'Two-pound cookies' or 'two pounds of cookies': Children's appreciation of quantity expressions

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

'Two-pound cookies' or 'two pounds of cookies': Children's appreciation of quantity expressions

Abstract

Inspired by Syrett (2013), three experiments explored children's ability to distinguish *attributives* (e.g., "three-pound strawberries," where MPs as adjectives signal reference to attributes) versus *pseudopartitives* (e.g., "three pounds of strawberries," where MPs combine with "of" to signal part-whole relations). Given the systematic nature of the syntax-semantics mapping, we asked whether children are able to use syntax to interpret how entities are quantified. In Experiment 1, 4 and 5-year-olds were asked to choose between two characters for the one who was selling appropriate items matching an attributive or pseudopartitive expression. In Experiment 2, children of the same age heard items described with a phrase using either an attributive, a pseudopartitive, "each" ("each weighs three pounds"), or "all together" ("all together they weigh three pounds"). At test, with some items removed, children were asked whether the same phrase applied to the remaining items (e.g., "Does Dora still have *three-pound strawberries*?"). Children did not distinguish between attributives and pseudopartitives, but did so for "each" and "all together." Experiment 3 extends the age range with a third experimental design. Children heard "each" or "all together" descriptions (e.g., "each strawberry weighs three pounds"), and judged, at test, which of two characters 'said it better' (i.e. "Mickey says 'these are *two pounds of strawberries*' but Donald says 'these are *two-pound strawberries.*'"). Children under six were at chance. Together, the three experiments suggest that despite its systematicity, children do not automatically appreciate the mapping between syntax and semantics.

We use number every day: to keep time and count money, to describe distances, and to measure ingredients. The developmental literature on children's numerical understanding, however, has focused primarily on a single, albeit fundamental, aspect of number use: children's understanding of how the count list is used to determine the numerosity of a set. This developmental understanding is deeply rooted in and built upon children's prelinguistic tendency to track and enumerate spatio-temporal individuals, and may consequently be a reason why children exhibit surprising behaviors when the entities to be enumerated are not typical individuals (Shipley & Shepperson, 1990; Huntley-Fenner, 1995; Sophian & Kalihiwa, 1998; Wagner & Carey, 2003; Brooks, Pogue, & Barner, 2011; Giralt & Bloom, 2001; Srinivasan, Chestnut, Li, & Barner, 2013). For example, when asked to count collections (e.g., "How many families?") or kinds (e.g., "How many kinds of animals?"), children count the individual members rather than the groups. When asked to count objects, some of which are broken, children count the individual pieces (e.g., describing two forks, one of which is broken into three pieces, as "five forks"). Children's failure on these sorts of counting tasks illuminates an important aspect of understanding number use. The units over which numbers apply are crucial. As poet Shel Silverstein pointed out, swapping one dollar for two quarters is not a good idea, even if *two* is greater than *one* (Silverstein, 1974). To achieve an adult-like understanding of number, children must come to understand not only the referents picked out by different units (e.g. a whole fork, and not a piece of one), but also the diverse ways in which numerals combine with units to represent quantities that are not about discrete individuals.

An important example of this, and the focus of our paper, is our use of number to quantify along continuous dimensions like time, length, volume, and weight. This more abstract use of number does not come easily to children, despite demonstrating sensitivity to changes along these dimensions as infants (Clearfield & Mix, 1999, 2001; Feigenson, Carey, & Hauser, 2002; Feigenson, Carey, & Spelke, 2002). For example, prelinguistic infants track changes in area or volume of non-cohesive substances like sand, but not the number of sand piles. This disregard for number as a way to quantify substances persists into preschool. In Huntley-Fenner (2001), when presented with two boxes containing different numbers of sand-filled cups, preschool children did not use the cups to quantify the amounts of sand, despite the fact that the cups were identical in size and were each filled to the brim with sand. Three-to five-year-old children could correctly answer the question "Which box has more cups?" regarding the discrete quantification of typical individuals, but performed more poorly when asked "Which box has more sand?" In English, as well as cross-linguistically (Corbett, 2000), we frequently use numerals in measure phrases, where the unit of quantification is specified. Thus, an adult in Huntley-Fenner's (2001) task would likely have described the arrays in terms of their numbers of "cups of sand," comparing the amounts of sand by comparing the number of full cups.

Several studies interested in children's application of discrete quantification to continuous extent have investigated their understanding of standard measures like inches, pounds, teaspoons, and minutes (e.g., Gal'perin & Georgiev, 1969; Piaget, Inhelder, & Szeminska, 1960; Nunes & Bryant, 1996; Clements & Sarama, 2009). Together these studies suggest that children do not fully grasp the meanings of such units until much after six or seven years of age. However, there is evidence that children are able to apply discrete quantification over continuous extent (e.g., Wang, Li, & Carey, 2013; Barner & Snedeker, 2006; Syrett, 2013) before they have fully grasped the meanings of standard units of measurement, which generally comes with formal schooling.

Preschool-age children may not know exactly what it takes for a pile of sugar to be "three tablespoons of sugar," or for a boulder to be "a fifty-pound rock," but what *do* they understand about measure phrase expressions like these? Following up on Syrett's (2013) work, the current paper explores English-speaking children's understanding of such expressions, with the view that this has the potential to be very revealing of their linguistic sophistication more broadly. Measure expressions in English are distinctly qualified to test children's appreciation of the combinatorial power of language because the quantities described by superficially similar expressions like "two-pound strawberries" (an attributive construction) and "two pounds of strawberries" (a pseudopartitive construction) are unambiguous to adults. In addition, the semantic relationships they represent do not rely on understanding anything beyond the dimension of the unit, but rather can be inferred from the syntax. In attributive measure phrases, the numeral combines with the measure word in an adjectival position to describe a property of individuals (each strawberry weighs two pounds), whereas with pseudopartitive phrases, the numeral and measure word combine with "of" to express a part-whole relation (the strawberries all together weigh two pounds). Children's comprehension of numerals in these contexts, then, can offer unique insight into their understanding of the mappings between syntax and semantics.

There are very few studies to date that have explored children's understanding of measure expressions. Wang, Li, and Carey (2013) investigated when children are able to coordinate number information with a measure and a noun in phrases like "two cups of rice." They found that whereas four-year-olds reject expressions like "two cups of rice" for scenarios in which there are two cups, and one cup is filled with rice and the other is filled with sand, three-year-olds often accept such expressions, as long as the numeral matches. Srinivasan et al. (2013) also found improvement between three and four years of age in children's understanding of the contrast between expressions like "two pieces of a shoe" and "two shoes" or "two half-shoes" and "two whole shoes."

Syrett (2010) conducted two experiments on measure phrases that were minimally different in wording on the surface. In the first, children were asked to choose between two pictures (e.g., a card with

two cups on it or two cards with five cups on each) for the one that either matched the attributive expressions ("two-cup card") or noun-noun compounds ("two cup-cards"). Though four-year-olds, but not three-year-olds, were above chance, it remains to be seen whether children could truly interpret these expressions, since they could succeed merely by attending to the plural morphology on the second noun ("card" vs. "cards"). In the second experiment, Syrett (2010) tested four-year-olds and asked about their acquisition of attributive (e.g., "two-cup cards") and pseudopartitive (e.g., "two cups of cards") expressions. Like Wang et al. (2013), the majority of four-year-olds could identify the referent for the pseudopartitive expression, picking the pseudopartitive referent (two cups each filled with cards) over the attributive foil (two cards each picturing two cups). However, when asked about the attributive ("two-cup cards"), four-year-olds chose randomly between the attributive referent (two cards each picturing two cups) and the pseudopartitive foil (two cups each filled with cards). In a third condition, performance on the attributives was even below chance; children who were asked for the "two-cup card" chose the pseudopartitive foil with two cups of cards over the attributive referent, a card with two cups on it. Syrett suggested that when children heard "two," they were lured into looking for two spatio-temporally discrete objects. This finding is quite similar to what Wang et al. found: three-year-olds were often lured by number alone, accepting "two cups of rice" for any situation involving a pair of cups, regardless of whether they contained another substance or were empty. These studies leave open the question of whether children truly understand attributive expressions.

