).

Together with the other tasks, we provide additional statistical analysis in the next section. The binomial test revealed that in this task, there was a difference between the two forms, as more children produced distributive responses following the distributive form than collective responses following the ambiguous form. This was the only task where this difference between the conditions was observed.

COMPARISONS BETWEEN THE THREE EXPERIMENTS

Chance-level performance, as reported in the previous sections, indicates different performance patterns between the different tasks. Here, we further report a direct statistical comparison of these differences. To compare performance across tasks, we examined the fit between the ambiguous form and collective events and the fit between the distributive form and distributive events. A "correct fit" based on the adults' responses means, for example, choosing the picture of the collective event when hearing the ambiguous form in Experiment 1, the ambiguous form when seeing a picture of a collective event in Experiment 2, and drawing a collective event when hearing the ambiguous form in Experiment 3. This fit was the dependent variable in our analysis.

We conducted a logistic mixed-effects regression analysis (using the lmerTest package in R), where we modeled the rates of correct fit. The reference level for Collectivity/Distributivity was the distributive condition and for task, the picture-matching task. We tested for main effects of collectivity/distributivity and task (picture-matching, sentence-matching, and drawing) and the interaction between them. A maximal random effect structure, including interaction slopes for both fixed effects (i.e., collectivity/distributivity and task) and random effects (i.e., participants and items) failed to converge. Therefore, we removed the interaction slopes from the random effects. There were significant effects of both collectivity/distributivity and task and a significant interaction between them (**Table 4**)[2].

---

[2]There were significant age differences between the groups, due to 3 older children in Experiment 3. Removing these children from the analysis resulted in no age differences and the same result pattern reported.

Follow-up pairwise comparisons showed that children performed equally well in the collective/ambiguous and the distributive conditions in the picture-matching task, with high rates of correct fit in both conditions. In Experiments 2 and 3, differences were observed in the two conditions, in opposite directions (**Table 5**). In the sentence-matching task, children had higher rates of correct fit for the collective/ambiguous condition than for the distributive, despite generally low rates of adult-like responses. In contrast, in the drawing task, children had significantly higher rates of correct fit in the distributive condition than in the collective condition.

We further looked at the correct fit of each condition separately, across tasks. Children had higher rates of correct distributive responses in the drawing task than in the other two tasks, but they also had many more such responses in the picture-matching task than in the sentence-matching task. In contrast, their rates of correct collective/ambiguous responses were lower in the drawing task than in the other two tasks, with stronger effects when compared with the picture-matching task.

DISCUSSION

The main goal of the current study was to assess children's comprehension of the collective/distributive distinction as expressed by universal quantifiers. We first tested whether Hebrew-speaking preschoolers were able to make this distinction. This is because Hebrew represents a special case of the distinction, as it provides only morphosyntactic and not lexical cues. Hebrew has a single universal quantifier as opposed to several quantifiers in English (e.g., "all" or "each"). We used a conventional picture-matching task (Experiment 1), similar to previous studies testing the collective/distributive distinction in other languages (e.g., Brooks and Braine, 1996; Brooks et al., 2001; van Koert et al., submitted). We showed that children had high rates of "correct fit" in this task (i.e., matching the ambiguous form with collective events and the distributive form with distributive events). This finding indicates that children distinguished the ambiguous form and the distributive form and attached a different interpretation to each. This conclusion is also justified by the differential performance in the drawing task (Experiment 3), where children exhibited different response patterns for the two forms. In other words, children were able to make the collective/distributive distinction in a language that uses morphosyntactic cues alone.

Experiment 1 was also aimed to determine whether one of the two readings of the ambiguous form (kol + definite N_{plu}) is preferred in Hebrew. Previous research has shown that the preference pattern is different across languages: English-speakers, both adults and children, prefer the collective reading, whereas Dutch speakers do not show a preference for either reading. Hebrew-speakers, adults and children, preferred to choose collective events when presented with the ambiguous form. This indicates that in Hebrew, the ambiguous form is typically interpreted as collective. That is, the ambiguous form can have a strong preference for the collective reading both when it is marked lexically (e.g., in English; e.g., Brooks and Braine,

TABLE 4 | Summary of mixed-effects model for the rates of correct fit in all three experiments.

