

Published in final edited form as:

Cognition. 2004 July ; 92(3): 329–352. doi:10.1016/j.cognition.2003.10.001.

Six does not just mean *a lot*: preschoolers see number words as specific

Barbara W. Sarnecka* and Susan A. Gelman
University of Michigan, Ann Arbor, MI, USA

Abstract

This paper examines what children believe about unmapped number words – those number words whose exact meanings children have not yet learned. In Study 1, 31 children (ages 2–10 to 4–2) judged that the application of *five* and *six* changes when numerosity changes, although they did not know that equal sets must have the same number word. In Study 2, 15 children (ages 2–5 to 3–6) judged that *six* plus more is no longer *six*, but that *a lot* plus more is still *a lot*. Findings support the hypothesis that children treat number words as referring to specific, unique numerosities even before they know exactly which numerosity each word refers to.

Keywords

Language development; Number words; Numbers; Children; Concepts; Cognitive development

1. Introduction

“I like some *plenty*! I like some *too much*! I like some *lot*! I like some *eight*!” This was the protest of Teddy (age 2.5) upon receiving a small (clearly too small) amount of M&Ms. Teddy’s use of *eight* was ambiguous. On the one hand, maybe he was using *eight* as an approximate quantifier, like *plenty*, *too much*, and *a lot*. On the other hand, maybe he knew that *eight* was a specific number, and thought that *eight* M&Ms would be more than he had in his hand.

Researchers have long studied number-word learning as a window on the development of number concepts. Gelman and Gallistel (1978) argued that even 2- and 3-year-olds use number words in systematic ways, and that children’s use of these words can tell us a lot about their numerical thinking. Sometimes, though, it is hard to know what children mean by the number words they use. Children Teddy’s age use number words whose exact meanings they do not know. But these *unmapped* number words are not completely without meaning – in the example above, Teddy used the word *eight* in reference to a quantity, indicating that he knew something about the word. The question is, what?

1.1. Early uses of number words

By the time they turn 3, most children can ‘count’ (and we use the term loosely) rows of about five objects – that is, they can say the number words up to *five* while pointing to one object at a time (Fuson, 1988; Gelman & Gallistel, 1978; Wynn, 1992). However, they do not use counting to determine the number of things in a set (its *numerosity*). We know this

because when children are asked, immediately after counting, *how many* objects the row has, they do not answer with the last word they used in counting. Instead they may give another number word, or several number words, or they may interpret the question as a prompt to ‘count’ the row again (Fuson, 1988; Wynn, 1990). When young 3-year-olds are shown a set of jars and lids, and are asked whether there are enough lids for every jar to have a lid, most children do not even try to count the jars or the lids in order to answer the question (Michie, 1984; Mix, 1999a; see also Mix, Huttenlocher, & Levine, 1996; Sophian & Adams, 1987). Neither do they use counting when asked to hand over a certain number of objects (the *Give-A-Number* task). Instead, most children simply grab a handful (Wynn, 1990; see also Schaeffer, Eggleston, & Scott, 1974; Sophian, 1987).

In two studies using the Give-A-Number task, Karen Wynn (1990, 1992) found that by 2.5 years of age, children give 1 item when asked for *one*, and multiple items when asked for *two*, *three*, *four*, *five*, or *six*, although they do not distinguish among these larger numbers. We will call this period in development Performance Level I. Around 3–3.5 years of age, children differentiate *two* from the other words. So they give 1 for *one*, 2 for *two*, and a handful (but not 1 or 2) of 3 or more for *three*, *four*, *five*, and *six*. We will call this Performance Level II. At the next level (Performance Level III), children give 1 for *one*, 2 for *two*, 3 for *three*, and a handful (but not 1, 2, or 3) for other number words. Finally, around 3.5–4 years of age, children seem to experience a sort of epiphany. Suddenly, the relationship of counting to numerosity (also called the *cardinal principle* or *cardinality principle*) is understood. At this point, which we will call Performance Level IV, the child can use counting to generate sets of *four*, *five*, *six*, and so on.