Note, however, these studies primarily examined cases in which the numeral combined with the measure picks out a set of spatio-temporally discrete individuals. To date, only one study has explored how preschool children interpret phrases used to quantify continuous extent where the numeral does not enumerate discrete individuals. Syrett (2013) examined whether young children understand that the differences in the syntax of measure phrases result in distinct semantic properties. She looked at attributive (e.g. "three-pound strawberries"; "three-inch ribbon") versus pseudopartitive (e.g. "three pounds of strawberries"; "three inches of ribbon") measure phrases in two experiments.

Syrett (2013) interpreted her experiments as showing convergent evidence that preschoolers are beginning to appreciate the syntax-semantic mappings of these measure phrases. In Syrett's (2013) first experiment, participants were asked to judge the puppet's explanation of a scenario. In that experiment, participants were introduced to Dora, a girl who was shopping at a market. Dora always bought items that were best described with a pseudopartitive phrase (e.g. "three pounds of strawberries"), but the puppet always used an attributive phrase (e.g. "She bought three-pound strawberries") to describe what had been bought. There were two item types: 'count' (e.g. strawberries) and 'mass,' (e.g. four feet of ribbon/four-foot ribbon), and two orders, one in which children saw the count items first, followed by the mass items,

and the reverse. While they were overall at chance, children in the order that mass nouns followed count nouns rejected the puppet's utterance at slightly above chance level.[1] This was taken to be promising evidence that 4-year-olds are beginning to map attributive measure phrases to a property of individuals. Syrett posited that the poor performance for the mass nouns before count nouns group was due to a potential ambiguity of the mass items that she had initially overlooked; for example, "two-inch ribbon" could mean not only a ribbon that is two inches wide, but also a ribbon of the type that is cut into two inches.

Vignettes Shown

What Participants Heard

Attributive Condition

These are my **three-pound strawberries**. They **each weigh three pounds**.
If a mouse comes along and nibbles some, **do I still have three-pound strawberries**?

Pseudopartitive Condition

That's **three pounds of strawberries** – **all of that together is three pounds of strawberries**.
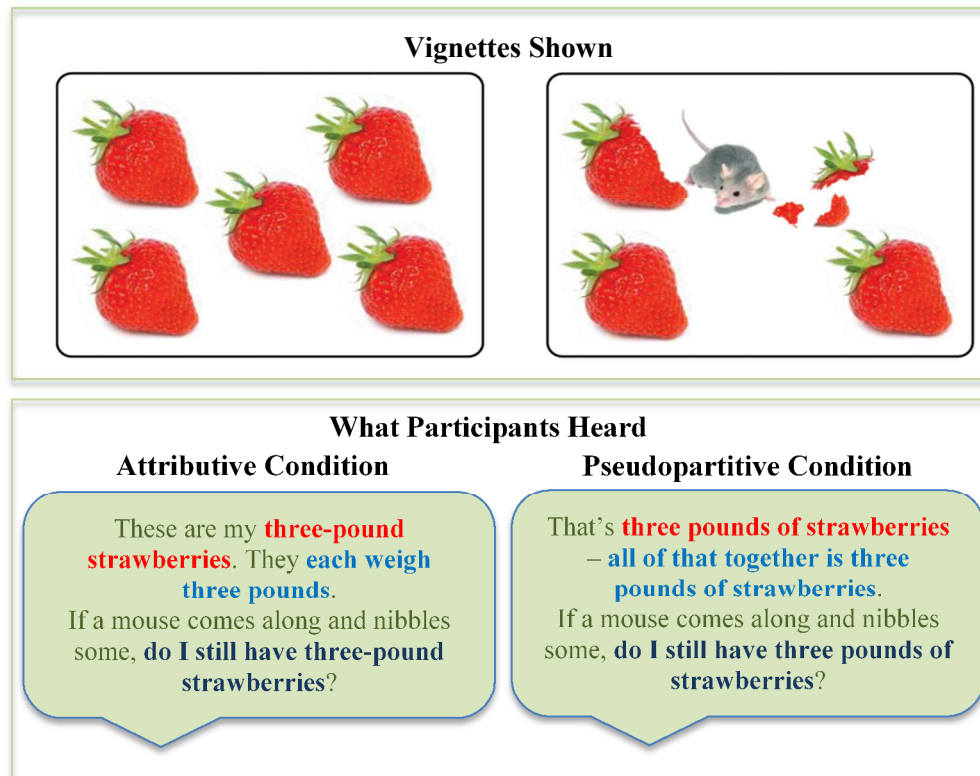If a mouse comes along and nibbles some, **do I still have three pounds of strawberries**?

Figure 1. Relevant snippets of a sample trial from Syrett's (2013) Experiment 2, which involved a "subtraction" of quantities. Children were presented with a story via picture slides with either attributive or pseudopartitive information. For more details, see Syrett's (2013) Appendix.

In her second experiment, children were introduced to an array of items. One group of children was introduced to an array described by an experimenter using "each" (e.g. "The strawberries were

---

[1] The children in the mass nouns followed by count nouns group correctly rejected the puppet's utterance 63.9% of the time. Syrett did not report whether this percentage is statistically above 50% chance. Instead, the percentage she reported as being statistically above chance was the percentage after removing children who had an (incorrect) response bias to accept the puppet's utterances. However, it is unclear whether removing such children is justified as doing so would favor a false positive.

enormous. They each weighed three pounds! These are my three-pound strawberries."), while the other group heard the array described using "all" (e.g. "Look, that's three pounds of strawberries – all of that together is three pounds of strawberries."). In the test trials, some of the items were removed (see Figure 1 for an example). The first group then responded to a question with an attributive construction ("Do I still have three-pound strawberries?"), while the second group answered a question containing a pseudopartitive expression ("Do I still have three pounds of strawberries?"). Taking away items from a set does not change the weight of each individual in the set, but changes the total weight of the set. Hence the first group should answer positively, while the second group should answer negatively. Children accepted the attributive description in greater proportion than they did the pseudopartitive one (55% vs. 23%).

If replicable, Syrett's findings are remarkable in two ways. First, the findings suggest that preschoolers, even before any formal instruction and education on standard measures, appreciate whether and how measured quantities are conserved under various transformations. Her children had to know which dimension of quantity is preserved or changed when objects were removed from a set. This is in contrast to developmental studies showing that same-age preschoolers often fail to realize which kind of transformation affects quantity (Piaget, 1965; Piaget et al., 1960; Elkind, 1967). That is, most preschoolers often incorrectly reason that a ball of clay flattened is "more clay" than before, or a cup of water poured into a thinner container is now "more water." Perhaps if developmental psychologists had been asking children about quantities with number and measure words, they would have discovered that children know more about transformation and conservation than initially believed (e.g., see Lipton & Spelke, 2006). The reason behind children's responses in those Piagetian studies could be related to the difficulty children have in understanding the word "more." After all, "more" is used in many different ways with different meanings (H. Clark, 1970; Gathercole, 1985; Barner & Snedeker, 2005). Sometimes "more" can be used in an additive manner and other times it can be used as a comparative, and the dimension of comparison can vary (Barner & Snedeker, 2005).

Second, Syrett's findings also suggest that as soon as children are able to appreciate notions of measurement, they can make sense of how syntax is used to distinguish the attributive meaning from the pseudopartitive one. However, these results are inconclusive with respect to children's understanding of the attributive and pseudopartitive constructions. Recall that children were explicitly given the meanings of the measure phrases in the introduction (e.g., "The strawberries were enormous. They each weighed three pounds! These are my three-pound strawberries."). Thus, as acknowledged by Syrett (2013, p. 236), children could have answered correctly regarding what the experimenter "still" possessed on the basis of

their understanding of "each" or "all together" and not on the basis of their understanding of the measure phrases.

In the current paper, we therefore asked whether we could replicate Syrett's findings. We start with a modified version of her Experiment 1, in which we tested children exclusively on the count nouns, where her children performed best. We also adapted her Experiment 2 specifically to investigate the potential comprehension aid provided by the use of "each" and "all together" in the test prompt. Finally, in Experiment 3, we used a "Who said it better" task with a wider age range to probe the developmental trajectory of children's ability to map different measure expressions to the sets they describe.