	β	SE	z-value	p
Collectivity/Distributivity (Distributive): Collective/Ambiguous	−1.0074	0.2515	−4.006	<0.00001
Task (Picture-matching-Exp1): sentence-matching-Exp2	−1.2904	0.2693	−4.792	<0.00001
Drawing-Exp3	−0.1337	0.3825	−0.350	0.72670
Collectivity/Distributivity (Distributive): Collective/Ambiguous X Task (Picture-matching-Exp1): sentence-matching-Exp2	1.4320	0.4824	2.969	0.003
Collectivity/Distributivity (Distributive): Collective/Ambiguous X Task (picture-matching-Exp1): sentence-matching-Exp2	−2.9773	0.6880	−4.327	<0.00001

TABLE 5 | Summary of pairwise comparisons.

Within-Experiment effects: Distributive vs. collective/ambiguous	β	SE	z-value	P
Exp1: Picture-matching	0.4923	0.3668	1.342	0.7518
Exp2: Sentence-matching	−0.9397	0.3129	−3.003	0.03
Exp3: Drawing	3.4696	0.5828	5.954	<0.001
DISTRIBUTIVE FIT				
Exp1 vs. Exp2	2.0064	0.3695	5.429	<0.001
Exp1 vs. Exp3	−1.3550	0.6108	−2.218	0.2192
Exp2 vs. Exp3	−3.3614	0.5940	−5.659	<0.001
COLLECTIVE FIT				
Exp1 vs. Exp2	0.5744	0.3533	1.626	0.5687
Exp1 vs. Exp3	1.6223	0.3953	4.104	<0.001
Exp2 vs. Exp3	1.0479	0.3734	2.807	0.0530

1996; Brooks et al., 2001, but see Dutch for a different pattern, Rouweler and Hollebrandse, 2015; van Koert et al., submitted) and morphosyntactically (in Hebrew).

Once we established a strong preference pattern for the collective reading of the ambiguous form in Hebrew, we asked whether this preference is stable across tasks in children, by using two additional preference tasks. Overall, we showed that adults, who were tested as a reference, gave consistent responses across tasks (with high rates of "correct fit"), and were minimally influenced by the task. In contrast, preschoolers were significantly affected by the changes in the experimental settings: They had high rates of "correct fit" in the conventional picture-matching task, which required them to match one of two pictures to one sentence (Experiment 1). However, their rates of "correct fit" were lower in the other two tasks. In the drawing task, which required adding drawings to pictures based on instructions (Experiment 3), children had high rates of "correct fit" when the instructions were given in the distributive form, but chance-level performance when the instructions were given in the ambiguous form. Finally, in the sentence-matching task (Experiment 2), where the two linguistic forms were available to be matched with one picture, most children performed at chance-level in both conditions. Taken together, the findings from the three tasks indicate that children's performance was affected by the task.

These findings also speak to our questions regarding the role of contrast in the comprehension of the collective/distributive distinction. We hypothesized that children's performance would be affected by the presence of contrast, such that it would be less adult-like when no contrast was present in the task (i.e., the drawing task) than when it was present (i.e., picture- and

sentence-matching tasks). We proposed that facilitation of adult-like performance might occur, as the alternatives are explicitly present and can be contrasted by the participants in the case of picture- and sentence-matching tasks (see also Katsos and Bishop, 2011). In the drawing task, participants might struggle to construct the correct interpretation by themselves. However, our findings did not completely align with this hypothesis; children performed worst on the sentence-matching task, which included a contrast in the linguistic form. This clearly suggests that the mere presence of contrast does not provide enough support for children performance. Further discussion on the differences between the tasks is given in the section named "differences performance across tasks".

Thus, our results show that the knowledge of the collective/distributive distinction in Hebrew-speaking preschoolers is incomplete by age six. Even so, under certain experimental settings, such as a picture-matching task, their knowledge appears almost adult-like. This finding indicates that different linguistic tasks that tap into the comprehension of the same linguistic feature may be affected by different aspects of children's knowledge. It further emphasizes the need to use multiple assessments to fully understand children's linguistic knowledge.