1.2. Specific/Unique Numerosities view of number-word learning

What can we conclude from the pattern described above? Wynn proposes an interpretation (Bloom & Wynn, 1997; Freeman, Antonucci, & Lewis, 2000; see also Whalen, Gallistel, & Gelman, 1999; Wynn, 1992) that we will call the Specific/Unique Numerosities (SUN) view. In this view, the important thing to notice is that children contrast mapped with unmapped number words so early on. After all, as early as Level I, children seem to realize that the words *two*, *three*, *four*, *five*, and *six*, whatever they mean, cannot mean the same thing as *one*. We know this because they never give 1 item when asked for another number word. Children at Levels II and III also avoid giving 2 or 3 when asked for higher number words. In the SUN view, this pattern indicates that children already know that number words correspond to specific, unique numerosities. They know this even before they map each word to its numerosity.

One concern with the Give-A-Number task might be that it could underestimate children’s knowledge. A child might actually know the exact meaning of *five*, for example, but try to get a set of 5 by estimating rather than counting. However, children also contrast number words on another task (the *Point-to-X* task) where they need only point to the target numerosity (Wynn, 1992). Results on this task mirror those on the Give-A-Number task, implying that both tasks are valid indicators of the number-word meanings a child actually knows.

An objection to the SUN view might be that it is perfectly possible for children to contrast unmapped with mapped number words, and still remain agnostic about how the unmapped words relate to each other. For example, Level II children could consider *one* and *two* to be distinct from all other number words, while still viewing the other number words as equivalent. A more convincing demonstration of the SUN hypothesis would show that children distinguish *among* unmapped number words, in addition to contrasting the mapped with the unmapped. And in fact, there was a recent study (Condry, Cayton, & Spelke, 2002) investigating this question. In that study, children expected mapped number words, but not

unmapped number words, to change when numerosity changed. This finding supports what we call the Approximate Numerosities/Bootstrapping view (ANB).

1.3. ANB view of number-word learning

According to the ANB view, children believe that unmapped number words mean something like *a lot* (Carey, 2001; Spelke & Tsivkin, 2001). This idea makes perfect sense when we consider how numbers are perceived. From infancy, children recognize small, exact numerosities like 2 and 3 (Antell & Keating, 1983; Starkey & Cooper, 1980; Strauss & Curtis, 1981; Van Loosbroek & Smitsman, 1990). They also recognize large, approximate numerosities. For example, babies can distinguish between 8 and 16 objects, but not between 8 and 12 (Xu & Spelke, 2000). If children assumed that number words simply label the numbers we perceive, then it would make sense to have exact words for 1, 2, 3, and perhaps 4, but approximate words for bigger sets (*plenty, too much, a lot... eight?*). According to the ANB view, children do not necessarily conceive of the existence of *large, exact numerosities* until, having reached Level III, they notice that each successive counting word (e.g. from *one* to *two*, from *two* to *three*) corresponds to a numerosity increase of 1 (Carey, 2001). Language thus allows children to ‘bootstrap’ their way to representations of large, exact numbers – concepts that were inconceivable before.

1.4. The present studies

The present studies tried to find out whether young children think that unmapped number words refer to specific numerosities (as in the SUN view), or to approximate numerosities (as in the ANB view). Study 1 was designed to find out whether children were aware of two things:

1. *Changing the numerosity of a set will also change its number word; changes that do not affect the numerosity of a set will not affect the number word.* Prior research has shown that children as young as 2.5 know which transformations affect quantity (Gelman & Gallistel, 1978; McGarrigle & Donaldson, 1974; Mehler & Bever, 1967). Here, we try to learn whether children believe that these changes affect the number word as well.
2. *Sets of equal numerosity have the same number word; sets of unequal numerosity have different number words.* Prior research has shown that preschoolers can recognize numerical equivalence without counting (Huttenlocher, Jordan, & Levine, 1994; Mix, 1999a,b; Mix et al., 1996; Sophian & Adams, 1987). Here, we try to learn whether children believe that equal sets share the same number word.