Experiment 1: Sentence Matching

Experiment 1, like Syrett's Experiment 1, involved a character who went shopping at a market and encountered two vendors. One vendor sold items that matched the pseudopartitive expression (e.g., two pounds of strawberries: small strawberries that weighed two pounds all together) and one sold items that matched the attributive expression (two-pound strawberries: large strawberries that weighed two pounds each). In Syrett's study, the customer always bought the quantity that matched the pseudopartitives (i.e., small strawberries that weighed two pounds all together), and children had to judge whether the puppet could use the attributive to describe the purchased quantity. A response bias, always preferring to reject the puppet's utterances (or always preferring to accept them), is thus confounded with the correctness of the response; one cannot be certain that a child's rejection of the puppet's utterance across trials is because the participant truly understood the attributive construction or whether the child just has a bias to say, "No." Instead of a truth-value-judgment task, our children were asked to choose between two vendors for the one who is selling what the character wanted to buy (e.g., "two-pound strawberries"). This design allows us to counterbalance what the vendors are selling and assign each vendor be the correct only half of the time, and thus to dissociate response bias (always preferring one vendor) from the correct response. Additionally, whereas Syrett only asked children to judge the attributive expressions, children in our experiment were asked to judge the pseudopartitive expressions (e.g., "two pounds of strawberries") as well, which allowed for a more powerful comparison across the two expressions within participants. Children were told and reminded each time that the character would buy up everything offered by one vendor. This ensured that when it came to buying two pounds of strawberries, the character would not be buying from the two-pound strawberry vendor, who had more than two pounds of strawberries.

Methods

Participants. Syrett tested children between 3;6 to 5;3, with a mean age of 4;3. Given that her children, as a group, performed at chance, we tested children of a slightly older age range. 10 children between the ages of 4;1 and 5;10 (mean: 5;0, SD: 7.0 months; 4 females, 6 males) participated, and were randomly assigned to one of two random test orders. Children were recruited from public parks in the Boston area or through the laboratory database, and tested individually at a quiet and non-distracting area at the park or laboratory.

Procedure. Children were introduced to three characters (SpongeBob, Sandy and Patrick) in a picture book. In the story, SpongeBob went to visit his two friends, Sandy and Patrick, who were salesclerks at a mall. To help Sandy and Patrick finish selling their goods so they could all go play, they made a game out of shopping. On each trial, Sandy and Patrick each presented goods (e.g., Sandy might present two tires while Patrick presents three) for SpongeBob to purchase. Children were told that SpongeBob would have to buy "everything sold by one friend" each time. SpongeBob then indicated what he wanted to purchase (e.g., "I want to buy only *two tires*"), and the child would have to point to which of the two friends SpongeBob should choose. In this example, Sandy would be the correct choice and not Patrick, even if Patrick did have two tires (out of three total) for sale.

All together, there were 12 trials consisting of four controls and eight test trials (four pseudopartitives and four attributives). The control trials were interspersed, such that the first two were in the beginning of the twelve trials while the third and fourth trials were at the middle and end of the experiment. These trials consisted of asking children to pick on the basis of properties: number ("*two* tires"), color ("*red* peppers"), or kind ("two scoops of *chocolate ice cream*"; "two red *boats*"). These trials were chosen to be ones that children in this age range could easily understand and respond to correctly. The beginning trials served as warm-ups to ensure that children understood the procedure, the middle and end control trials verified that children were still paying attention.
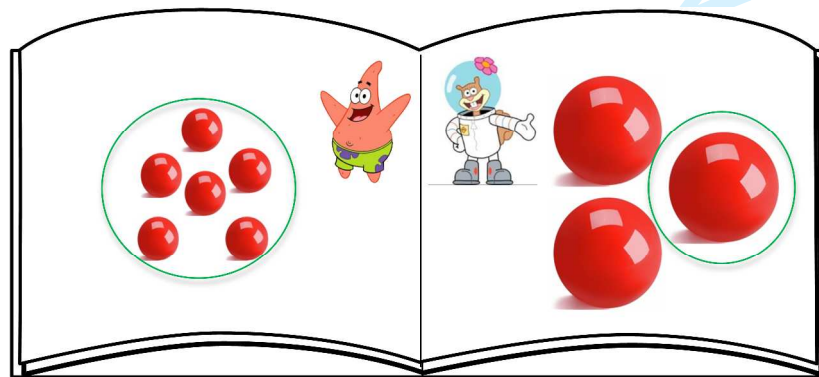

Figure 2. Sample stimuli from a trial of Experiment 1. For the page on the left, children were told the weight of the entire set, and for the page on the right, children were told the weight of each individual.

The test trials were randomly interspersed across the remaining eight positions in the twelve trials. For each test trial, one friend would sell items that matched a pseudopartitive description while another would sell items that matched the attributive description. For example, Figure 2 shows an example where children were told that Patrick was selling "small balls that all together weighed two grams" (pseudopartitive-match) and that Sandy was selling "huge balls that each weighed two grams" (attributive-match). To emphasize that the weight was all together two pounds, a green circle encircled all the balls. The experimenter also used her index finger to draw a circle around the entire set ("Remember. All together, the balls weigh two grams."). To emphasize that the weight was about each individual, a green circle surrounded one individual ball, and the experimenter circled each object with her index finger ("Remember. This one weighs two grams. This one weighs two grams… Each weighs two grams"). After introducing the child to the items offered, SpongeBob then made his request ("I want to buy only *two grams of balls*"). The child was then asked to choose the person who was selling what SpongeBob wanted ("If SpongeBob wants to only buy *two grams of balls*, who should he buy from?").

Whether the test trial was considered pseudopartitive or attributive was determined by what SpongeBob said he wanted (e.g., pseudopartitive: "I want to buy only *two grams of balls*" or attributive: "I want to buy only *two-gram marbles*"). Whether Sandy or Patrick had the pseudopartitive- or the attributive-matched items was counterbalanced across trials. Half of the time Sandy had what SpongeBob requested and half of the time Patrick did. The test constructions always had the numeral "two," which is a high-frequency number word and a known number word to children this age (i.e., "two M of Ns" and "two-M Ns"; M = measure word, N = noun; Wynn, 1991). To provide the child with the best possibility of understanding the expression and succeeding, the number of objects that Sandy and Patrick sold always exceeded two. This removed the possibility that children would (mistakenly) choose on the basis of the number of objects in the set. Additionally, our test constructions always had a weight measure and the nouns were count nouns, as Syrett (2013) found that children performed worse with length measures and mass nouns than with weight measures and count nouns. Furthermore, in Syrett's study, length uses of attributive phrases were potentially ambiguous. Thus, by choosing to only ask about weight, we bypassed the ambiguity problem.

The measures phrases tested were "two pounds of strawberries," "two-pound cherries," "two tons of blocks," "two-ton bricks," "two tons of blocks," "two ounces of crackers," "two-ounce cookies", "two grams of balls" and "two-gram marbles."