Differential Performance Across Tasks

In this section we will discuss the differential performance in the three different tasks. We first consider the differences between the picture-matching task and drawing task, as these differences are in line with our hypothesis that the presence of contrast provides some support for children. We then turn to discuss

the differences between the picture-matching task the sentence-matching task, which are not in line with our hypothesis.

The Picture-Matching Task vs. the Drawing Task

As stated above, children had higher rates of correct fit/adult-like performance in the picture-matching task. It is possible that this is due to the contrast of meaning present in the picture-matching task. The pictures in the picture-matching task depicted two similar but distinct situations. The contrast between them provided a clear cue for a meaning distinction, which could have facilitated children's performance (e.g., through the principle of contrast, Clark, 1987). No such cue was available for the drawing task. In other words, the experimental settings in the picture-matching task, and not in the drawing task, may have highlighted the two possible meanings and signaled to the children that both meanings should be considered in the response. This could have helped them in attaching each form with its (preferred) meaning.

Previous studies that tested children's lexical-semantic knowledge also showed an advantage for a picture-matching task over a variation of an act-out task (specifically, coloring book task, Pinto and Zuckerman, 2018, Experiment 2). Interestingly, the opposite pattern was observed when testing syntactic knowledge, showing disadvantages for a picture-matching task (Frizelle et al., 2017b; Pinto and Zuckerman, 2018; Experiment 1). Frizelle et al. (2017b) argued that a picture-matching task might test factors other than linguistic abilities, such as memory load required to represent the distractors (i.e., the other pictures). One explanation for the better performance in the picture-matching task in the current study is that we used one distractor rather than two or three as used in the other studies, limiting the memory load needed to perform the task. However, we find it more reasonable that the tested linguistic features, syntactic vs. semantic knowledge, are affected differently by task demands. This is supported by the dissociation between syntactic and vocabulary knowledge found in Pinto and Zukerman's study (2018). Nevertheless, our findings join those from other studies advocating for exploring task sensitivity.

The Picture-Matching Task vs. the Sentence-Matching Task

Children had low rates of correct fit in the sentence-matching task, in comparison to the high rates of correct fit in the picture-matching task. This is reflected not only in the low rates of assigning the collective reading to the ambiguous form, but also in erroneously assigning the collective reading to the distributive form, which does not allow such reading. Here, we consider three possible explanations for this finding: a subtle linguistic distinction, task experience and working memory.

We hypothesized that children will show stronger preference for the collective reading of the ambiguous form in the sentence-matching task based on previous results from the same task testing another aspect of quantifier knowledge- that of scalar implicatures (Chierchia et al., 2001), where children showed adult-like performance. We made this prediction based on the suggestion that ambiguity resolution in the case of the ambiguous form involves an implicature (Dotlačil, 2010; Pagliarini et al.,

2012). Thus, our finding might suggest that this is not the case for Hebrew. However, it is important to note that in the case of scalar implicature studies, the linguistic contrast included two different lexical items (i.e., "all" and "some"), whereas in the present study, the linguistic contrast included the same lexical items in different morphosyntactic forms. It is possible that a morphosyntactic contrast is subtler than a lexical contrast. If so, it might have been harder for children to detect the contrast, as well as parse and maintain the differences in meaning between two sentences when relying on morphosyntactic cues rather than lexical-semantic cues. This explanation awaits future research and can be explored by testing the collectivity/distributivity distinction in a sentence-matching task in languages where the distinction is made lexically (e.g., English which uses "all" and "each").