2. Study 1

2.1. Method

2.1.1. Participants—Eighty-four children (41 girls, 43 boys) were screened for possible participation in the complete series of tasks. Of these, 25 were dropped because they did not yet know at least two number words (see Section 2.1.2.2); three children were dropped because they failed too many memory checks (see Sections 2.1.2.5 and 2.1.2.7 for more detail); and two children were dropped because they did not complete the tasks. This left a total of 54 children (26 girls, 28 boys) ranging in age from 2–10 to 4–1 (mean age 3–6).

All children were monolingual and native speakers of English. No questions were asked about socio-economic status, race, or ethnicity. However, participants were recruited from university-affiliated and private day-care centers serving primarily middle-class families, and were presumably representative of the Midwestern university community from which they were drawn.

2.1.2. Tasks

2.1.2.1. Give-A-Number task: The purpose of this task (based on Wynn, 1992) was to determine which number-word meanings each child knew. This information was used to determine performance levels for subsequent tasks. A pile of 15 erasers was placed in front of the child, who was asked to give a certain number of them to a puppet. Requests were of the form “Can you give *five* apples to the monkey? Just take *five* and put them right here on the table in front of him.” After responding, the child was asked a single follow-up question, of the form “Is that *five*?” which repeated the initial number word asked for. If a child responded “no” to the follow-up question, the original request was repeated.

Each child received 15 trials: five trials with each of three puppets (a monkey, a dog, and a snake). A different type of eraser was used for each puppet (apples, flowers, and teeth). *One* object was always requested first, then *two* and *three* in counterbalanced order, then *five* and *six* in counterbalanced order.

2.1.2.2. Sorting of children into performance levels: The Give-A-Number task formed the basis for the performance levels used in the rest of the analysis. Children were grouped according to the highest number word whose exact meaning they knew. Scoring criteria were taken from Wynn (1992): children were given credit for knowing a number word if they (a) gave the correct number of items for that word on at least two of the three trials, and (b) gave that number of items no more than once for any other word. For example, a child who gave 1 and 2 (but not 3) upon request was placed in Performance Level II.

With few exceptions, results followed the pattern reported by Wynn (1992). Children succeeded up to a certain number, and failed at all higher numbers. There were only four children who did not follow this pattern. Two of them correctly counted out sets of 5 and 6 but failed at one of the lower numbers. These children were placed in Level IV. The other two children succeeded at 1 and 2, failed at 3 and 5, but then happened to grab 6 (without counting) on the *six* trials. These children were placed in Level II.

2.1.2.3. Counting task: In order to keep the procedure as similar as possible to that used by Wynn (1990), children were also given a warm-up task in which they counted arrays of 2, 3, 5, and 6 objects glued to a board. The order of tasks was counterbalanced across participants.

2.1.2.4. Transform-Sets task: The purpose of this task was to learn what transformations children consider relevant to number words. Materials included a metal box with hinged lid (approximately 14 cm × 9 cm × 6 cm) and rubber erasers (moons, soccer balls, cars, and eyeballs). The experimenter placed the box on the table and said, “The way we play this game is, I will put something inside the box, and you try to remember what’s in there.” On the warm-up trials, the experimenter placed 1 eraser inside the box, saying for example “Here’s a moon”. The experimenter then closed the lid and asked the child “OK, what’s in the box?” After the child answered, the experimenter either shook the box vigorously, rotated it 360 degrees on the table, or removed the eraser (right in front of the child) and replaced it with one of a different shape. After completing the action, the experimenter said, “OK, *now* what’s in the box?”

After the four warm-up trials, the experimenter said, “Now you will guess *how many* things are in the box.” As the child watched, the experimenter then placed 2, 3, 5 or 6 identical erasers in the box, telling the child the correct number word (e.g. “Here are *five* moons.”) Then the experimenter closed the lid and asked the memory-check question – “How many moons?” No correction was offered on the memory-check question; trials where the child answered incorrectly were excluded from the analysis.