Results

Children scored 100 percent correct on the control trials, indicating that they understood the task and were engaged throughout the experiment. They scored 48.8% correct on the test trials, with 60.0% correct on the attributive and 37.5% correct on the pseudopartitive trials. Performance on these two types of test trials did not differ from each other (paired $t(9) = 1.65$, $p = .13$, n.s.; Wilcoxon $Z = -1.48$, $p = .14$, n.s.), nor were they above chance (60% vs. 50%: $t(9) = 1.0$, $p = .34$, n.s., Wilcoxon $Z = -1.13$, $p = .26$; 37.5% vs. 50%: $t(9) = -1.63$, $p = .14$, n.s., Wilcoxon $Z = -1.51$, $p = .13$).[2]
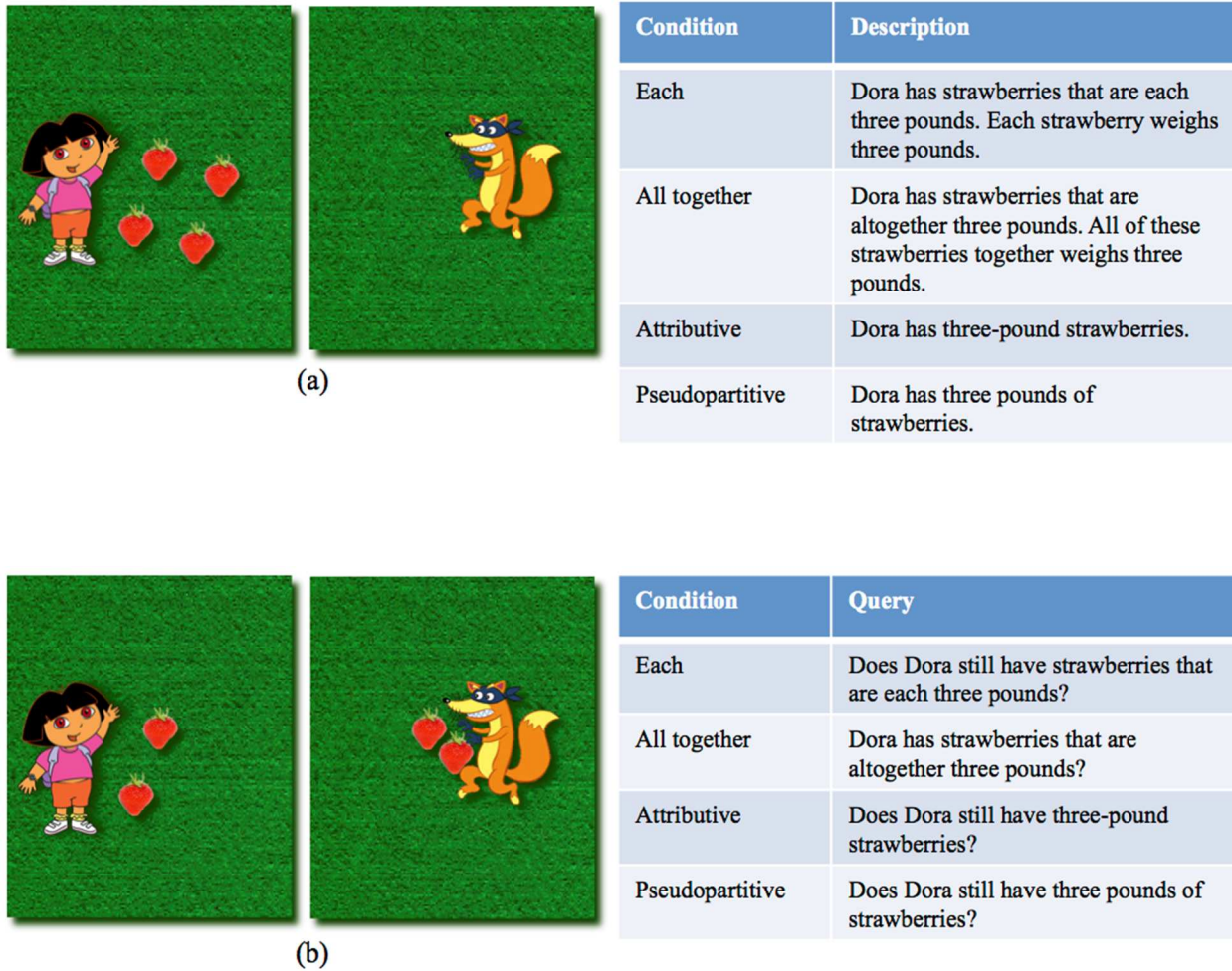
Discussions

Children's responses on the control questions indicated that they understood the task. For example, when told that SpongeBob wanted to buy "two tires," children correctly chose Sandy, who had two tires, and not Patrick, who had three. They correctly selected on the basis of numerical information, adjectival information (*red* and not *yellow*), noun information (*boat* and not *train*), and a combination of adjective and noun information (*chocolate ice cream* and not *vanilla ice cream*). However, children did not perform above chance level on either the pseudopartitives or the attributives. We chose to test children who were slightly older than Syrett (2013) had. We also chose to test children using the same count nouns and measure phrases as Syrett's, because those were the items where children showed a greater promise of understanding. We made sure the number ("two") used in the test expressions did not match the number of individuals depicted. We also emphasized to the children that these standard measures were about weight. None of these manipulations seemed to have helped. The glimmer of success in Syrett's data is likely not robust. We probed further with a second experiment as replication, by adopting her Experiment 2, in which children showed some evidence of understanding. However, recall that the differentiation between pseudopartitive and attributive trials could have been due to the extra information that children were told about these two kinds of constructions. We tested whether this was a possibility in the next experiment.

Experiment 2: Subtraction Method

Experiment 2 made use of the clever subtraction method devised by Syrett (her Experiment 2; see Figure 1 for Syrett's design) to test children's understanding of the link between the measure phrases and monotonic part/whole relations. The task involved a character, Dora, who on every trial had a set of items that was described with either an attributive or a pseudopartitive phrase. Each time, some of the items

---

[2] With a small N and with each children being tested on only four questions per condition, we conducted distribution-free non-parametric tests. However, we included the more common parametric t-tests as Syrett (2013) had done.

were taken away from her. Children then had to decide whether what remained could be described by the same phrase (e.g., "Does Dora still have three-pound strawberries?"). Unlike Syrett, we had four types of trials (see Figure 3 and compare with Figure 1) that crossed two levels of linguistic expressions (measure phrases or quantifiers) with two levels of meaning (whether the expression was about *each* or about *all*). The design allowed us to tease apart whether children at this age are really distinguishing between pseudopartives and attributives, or whether they are only succeeding on the basis of the quantifier descriptions ("each strawberry weighs three pounds," "all strawberries together weigh three pounds") that were given. Syrett suggests the possibility that the quantifier descriptions and measure phrases could have worked in tandem in contributing to children's success. Presenting the two types of instruction in isolation enables us to efficiently determine whether in fact the quantifier descriptions alone could have been responsible for children's performance. Furthermore, presenting the two types of instruction in blocks counterbalanced across participants allows us to look at order effects revealing how hearing one form of describing the relevant quantification concepts might make those concepts more or less accessible for their description in the other form.

Condition	Description
Each	Dora has strawberries that are each three pounds. Each strawberry weighs three pounds.
All together	Dora has strawberries that are altogether three pounds. All of these strawberries together weighs three pounds.
Attributive	Dora has three-pound strawberries.
Pseudopartitive	Dora has three pounds of strawberries.

(a)

Condition	Query
Each	Does Dora still have strawberries that are each three pounds?
All together	Dora has strawberries that are altogether three pounds?
Attributive	Does Dora still have three-pound strawberries?
Pseudopartitive	Does Dora still have three pounds of strawberries?

(b)

Figure 3. Depiction of a sample trial in Experiment 2. A trial begins as (a), with Dora possessing a set of items (e.g., strawberries) on one placemat and Swiper the Fox on the other placemat. Children heard descriptions about Dora's possessions that differed by condition. Swiper then stealthily stole some items from Dora, returning with them to his placemat side (b). Finally, children heard their condition-specific query.

Methods

Participants. 24 children matching the age range targeted by Syrett's study participated. The children ranged between 4;1 and 6;2 (mean = 4;10; SD = 6.0 months; 12 females, 12 males), were recruited through the laboratory database or through daycares in the greater Boston area, and were tested individually at a quiet and non-distracting area in the laboratory or at school.

Procedure. Children were introduced to two characters, Dora and Swiper. The characters are from a popular animated TV show, where Swiper is a thieving fox who often steals from Dora. On each trial, toy objects would first be given to Dora (e.g., "Dora has three pounds of strawberries"). Then

Swiper would come along and steal some of the objects from Dora. The child would then have to answer a 'yes/no' question regarding what Dora has left (e.g., "Does Dora still have *three pounds of strawberries*?"). The study consisted of a block of eight quantifier trials and a block of eight measure phrase trials with the blocks counterbalanced across children. The eight quantifier trials consisted of four "each" trials and four "all together" trials (see Figure 3 for an example of what children saw and heard on each of the four types of trials). The eight measure phrase trials consisted of four pseudopartitive trials and four attributive trials. The measure phrases were in the form of "three M of Ns" and "three-M Ns" (M = measure, and N = Noun). Additionally, the subtraction always involved four objects minus two, thus the number of objects initially and after removal never equaled to three. This manipulation was intended to push children to more carefully analyze the measure phrase and not be lured by matching the numeral with the number of individuals in the set. The measures and count nouns were the same as those in Experiment 1. The order of the two kinds of semantic trials within each block, when the semantics was about the group (pseudopartitive/all together) and when the semantics was about the individuals (attributive/each), were randomized to create two orders, one the reverse of the other.