Other explanations are available for the differences in performance between the sentence-matching task and the picture-matching (and drawing) task. Such explanations are related to task experience or task demands, rather than to linguistic factors. It is possible that children have less experience with a sentence-matching task, as they do not engage with this activity in the real world. During language comprehension, children are typically required to map structure to meaning (with or without visual cues), similar to what they are assumed to do in the drawing task. Children also encounter situations in which they are requested to choose one object from an array of objects, similar to what they are assumed to do in the picture-matching task (for example, in natural discourse where multiple objects of different colors and shapes are in view, children often hear sentences such as "give me the red circle"). Children's experience with this task is reflected in many standardized language tests: Language comprehension is often tested by asking the child to point to one of several pictures or objects (e.g., Zimmerman et al., 2010), which indicates that children are indeed capable of performing this task. In contrast, we assume that situations in which children are asked to apply two (or more) linguistic structures to the same situation, and to choose between them, are less frequent in children's language learning and experience. It is therefore possible that children did not succeed in the sentence-matching task because they lack the experience with this linguistic task. This finding is supported by the high exclusion rate of participants in this task and by the high error rates of assigning the collective reading to the distributive form. If so, the other tasks used in the current study may be better indicators of children's knowledge. This explanation can be explored by testing bilingual children, who have extensive experience with applying multiple forms of the different languages they speak to the same real-world situations (e.g., Degani et al., 2019).

The last explanation concerns task demands and suggests that differences between the tasks are governed by working memory. In the sentence-matching task, children had to remember two sentences and keep them active in their working memory before making a decision. In the picture-matching task and the drawing task, children had to hold only one sentence in their working memory. Furthermore, in the picture-matching task, the two contrasting pictures were still available while making the decision. By contrast, in the sentence-matching task the auditory signal disappeared over time. In fact, several studies

tied working memory with language development, showing correlations between better performance in working memory tasks and linguistic abilities (e.g., Gathercole and Baddeley, 1990; Alloway et al., 2005). This explanation can be tested in future studies by a direct examination of the relationship between measures of working memory and performance in a sentence-matching task of quantifiers.

We cannot determine which of the above explanations applies in our study. While we think that the explanations involving task experience and task demands are less plausible because English-speaking children succeeded in a sentence-matching task (Chierchia et al., 2001), future research along the line suggested in this section can help uncover the factors that influenced children's performance in the current study.

Implications for Clinical Assessment

In the current study, different comprehension tasks revealed significantly different levels of knowledge of the same linguistic feature, which is critical for clinical assessments (de Villiers and de Villiers, 2010). Several previous studies have focused on the differences between comprehension and production of a specific aspect of language (e.g., Brandt-Kobele and Höhle, 2010; Frizelle et al., 2017a), an approach which influenced the structure of language assessment tools. For example, the Preschool Language Scale test (Zimmerman et al., 2010) is structured to test language comprehension, language production (i.e., expressive language) and a combined language age. Our results stress that having different tasks within the same domain (i.e., comprehension) is critical and can provide a more sensitive measurement that uncovers different levels of knowledge.

Clearly, using one task only cannot capture all levels of a child's knowledge, and more importantly, it can lead to over- or under-estimation of their knowledge. When performing an assessment for a child in need of intervention, it is imperative to select a highly sensitive measurement. In some cases, the assessment task itself could provide scaffolding for the child's performance (as in our picture-matching task). This could lead to high scores in the assessment task that do not reflect the child's knowledge as it is manifested in daily communication. The differential performance across the tasks in the current study suggests that when developing assessment tools, one should consider not only which linguistic feature to test, but also which task to use. Let us take for example the tasks explored here: If the aim of the

language test is to assess whether the child has acquired a certain meaning or interpretation (in our case, that of the universal quantifier in its different forms), a picture-matching task would be appropriate. However, if the aim of the assessment is to test the depth of the child's knowledge, the drawing task, which is more ecologically-valid (i.e., closer to how this knowledge is processed in real life) might be more sensitive (see also the coloring book task, Pinto and Zuckerman, 2018). Our findings suggest that the type of task selected in research and in the clinic should be based on the research, assessment, or intervention goals (de Villiers and de Villiers, 2010).

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

All procedures performed in this study involving human participants were in accordance with the ethical standards of the University of Haifa and the Tel Aviv University ethics committees.