After the memory-check question, the experimenter ‘transformed’ the set in one of four ways: (a) by picking up the (closed) box and shaking it vigorously; (b) by rotating the (closed) box 360 degrees on the table; (c) by opening the box and adding 1 eraser to it; or (d) by opening the box and removing 1 eraser from it. (An eraser was always added to sets of 2 or 5; an eraser was always removed from sets of 3 or 6.) Finally, the experimenter asked the test question – “Now how many moons – is it *five* or *six*?” (For trials with initial sets of 2 or 3, the question was “... is it *two* or *three*?”.) The form of the questions was kept the same, so the correct choice was offered first exactly half of the time.

Each child received eight trials: two trials starting with *two* objects, two trials starting with *three* objects, two trials starting with *five* objects, and two trials starting with *six* objects. One of the trials for each number word involved a numerosity-irrelevant transformation (shaking or rotating the box), and the other involved a numerosity-relevant transformation (adding or subtracting an item). The order of trials was randomized across participants.

For scoring, the eight trials were collapsed to produce four measures: *low-same*, *low-different*, *high-same*, and *high-different*. The *low-same* measure was the composite score from trials where initial sets of 2 and 3 were shaken or rotated. The *low-different* measure was the composite score from trials where initial sets of 2 and 3 gained or lost an object. The *high-same* measure was the composite score from trials where initial sets of 5 and 6 were shaken or rotated. The *high-different* measure was the composite score from trials where initial sets of 5 and 6 gained or lost an object.

Scores reflected whether or not the child still gave the initial number word after the set was transformed (rather than being scored correct/incorrect). Each trial was assigned a score of 1 if the child had answered the test question (“OK, *now* how many moons?”) with the same number word as the trial began with. If the child gave a different number word after the transformation, a score of 0 was assigned. If children see number words as specific, then they should give the initial number word more often on *same* trials (where the box was merely shaken or rotated) than on *different* trials (where an item was gained or lost).

2.1.2.5. Transform-Sets memory-check criterion: Two children failed the memory-check question on at least half the trials. These children’s data were dropped from the analysis (see Section 2.1.1). Of the other children’s data, 395 trials (95%) had valid memory-checks and were included in the analysis; 21 trials (5%) had failed memory checks and were excluded. (Thirty-eight children missed no trials; 14 children contributed to the trials that had to be dropped. Only three children missed more than two trials.) When a trial was excluded, the composite measure (e.g. *low-same*) was computed by doubling the score of the other, valid trial in the measure. If a child missed both memory checks of a given pair, no data were entered for that composite score (this happened only once).

2.1.2.6. Compare-Sets task: The purpose of this task was to learn whether children expect numerically equal sets to have the same number word. Materials included two puppets (a frog and a lion) and pictures representing snacks, approximately 20 cm × 6 cm. Each picture showed a row of identical food items (peaches, brownies, hamburgers, etc.) framed by a dark border. In half of the trials, Frog’s snack (picture) was identical to Lion’s. We will call these *same* trials. In the remaining trials, Frog’s snack (picture) differed from Lion’s snack (picture) by one item. We will call these *different* trials. On the *different* trials, a large, empty circle was drawn around the blank space (always at the end of a row) where an item was missing from one of the pictures. It was easy to see that one snack had more than the other. The experimenter introduced the task by saying

Now I want to tell you a story about when Frog and Lion came to my house and I gave them some snacks. I tried to make their snacks just exactly the same, because

they like their snacks to be the same. But sometimes I made a mistake, and their snacks were not the same. Like sometimes I gave more to Lion, or I gave more to Frog. So I want to show you the snacks I gave them, and you can tell me if their snacks are just the same, or if I made a mistake.