In addition there were four control questions to make a total of 20 trials. The control questions were easy questions that served as a check to see if children understood the instructions and were paying attention. Half of the control trials had "yes" as the expected adult answer (e.g., "Does Dora still have small balloons?" when Dora started out with four small balloons and Swiper took away two) and half had "no" as the expected adult answer (e.g., "Does Dora still have small balloons?" when Dora started out with two big balloons and two small balloons and Swiper took away the small ones). The "yes" control trials were thereby designed to mirror the attributive test trials such that the property queried still held for the set of remaining objects. The "no" control trials mirrored the pseudopartitives in that the property being asked about was no longer true of the set after removal. Two control trials, one with an expected "yes" and one with an expected "no" response began the 20 trials to familiarize children to the structure of the task. The remaining two control trials were placed within the remaining trials.
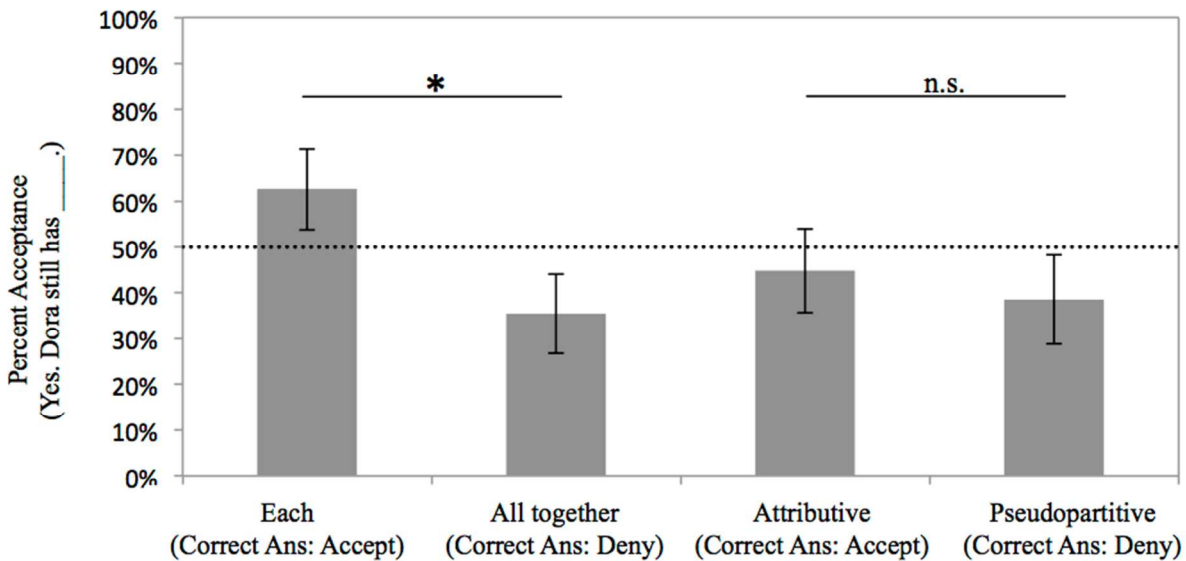
Figure 4. Experiment 2 Results. Results are presented by percentage of acceptance. "Each" and Attributive conditions should receive a high rate of acceptance while "All together" and Pseudopartitive conditions should receive a high rate of rejection. No conditions were by themselves significantly different than chance (t-tests and Wilcoxons). When compared across conditions via paired t-tests and Wilcoxons, * indicates p < .05.

Results

       Children performed well above chance on the control trials (99.0% correct; $t(23) = 47$, $p < .001$; Wilcoxon $Z = -4.81$, $p < .001$). Their performance on the early control trials verified that they understood the task, while their performance on the middle and late control trials showed that they were paying attention throughout the experiment. Figure 4 plots the data for the remaining trial types (each, all together, attributives, and pseudopartitives). Following Syrett (2013), Figure 4 plots the percentage of trials in which children accepted the linguistic expression as still describing the scene after subtraction. Removing items does not affect the weight of individuals, but affects the weight of the entire group. If children understood the linguistic expressions, they should still accept the expressions for the each/attributive (weight about individuals) trials, but not for the all together/pseudopartitive (weight about the group) trials. A generalized estimating equation logistic regression[3] of construction type (quantifiers vs. measure phrases) x semantics (expressions on weight about individuals vs. expressions on weight about the group) corroborated that children tended to accept expressions concerning weight about individuals more often than expressions concerning weight of the group after some individuals were

---

[3] The effect of block order (MPs first or quantifiers first) was not included in the model because initial analysis showed that order did not affect tendency to accept or reject the linguistic expression.

removed (individuals: 53.6% accept vs. group: 37.0%; Wald $\chi^2(1) = 12.52$, $p < .001$). There was no difference between the likelihood of acceptance by construction type (quantifiers: 49.0% vs. MPs: 41.7%; Wald $\chi^2(1) = 1.35$, $p = .25$). However, there was a significant interaction between construction and semantics (Wald $\chi^2(1) = 12.16$, $p < .001$). In particular, acceptance rate differed for the quantifiers (Each: 62.5% vs. All together: 35.4%; $t(23) = 4.03$, $p = .001$; Wilcoxon $Z = -3.37$, $p = .001$), but not for the measure phrases (Attributive: 44.8% vs. Pseudopartitive: 38.5%; $t(23) = 1.66$, $p = .11$; Wilcoxon $Z = 1.67$, $p = .10$). When scored in terms of percentage correct, children performed better on quantifiers than measure phrases (quantifiers: 63.5% correct vs. MPs: 53.1%; $t(23) = 3.62$, $p = .001$; Wilcoxon $Z = -2.98$, $p < .01$), and were above chance on the quantifiers ($t(23) = 4.03$, $p = .001$; Wilcoxon $Z = -3.37$, $p = .001$), but not the measure phrases ($t(23) = 1.66$, $p = .11$; Wilcoxon $Z = -1.67$, $p = .10$). However, despite better performance on the quantifiers than measure phrases, the two were correlated ($r = .52$, $p = .01$). A cross-tabulation using 75% correct as passing indicated that while there were passers of the quantifiers trials who failed the measure phrase trials (5 out of 24 children), there were no children who passed the measure phrases and failed the quantifier trials.

Discussions

The control trials followed the same structure as the test trials in that when some objects were removed, for half of the trials, the queried expression still matched the remaining objects while for the other half of the trials, it did not. Children had no problems on these questions, scoring almost perfectly on them. In contrast, most children struggled with the test questions. For the test trials, performance on the quantifiers was better than on the measure phrases. As a group, they more often correctly accepted the "each" question, while rejecting the "all together" question. In comparison, they responded basically equivalently for pseudopartitives and attributives, accepting both at similar rates.

What do we make of the relative performance on quantifiers and measure phrases? While performance on the quantifiers correlated with that on the measure phrases, there were some children who were better at the quantifiers than they were at the measure phrases, but none who were better at the measure phrases than the quantifiers. This suggests that whereas children may have notions about how weight changes (which they must in order to answer the each/all together questions correctly), they have not yet made the connection of how the syntax maps onto those notions. Consistent with this, we found no effects of order. Had it been the case that the quantifier descriptions in Syrett's experiment *scaffolded* children's understanding of the measure phrases, as she suggests, we might have expected a benefit on the attributive/pseudopartitive questions for the group who was tested with *each*/*all together* questions during the first block relative to the group who was exposed to the attributive/pseudopartitive questions first. As

it is, however, children performed equivalently on the measure phrase block, regardless of whether it came first or second (first block: 53.1% vs. second block: 54.2% correct).

In relation to Syrett's data, ours suggest that children's differential acceptance of her attributive and pseudopartitive sentences is attributable to the extralinguistic information about "each" and "all .... together" that children were given when introduced to the sentences. We found no evidence that children at this age are truly on their way to understanding the two constructions, which is consistent with the results of our Experiment 1 and possibly with the results of Syrett's Experiment 1.