AUTHOR CONTRIBUTIONS

ES and RN contributed equally to this study, including concept, experimental design, data collection supervision, data analysis, and writing.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Alloway, T. P., Gathercole, S. E., Adams, A. M., Willis, C., Eaglen, R., and Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *Br. J. Dev. Psychol.* 23, 417–426. doi: 10.1348/026151005X26804

Aravind, A., and de Villiers, J. G. (2015). "Implicit alternatives insufficient for children's implicatures with some," in *Proceedings of the 39th Annual Boston University Conference on Language Development (BUCLD)*. Somerville, MA: Cascadilla.

Avrutin, S., and Thornton, R. (1994). Distributivity and binding in child grammar. *Ling. Inquiry* 25, 165–171.

Beghelli, F., and Stowell, T. (1997). "Distributivity and negation: the syntax of each and every," in *Ways of Scope Taking*, ed Anna Szabolcsi (Netherlands: Springer), 71–107. doi: 10.1007/978-94-011-5814-5_3

Brandt-Kobele, O. C., and Höhle, B. (2010). What asymmetries within comprehension reveal about asymmetries between comprehension and production: the case of verb inflection in language acquisition. *Lingua* 120, 1910–1925. doi: 10.1016/j.lingua.2010.02.008

Brisson, C. (2003). Plurals, all, and the non-uniformity of collective predication. *Linguist. Philos.* 26, 129–184. doi: 10.1023/A:1022771705575

Brooks, P. J., Braine, M. D., Jia, G., and Dias, M. D. G. (2001). "Early representations for all, each, and their counterparts in Mandarin Chinese and Portuguese," in *Language Acquisition and Conceptual Development*, ed S. C. Levinson (Cambridge: Cambridge University Press), 316–339. doi: 10.1017/CBO9780511620669.013

Brooks, P. J., and Braine, M. D. S. (1996). What do children know about the universal quantifiers all and each? *Cognition* 60, 235–268. doi: 10.1016/0010-0277(96)00712-3

Champollion, L. (2014). "Each vs. all: distributivity, cumulativity, and two types of collectivity," in *Handouts for a Talk at the 4th Cornell Workshop in Linguistics and Philosophy* (Ithaca, NY).

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., and Meroni, L. (2001). "The acquisition of disjunction: evidence for a grammatical view of scalar implicatures," in *Proceedings of the 25th Boston University Conference on Language Development,* eds A. H. J. Do, L. Domingues, and A. Johansen (Somerville, MA: Cascadilla Press), 157–168.

Chierchia, G., and McConnel-Ginet, S. (2000). *Meaning and Grammar: An Introduction Into Semantics, 2nd edn*. Cambridge, MA: MIT Press.

Clark, E. V. (1987). "The principle of contrast: a constraint on language acquisition," in *Mechanisms of Language Acquisition,* ed B. MacWhinney (Hillsdale, NJ: Lawrence Erlbaum Associates), 1–33.

Crain, S., and Fodor, J. D. (1989). "Competence and performance in child language," in *Language and Cognition: A Developmental Perspective,* ed E. Dromi (Norwood, NJ: Ablex), 141–171.

Crain, S., and Thornton, R. (2000). *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, MA: MIT Press.

de Villiers, P. A., and de Villiers, J. G. (2010). Assessment of language acquisition. *WIREs Cogn. Sci.* 1, 230–244. doi: 10.1002/wcs.30

Degani, T., Krizer, V., and Novogrodsky, R. (2019). The joint effects of bilingualism, DLD and item-frequency on children's lexical retrieval performance. *Int. J. Lang. Commun. Disord.* 54, 485–498. doi: 10.1111/1460-6984.12454

Dotlačil, J. (2010). *Anaphora and Distributivity*. A study of same, different, reciprocals and others (Doctoral Dissertation). Utrecht Institute for Linguistics OTS, LOT Series.