Before each trial, the experimenter said: “The next snack I gave them was [strawberries]. This is Frog’s snack...” (placing the picture of 5 strawberries in front of Frog) “...and this is Lion’s snack...” (placing the picture of 6 strawberries in front of Lion). The experimenter then lined up the pictures one above the other, so that they were easily compared and asked, “Are their snacks just the same, or did I make a mistake?”

If the child responded correctly, the experimenter offered affirmation (“That’s right, I made a mistake, didn’t I? Can you see the empty place where I forgot to put a strawberry? Yes, there it is!”). If the child responded incorrectly, the experimenter offered correction (“Hmm. They *are* the same *kind* of snack, but... oh, no! I think I forgot to put a strawberry on Frog’s plate! Can you see the empty place where I forgot to put a strawberry? Yes, there it is!”).

The test trials were the same as the warm-up trials, except that after the child had answered the control question (and been corrected if necessary) the experimenter removed both pictures and asked the test question – “Frog had *five* strawberries. Do you think Lion had *five* or *six*?” On the low-number trials, the question was, “Do you think Lion had *two* or *three*?” The form of the question was always the same, so the first choice was correct exactly half the time. Finally, with the pictures still hidden from view, the experimenter asked the memory-check question, “And were their snacks just the same, or did I make a mistake?” No correction was offered on the memory-check question; trials where the child answered incorrectly were excluded from the analysis.

Each child received eight trials: two trials where Frog’s snack had *two* objects, two trials where Frog’s snack had *three* objects, two trials where Frog’s snack had *five* objects, and two trials where Frog’s snack had *six* objects. In half these trials (which we call *same* trials), Lion’s snack was identical to Frog’s. In the other half (which we call *different* trials), Lion’s snack differed from Frog’s by 1 item. The order of trials was randomized across participants.

For scoring, the eight trials were collapsed to produce four measures: *low-same*, *low-different*, *high-same*, and *high-different*. The *low-same* measure was the composite score from trials where Frog and Lion both had 2 items, or both had 3 items. The *low-different* measure was the composite score from trials where Frog had 2 and Lion had 3, or vice versa. The *high-same* measure was the composite score from trials where Frog and Lion both had 5 items, or both had 6 items. The *high-different* measure was the composite score from trials where Frog had 5 and Lion had 6, or vice versa.

Scores reflected whether or not the child gave the same number word for Lion’s set as the experimenter had given for Frog’s set (rather than being scored correct/incorrect). Each trial was assigned a score of 1 if the child had answered the test question (“Frog had *five* strawberries. How many did Lion have?”) with the same number word as Frog’s. If a child gave a different number word for Lion’s snack, a score of 0 was assigned. If children see number words as specific, then they should give the same number word more often on *same* trials (where the snacks were identical) than on *different* trials (where the snacks differed by 1).

2.1.2.7. Compare-Sets memory-check criterion: One child failed the memory-check question on half of her Compare-Sets trials. This child’s data were dropped from the analysis (see Section 2.1.1). Of the other children’s data, 400 trials (94%) had valid

memory-checks and were included in the analysis; 24 trials (6%) had failed memory checks and were excluded. (Thirty-eight children missed no trials; 16 children contributed to the trials that had to be dropped. Only one child missed more than two trials.) When a trial was excluded, the composite measure (e.g. *low-same*) was computed by doubling the score of the other, valid trial in the measure. If a child missed both memory checks of a given pair, no data were entered for that composite score (this happened three times).

2.1.3. Procedure—Data were collected in two sessions, no more than a week apart (mean 3 days apart). Subjects who were tested twice in the same day got the first session in the morning and the second session in the afternoon. Testing was done in a quiet room at the child’s school by a female experimenter. The first session included the Give-A-Number and Counting tasks; the second session included the Transform-Sets and Compare-Sets tasks. The order of tasks within the session was counterbalanced across participants.