We try to replicate our findings again in the next experiment. In our new experiment, though we like the cleverness of the subtraction task, we avoided testing children on any changes of weight, which may require them to have some understanding of weight conservation. Instead, we opted for a simpler design, using a "Who said it better?" task that contrasted two linguistic expressions (e.g., attributive vs. pseudopartitive) and asked the participant to pick the one that better described a given scenario. Also known as the felicity judgment task, this task was initially designed to test children's understanding of scalar implicatures (Chierchia et al. 2001). Because children were given the contrasting linguistic expressions (e.g., "Ernie said, 'Cookie Monster ate some of the cookies' and Bert said, 'Cookie Monster at all the cookies.' Who said it better?"), children tend to perform better on this task than the truth-value judgment task where they had to verify whether a single sentence applied to a given scenario (Chierchia et al. 2001; 2004; Foppolo et al. 2012). This "Who said it better?" task has also been successfully employed to test very young children. For example, it has been used to test two- and three-year-olds' knowledge of measure phrases such as "two pieces of sock" versus "two socks" (Srivinisan et al., 2013; Li et al., 2014). Thus, the "Who said it better?" task has the potential of revealing children's understanding of attributives and pseudopartitives. Lastly, in Experiment 3, we also extended the age range in order to see when children start to understand the measure phrase constructions.

Experiment 3: Who said it better?

Methods

Participants. Eleven four-year-olds (range = 4;2 to 4;11, mean = 4;7, SD = 3 months, 7 females, 4 males), 12 five-year-olds (range = 5;0 to 5;11, mean = 5;4, SD = 4 months, 3 females, 9 males), and 12 six-years-olds (range 6;0-7;0, mean = 6;7, SD = 3 months, 6 females, 6 males) participated, and half were assigned to one of two randomized orders (one the reverse order of the other). Children were recruited in the same manner as Experiment 1 and tested individually.
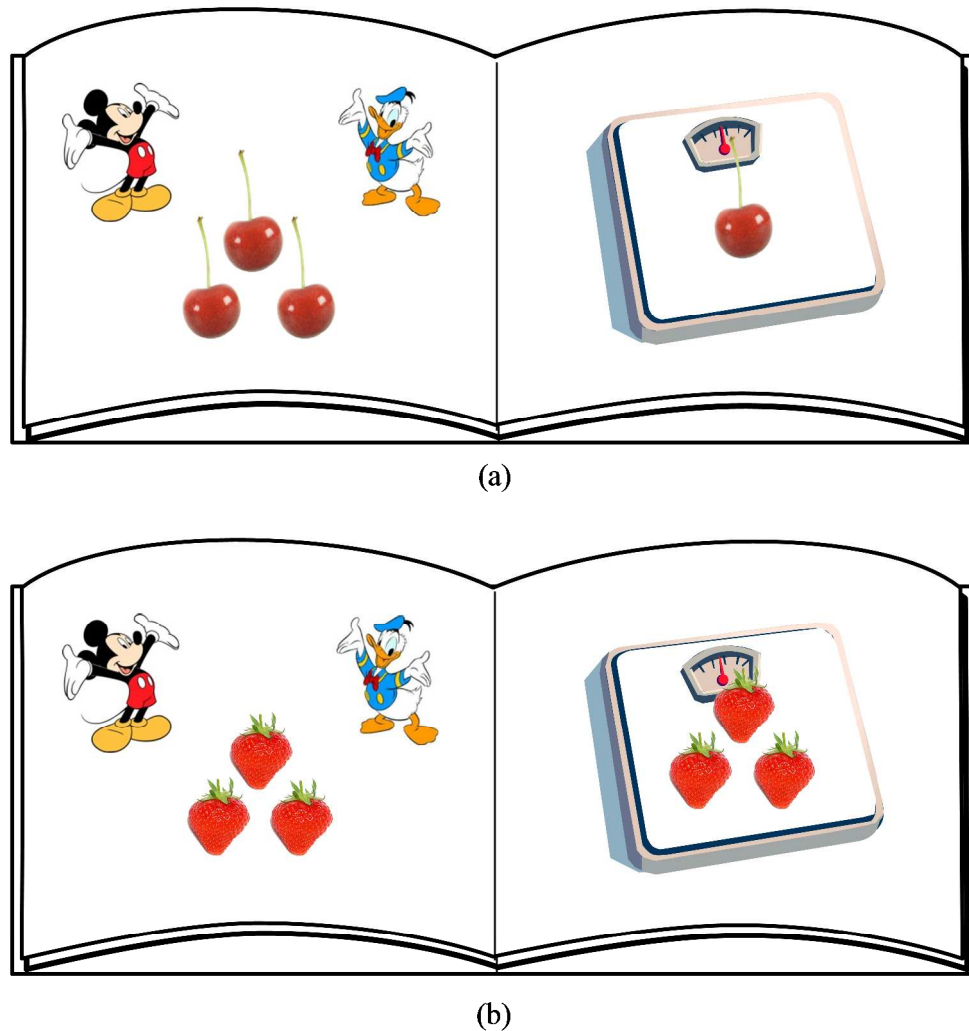
(a)

(b)

Figure 5. Sample trials of Experiment 2. Children were told about the weight of items. Either the items each weighed a particular amount (a), or else the items all together weighed a particular amount (b).

Procedure. Using a picture book, children were introduced to two characters (Mickey and Donald) from Disney, and told that they were going to play a game of "Who said it better?" Mickey and Donald would describe the pictures in the book and the child's job was to decide who did a better job describing the picture.

The children were tested on twelve trials, beginning with four control trials followed by eight test trials. For the test trials, one character would utter the attributive phrase while another character would utter the pseudopartitive phrase (four attributive-picture trials and four pseudopartitive-picture trials). Figures 5a and 5b provide examples for the attributive-picture and pseudopartitive-picture trials, respectively. One page shows Mickey and Donald with the items in question, and the facing page shows an item or items on a scale.

For the attributive picture trials, using Figure 5a as example, we told the children, "Look at these big cherries, they each weigh two pounds." Then we pointed to the scale with one cherry on it, "If you put one cherry on the scale, you can see, that one cherry weighs two pounds. Now Mickey says 'these are two pounds of cherries' but Donald says 'these are two-pound cherries'. Who said it better?"

For the pseudopartitive picture trials, using Figure 5b as example, we told the children, "Look at the strawberries, they weigh two pounds all together. If you put all strawberries on the scale, you can see that they are all together two pounds. Now Mickey says 'these are two pounds of strawberries' but Donald says 'these are two pound strawberries'. Who said it better?"

The measure words in the test phrases were the same ones as Experiments 1 and 2. Mickey always uttered the phrase first, but half of the time it was an attributive phrase and half of the time it was pseudopartitive phrase. Mickey was right on half of those trials while Donald was right on the other half.

The first two control trials familiarized children to the task with simple phrases that children are likely to distinguish (e.g., Now Mickey says "These are green peppers," but Donald says "These are yellow peppers"). The next two control trials had the same kind of picture and structure as the test trials. Children heard descriptions about the pictures just like the test trials, except instead of pitting the attributive phrase with the pseudopartitive phrase, children heard an "each" phrase pitted against an "all together" phrase. For example, children heard "Now Mickey says 'these bananas all together weigh two ounces' but Donald says 'these bananas each weigh two ounces'." For one trial, the "each" phrase was correct and for one trial, the "all together" phrase was correct. For the control trials, half of the time Mickey was correct.