Drozd, K. F. (2001). "Children's weak interpretations of universally quantified questions," in *Language Acquisition and Conceptual Development,* eds M. Bowerman and S. C. Levinson (Cambridge: Cambridge University Press), 340–345. doi: 10.1017/CBO9780511620669.014

Ferenz, K. S., and Prasada, S. (2002). Singular or plural? Children's knowledge of the factors that determine the appropriate form of count nouns. *J. Child Lang.* 29, 49–70. doi: 10.1017/S0305000901004937

Francez, I., and Goldring, K. (2012). "Quantifiers in modern hebrew," in *Handbook of Quantifiers in Natural Languages,* eds E. Keean and D. Paperno (Berlin: Springer), 347–398. doi: 10.1007/978-94-007-2681-9_7

Frizelle, P., O'Neill, C., and Bishop, D. V. (2017a). Assessing understanding of relative clauses: a comparison of multiple-choice comprehension versus sentence repetition. *J. Child Lang.* 44, 1435–1457. doi: 10.1017/S0305000916000635

Frizelle, P., Thompson, P. A., Duta, M., and Bishop, D. (2017b). Assessing children's understanding of complex syntax: a comparison of two methods. *Lang. Learn.* 69, 255–291. doi: 10.31219/osf.io/wujhy

Gathercole, S. E., and Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: is there a causal connection? *J. Mem. Lang.* 29, 336–360. doi: 10.1016/0749-596X(90)90004-J

Gil, D. (1995). Universal quantifiers and distributivity. *Quantif. Nat. Lang.* 2, 321–362. doi: 10.1007/978-94-017-2817-1_11

Goralnik, E. (1995). *Goralnik Diagnostic Test.* Even Yehuda: Matan.

Ioup, G. (1975). Some universals for quantifier scope. *Syntax Semant.* 4, 37–58.

Katsos, N., and Bishop, D. V. M. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition* 120, 67–81. doi: 10.1016/j.cognition.2011.02.015

Katsos, N., Cummins, C., Ezeizabarrena, M. J., Gavarró, A., Kraljević, J. K., Hrzica, G., et al. (2016). Cross-linguistic patterns in the acquisition of quantifiers. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9244–9249. doi: 10.1073/pnas.1601341113

Korat, O. (2016). "*Singular quantified names*" in Proceedings of the Semantics of African, Asian and Austronesian Languages, Vol. 2, 36–51.

Lidz, J., and Musolino, J. (2002). Children's command of quantification. *Cognition* 84, 113–154. doi: 10.1016/S0010-0277(02)00013-6

Meir, N., Walters, J., and Armon-Lotem, S. (2017). Bi-directional cross-linguistic influence in bilingual Russian-Hebrew speaking children. *Ling. Appr. Biling.* 7, 514–553. doi: 10.1075/lab.15007.mei

Pagliarini, E., Fiorin, G., and Dotlačil, J. (2012). "The acquisition of distributivity in pluralities," in *Proceedings of the Annual Boston University Conference on Language Development, Vol. 2* (Somerville, MA: Cascadilla Press), 387–399.

Pinto, M., and Zuckerman, S. (2018). Coloring book: a new method for testing language comprehension. *Behav. Res. Methods* 16, 1–20. doi: 10.3758/s13428-018-1114-8

Rouweler, L., and Hollebrandse, B. (2015). Distributive, collective and "everything" in between: interpretation of universal quantifiers in child and adult language. *Lingu. Netherlands* 32, 130–141. doi: 10.1075/avt.32.10rou

Schaeffer, J. (2003). "The characterization of specific language impairment," in *Language Competence Across Populations: Toward a Definition of Specific Language Impairment* (Mahwah, NJ: Psychological press; Erlbaum), 3–10.

Schmitt, C., and Miller, K. (2010). "Using comprehension methods in language acquisition research," in *Experimental Methods in Language Acquisition Research,* eds E. Blom and S. Unsworth (Amsterdam: John Benjamins Publishing), 35–56. doi: 10.1075/lllt.27.04sch

Tunstall, S. (1998). *The Interpretation of Quantifiers: Semantics and Processing.* (Doctoral dissertation). University of Massachusetts.

van der Ziel, M. E. (2008). *The Acquisition of the Weak Crossover Constraint: Evidence from Dutch Preschool Children*. MA thesis, Utrecht University.

Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca: Cornell University Press.

Zimmerman, I. L., Steiner, V. G., and Pond, R. E. (2010). *Preschool language scale*. San Antonio, TX: The Psychological Corporation. doi: 10.1037/t15141-000