2.2. Results

Performance Levels II and III were merged in the following analysis ($n = 31$, mean age 3–5, range 2–10 to 4-1). A separate analysis showed no significant difference between Levels II and III on any measure. Unlike Wynn (1992), we did have a few children ($n = 7$) who used counting to make sets of 5, but did not quite manage to do 6. In some cases, the problem was clearly procedural – the children lost track of which items they had already counted, etc. Several of the other children seemed to believe that six was a huge number. One child said, “Woo-hoo! Well I’m not a octopus, so I may not do too good on this.” These children typically gave all 15 items when asked for *six*. As another child reflected, “What is six? It’s a lot. It means all of them. One, two, three, four, five, six, seven, eight. It’s a big, long line of them.”

These children, along with children who succeeded at both numerosities 5 and 6, were placed in Performance Level IV ($n = 23$, mean age 3–8, range 2–11 to 4-1). A separate analysis showed no significant difference between the children who succeeded at 5 but not 6, and those who succeeded at both 5 and 6, on any measure.

2.2.1. Transform-Sets task—Trials were scored 1 or 0 according to whether or not the child repeated the original number word after the set was transformed. If the SUN view is correct and children do see number words as specific, then the scores should be higher for *same* trials (where the box was merely shaken or rotated) than for *different* trials (where an item was gained or lost). If the ANB view is correct, children should be equally likely to repeat the initial number word, regardless of the type of transformation (at-chance performance would mean a score of 1 on both *same* and *different* trials).

Fig. 1 illustrates the results. The upper portion of the figure shows the response patterns that would be predicted by the SUN and ANB views, respectively. Note that the predictions differ only for Level II–III, and only on high-number trials (trials involving the words *five* and *six*). Predictions do not differ for low-number trials (represented by the solid black and confetti-patterned bars in each cluster), since these trials involve the already-mapped number words *two* and *three*. Predictions also do not differ for Level IV children (not pictured in the upper portion of the figure), who have mapped the exact meanings of the high-number words and presumably should do well at the task.

The lower portion of the figure shows the actual responses of children at each performance level. As expected, all children did well on the low-number trials, and Level IV children did well on both low- and high-number trials. However, the Level II–III children also did well on the high-number trials. As predicted by the SUN view, they kept the same number word

for sets that had been shaken or rotated (solid white bar), but used a different number word for sets that had gained or lost an item (diagonal-patterned bar).

A 2 (level: II–III versus IV) \times 2 (number: low versus high) \times 2 (concordance: same versus different) repeated measures ANOVA revealed a main effect of concordance on percentage of same-number-word responses ($F(1, 51) = 124.61, P < 0.001$). This reflects the fact that, on the whole, the children as a group were more likely to say that the number word of a set changed when an item was gained or lost than when the set was merely shaken or rotated. There was no main effect for number, and no significant interactions.

Based on the original hypothesis, we also conducted a set of planned comparisons, focusing on *low-same* versus *low-different* and *high-same* versus *high-different* measures within subjects, by performance level. As expected, all children repeated the initial number word significantly more often on *low-same* trials than on *low-different* trials ($P_s < 0.001$ by paired *t*-test). Most importantly, the same was true for the measures *high-same* versus *high-different* ($P_s < 0.001$ by paired *t*-test). A separate analysis of Level II found the same results even for this youngest group ($P_s < 0.01$ by paired *t*-test). The mean of each performance level for each measure was also significantly different from chance (1.0) ($P_s < 0.05$ by one-sample *t*-test), with one exception: the *low-different* measure for Level II children alone was not significantly different from chance ($P = 0.16$).

2.2.2. Compare-Sets task—Some children initially interpreted the question “Are their snacks just the same?” as asking about snack *type* rather than snack *numerosity*, making comments like “Yes, they are the same, and Frog has more” (see Karmiloff-Smith, 1977 for a related finding). The warm-up trials were effective in clearing up this confusion; on the first warm-up trial 30 children (56%) answered incorrectly that the discrepant snacks were the same; on the first test trial, only one child (0.02%) made this error. A separate analysis found essentially no differences between the children who initially made this error and those who did not (see *Compare-Sets task* under *Results*, below, Section 2.2.2.).