Results

Figure 6 plots the percentage correct for each age group for the control trials (Figure 6a) and the test trials (Figure 6b). The control trials established that the children understood the task, with children of all three age groups scoring at above chance level (p's < .01 for both t-tests and Wilcoxon signed rank against 50%). There was no age difference across the three groups (see three rightmost columns of Figure 6a, 4s: 88.6% correct, 5s: 93.8%, 6s: 91.7%), but children scored better on the quantity irrelevant trials than the quantity relevant trials (98.6% vs. 84.0% correct; see also the first and second sets of three columns of Figure 6a, which breaks up the data by quantity-relevant and quantity-irrelevant trials respectively). The non-significant age effect and significant trial type effect were verified via a logistic regression model of whether the response was correct (age: Wald $\chi^2(2) = 2.08$, p = .35; trial type: Wald $\chi^2(1) = 7.53$, p = .01).
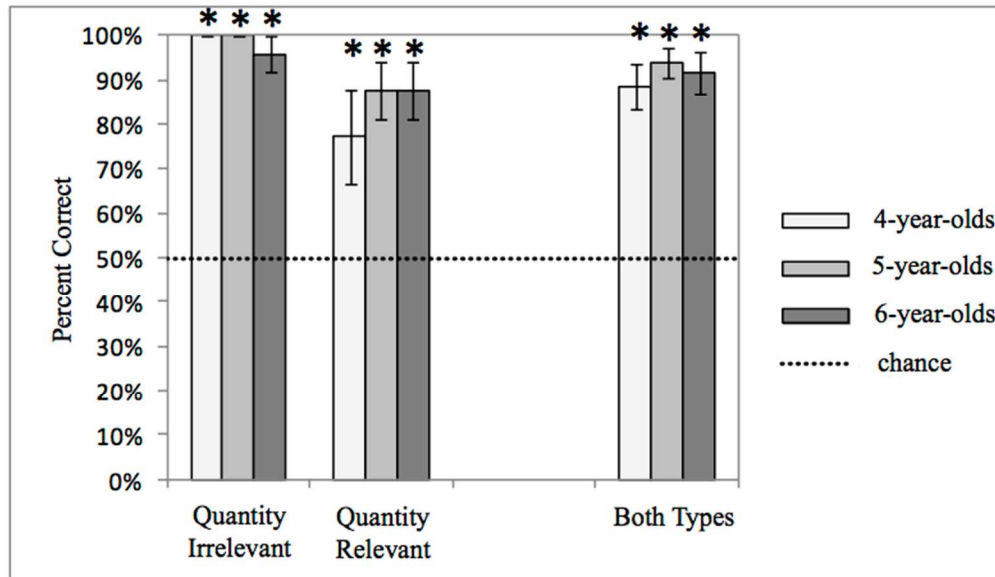
(a)
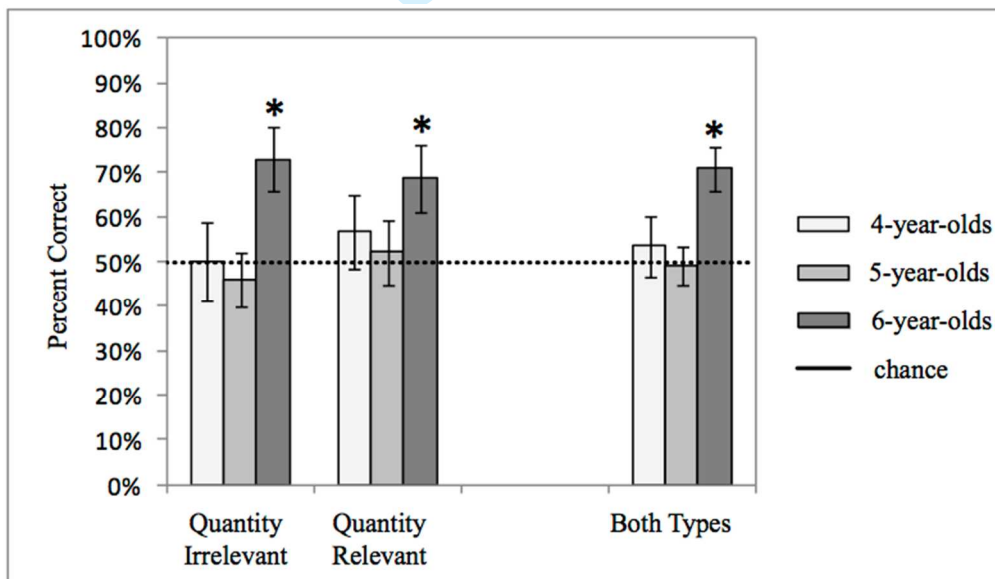
(b)

Figure 6. Experiment 3 Results. Percentage correct by age for the control trials and the test trials are separately depicted in (a) and (b) respectively. The rightmost three bars of each figure show the data collapsing across both types of control or test trial. * p < .05 for t-tests and Wilcoxons against chance.

Turning to the test trials (Figure 6b), 4- and 5-year-olds performed at chance level (see Figure 6b's three rightmost columns, 4s: 56.8%, 5s: 49.0%; p's > .61 for t-tests against 50% and p's > .66 for Wilcoxon signed rank tests against 50%) while 6-year-olds were above chance level (70.8% correct, t(11) = 4.21, p= .001 and Z = -2.72, p < .01), indicating an improvement with age on children's understanding

of measure phrases. Figure 6b also indicated that there was no difference between whether the children were better at the pseudopartitive or the attributive scenarios (compare the first set of three columns, average 56.3% correct, with the second set of three columns, average 59.2% correct). Whether children responded correctly was entered into a generalized linear logistic regression model, which verified the effect of Age (Wald $\chi^2(2)$ = 7.35, p = .025), and lack of effect of Trial Type (Wald $\chi^2(1)$ = .682, p = .41) and Trial Type by Age interaction (Wald $\chi^2(2)$ = .344, p = .84).

Discussions

Although children performed above chance on all control questions, interestingly, they made slightly more mistakes on the quantity-relevant control trials in which they had to simply choose the character who parroted the descriptions the experimenter provided about the picture. That is, when told "Look at these big bananas, they each weigh two ounces. If you put one banana on the scale, you can see that one banana weighs two ounces," the child simply had to choose the character who said "Each banana weighs two ounces" and not the character who said "The bananas all together weigh two ounces." The mistakes indicate that perhaps not all children understood the provided descriptions about weight, and that concepts about weight are difficult for children. Given this, children should find the attributive versus pseduopartitive distinctions even more difficult. This was evidenced by their chance performance, as well as by comments made by some of the younger children during the study, which revealed their lack of understanding that the two characters' descriptions were different. For example, one child (4;7) said "both are right" and another (4;6) said "but they both said the same thing." The children were oblivious to the subtle linguistic differences, and our findings show that it is not until six years of age that children begin to distinguish the two constructions.

General Discussions

In Experiment 1, we found that children could not reliably match a pseudopartitive or attributive measure expression to an array, even with the procedural adjustments we made in light of Syrett (2013). Despite testing older children, limiting our critical nouns to those where subjects had shown the most success in Syrett (2013), ensuring children could not confuse the numeral in the measure phrases as referring to the numerosity of the set, and emphasizing that the dimension of measurement was weight, children did not succeed. In Experiment 2, we sought to disentangle children's grasp of "each" and "all together" from their understanding of the semantics of measure phrases in an effort to explain the success of subjects in Syrett (2013). To do so, we adapted her subtraction task, and found that while children were sensitive to how the applicability of a weight description using "each" or "all together" varied following subtraction, they did not distinguish between attributive and pseudopartitive measure phrases. Finally, in

Experiment 3, we used a simplified method where children were asked to choose directly between an attributive and pseudopartitive description of an array. Children below age six performed at chance.

While contrary to Syrett's (2013) findings, our results are consistent with Piagetian suggestions that preschoolers do not grasp how different transformations, like the subtraction method used in this task, affect quantity (Piaget, 1965; Piaget et al., 1960; Elkind, 1967). In light of these, it is not surprising that children in our study did not make the distinction between the weight of individuals and the weight of a set. Young children incorrectly guess that a ball of clay, when mashed flat, will not weigh the same amount when replaced on a scale. If they fail to understand how weight is conserved in such a task, how can they be expected to understand transformations of sets whose descriptions hinge on an understanding of how weight is measured and conserved?

The above studies suggest that the mapping between syntax and semantics may not be immediately transparent to children. Experiment 2 suggests that while children make the semantic distinction between a property of an individual and a property of a set, they have yet to infer this semantic information from the syntactic structure. Namely, it seems they are not initially making the connection between the adjectival position of the measure phrase in an attributive expression and its quantification of an individual, and the "of" in a pseudopartitive expression and its description of a part-whole relation. However their potential success may have been masked by the processing demands imposed by reasoning about weight, in particular. Previous work suggests that weight may be an especially difficult dimension for children to grasp, and understood later than length (Brown et al., 1995; Gulkoa et al., 1988; Elkind, 1961). In Schrauf, Call, and Pauen (2011), for example, children younger than four did not have clear expectations of different weights' effects on a balance scale. It is possible that children would be better able to make use of the syntax-semantics mappings in our task were the content less trying.

To address this concern, we might additionally test children on their understanding of weight (Schrauf, Call, & Pauen, 2011; Smith, Carey, Wiser, 1985). If their failure on the present studies were due to conceptual difficulty, rather than insensitivity to the regular mappings between syntax and semantics, we would expect those children who understand weight to also understand the language. Along similar lines, we might also test children on a dimension for which they demonstrate an earlier understanding, like length, thereby lessening the processing load (Clements & Sarma, 2009; Nunes & Bryant, 1996; Sarama, Clements, Barrett, Van Dine, & McDonel, 2011). If the same methodologies were used, but the critical expressions involved units for measuring length, would children be able to take advantage of the syntax to infer the semantics earlier? To maintain the advantage of using count, as opposed to mass, nouns, reported by Syrett (2013), while also querying the more accessible dimension of length, we might try adapting the "Who Said it Better?" method of Experiment 3. For example, we might show children a

chain of paperclips, and in one scenario, explain that each paperclip is two inches long (Each one is two inches long. This one is two inches, so is this one…"). They would then be asked to choose who describes the scene most appropriately, between one character using an attributive measure phrase ("I see two-inch paperclips"), and another using a pseudopartitive ("I see two inches of paperclips"). In another scenario, children would be told that the whole chain, from end to end, was two inches long, and be asked to choose between the same two utterances.

As discussed above, the present experiments suggest that children may not immediately appreciate the mappings between syntax and semantics in expressions of quantity. Children in previous tests of their grasp of this relationship may have relied on their understanding of "each" and "all together" to distinguish between arrays later described using attributive or pseudopartitive measure expressions. Due to the limited nature of the stimuli, however, the current studies leave open the possibility that children are in fact able to generalize semantic relationships from syntactic structure in this type of expression, but failed here due to the difficulty of the concept of weight, a prospect that merits further investigation. As it is, the present study provides evidence, across a variety of methodologies, that children do not begin to appreciate the distinct ways in which attributive and pseudopartitive measure expressions quantify weight until after the preschool years.

References

Barner, D., & Snedeker, J. (2005). Quantity judgments and individuation: Evidence that mass nouns count. *Cognition*, 97(1), 41– 66

Barner, D., & Snedeker, J. (2006). Children's early understanding of mass-count syntax: Individuation, lexical content, and the number asymmetry hypothesis. *Language Learning and Development, 2,* 163-194.

Brooks, N., Pogue, A., & Barner, D. (2011). Piecing together numerical language: Children's use of default units in early counting and quantification. *Developmental Science*, 14, 44-57.

Brown, M., Blondel, E., Simon, S., & Black, P. (1995). Progression in measuring. Research Paper in Education, 10(2), 143-170.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. *Proceedings of the 25th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Chierchia, G., Guasti, M. T., Gualmini, A., Meroni, L., Crain, S., & Foppolo, F. (2004). Semantic and pragmatic competence in children's and adults' comprehension of or. In I. A. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 283–300). New York, NY: Palgrave Macmillan.

Clements, D. H., & Sarama, J. (2009). *Teaching and learning early math: The learning trajectories approach*. New York, NY: Routledge.

Clark, H. (1970). The primitive nature of children's relational concepts. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 269 –278). New York, NY: Wiley.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science*, 10, 408–411.

Clearfield, M. W., & Mix, K. S. (2001). Amount versus number: Infants' use of area and contour length to discriminate small sets. *Journal of Cognition and Development*, 2(3), 243–260.

Corbett, G. G. (2000). *Number*. Cambridge: CUP.

Elkind, D. (1961). Children's discovery of the conservation of mass, weight, and volume: Piaget Replication Study II. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 98(2), 219-227.

Elkind, D. (1967). *Piaget's conservation problems*. Child Development, 38(1), 15-27.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object-files versus analog magnitudes. *Psychological Science,* 13, 150-156.

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent.

Cognitive Psychology, 44(1), 33–66.

Feigenson, L (2007). The equality of quantity. *Trends in Cognitive Sciences*, 11, 185-187.

Foppolo, F., Guasti, T. M. & Chierchia, G. (2012). Scalar Implicatures in Child Language: Give Children a Chance. *Language Learning and Development*, 8, 365–394.

Gal'perin, P. Y., & Georgiev, L. S. (1969). The formation of elementary mathematical notions. *Soviet studies in the psychology of learning and teaching mathematics*, 1, 189–216.

Gathercole, V. C. (1985). More and more and more about "more." *Journal of Experimental Child Psychology*, 40, 73–104.

Giralt, N. & Bloom, P. (2000). How special are objects? Children's reasoning about objects, parts, and holes. *Psychological Science*, *11*, 497-501.

Gulkoa, J., Doylea, A. B., Serbina, L. A., & Whitea, D. R. (1988). Conservation skills: A replicated study of order of acquisition across tasks. *The Journal of Genetic Psychology: Research and Theory of Human Development*, 149(4), 425-439.

Huntley-Fenner, G. N. (1995). The effect of the whole object bias on preschoolers' understanding of collective nouns. Proceedings of the Twenty-seventh Child Language Research Forum, Stanford, CA.

Huntley-Fenner, G. (2001). Why count stuff? Young preschoolers do not use number for measurement in continuous dimensions. *Developmental Science*, 4: 456–462.

Li, P. Cheung, K. Aguayo, S. Carey (2014, Nov.) The Role of Language in Object Individuation and Identification: Insights from the Acquisition of Partitive Expressions. Poster presented at the 39th Boston University Child Language Development Conference.

Lipton, J. S., & Spelke, E. S. (2006). Preschool children master the logic of number word meanings. *Cognition*, 98, B57-B66.

Nunes, T., & Bryant, P. (1996). *Children doing mathematics*. Oxford, UK: Blackwell Publishers

Piaget, J. (1965). *The child's conception of number*. New York: W. Norton Company & Inc.

Piaget, J., Inhelder, B., & Szeminska, A. (1960). *The child's conception of geometry*. New York, NY: Basic Books

Piaget, J. & Inhelder, B. (1974). The child's construction of quantities. London: Routledge and Kegan Paul.

Sarama, J., Clements, D. H., Barrett, J., Van Dine, D. W., & McDonel, J. S. (2011). Evaluation of a learning trajectory for length in the early years. *ZDM*, 43, 667–680.

Schrauf, C., Call, J., & Pauen, S. (2011). The Effect of Plausible Versus Implausible Balance Scale Feedback on the Expectancies of 3- to-4-Year-Old Children. Journal of Cognition and Development, 12(4), 518-536.

Shipley, E., & Shepperson, B. (1990). Countable entities: Developmental changes. *Cognition, 34,* 109-136.

Silverstein, S. (1974). Smart. In *Where the sidewalk ends*, p. 35.

Smith, C., Carey, S. & Wiser, M. (1985). On differentiation: A case study of the development of concepts of size, weight, and density. *Cognition*, 21(3), 177-237.

Sophian, C., & Kalihiwa, C. (1998). Units of counting: Developmental changes. *Cognitive Development*, 13, 561-585.

Syrett, K. (2013). The role of cardinality in the interpretation of measurement expressions. Language Acquisition: A Journal of Developmental Linguistics, 20, 228-240.

Syrett, K. (2010). The representation and processing of measure phrases by four-year-olds. In K. Franich, K. Iserman, & L. Keil (Eds.), Proceedings of the 34th Annual Boston University Conference on Language Development (BUCLD) (pp. 421-432). Somerville, MA: Cascadilla Press.

Wagner, L. and Carey, S. (2003). Individuation of Objects and Events: A Developmental Study. Cognition, 90, 163-191.

Wynn, K. (1992). Children's acquisition of number words and the counting system. *Cognitive Psychology*, 24, 220-251.

Xu, F. (2007). Sortal concepts, object individuation, and language. *Trends in Cognitive Sciences*, 11, 400-406.