Trials were scored 1 or 0 according to whether or not the child gave the same number word for Lion’s snack as the experimenter had given for Frog’s snack. If the SUN view is correct and children do see number words as specific, then the scores should be higher for *same* trials (where the snacks were identical) than for *different* trials (where the snacks differed). If the ANB view is correct, children should be equally likely to give the same number word, regardless of the type of trial (chance would be a score of 1 on both *same* and *different* trials).

Fig. 2 illustrates the results. The upper portion of the figure shows the response patterns that would be predicted by the SUN and ANB views, respectively. Note that the predictions differ only for Level II–III, and only on high-number trials (trials involving the words *five* and *six*). Predictions do not differ for low-number trials (represented by the solid black and confetti-patterned bars in each cluster), since these trials involve the already-mapped number words *two* and *three*. Predictions also do not differ for Level IV children (not pictured in the upper portion of the figure), who have mapped the exact meanings of the high-number words and presumably should do well at the task.

The lower portion of the figure shows the actual responses of children at each performance level. As expected, all children did well on the low-number trials, and Level IV children did well on both low- and high-number trials. On this task, however, Level II–III children performed at chance on the high-number trials. As predicted by the ANB view, they were not more likely to apply the same number word to two identical sets (solid white bar) than to two visibly different sets (diagonal-patterned bar).

A 2 (level: II–III versus IV) \times 2 (concordance: same versus different) \times 2 (number: low versus high) repeated measures ANOVA revealed a main effect of concordance on percentage of same-number-word responses ($F(1, 49) = 45.38, P < 0.001$). This reflects the fact that, on the whole, the children as a group were more likely to apply the same number word to two identical sets than to two visibly different sets. There was also a Concordance \times Number interaction ($F(1, 49) = 12.55, P < 0.001$), reflecting children's greater success on low-number trials (involving the numbers 2 and 3) than on high-number trials (involving the numbers 5 and 6). Finally, there was a Concordance \times Level interaction ($F(1, 49) = 8.26, P < 0.01$), reflecting the fact that Level IV children performed better than Level II–III children overall. There was no main effect of number, and no significant interactions of Number \times Level or Concordance \times Number \times Level.

Based on the original hypothesis, we also conducted a set of planned comparisons, focusing on *low-same* versus *low-different* and *high-same* versus *high-different* measures within subjects, by performance level. As expected, all children were significantly more likely to give the same number word on *low-same* trials than on *low-different* trials ($P_s < 0.001$ by paired *t*-test). On the high-number trials, however, only Level IV children, who knew the exact meanings of *five* and/or *six*, predicted that identical sets should have the same number word and that differing sets should not ($t(20) = 4.93, P < 0.001$). Level II–III children were not more likely to apply the same number word to identical sets than to differing sets ($P = 0.48$). The mean responses of Level IV children were also significantly different from chance (0.1) on both measures ($P_s < 0.01$), whereas Level II–III children responded at chance.

A separate analysis compared the performance of children who misunderstood the control question on the first warm-up trial as asking about snack *type* rather than snack *numerosity* with the performance of children who had not made this error. This analysis found no differences in the performance of children at Level II–III. At Level IV, there was no difference found for three of the measures (low-same, low-different, or high-different). On the *high-same* measure, children who initially made the error ($n = 9$) actually performed better than the other children ($t(20) = -2.75, P = 0.01$).

2.3. Discussion

What do these findings tell us about how children use number words? In particular, do children expect unmapped number words to refer to specific, unique numerosities? Apparently, it depends on the task. Both the Transform-Sets task and the Compare-Sets task had significant results, but in opposite directions. The Transform-Sets results suggest that children do treat unmapped number words as unique and specific. The fact that children who did not know the meanings of *five* and *six* nonetheless expected these words to change with the addition or loss of one item is consistent with what we have termed the SUN view. On the other hand, the Compare-Sets results are more consistent with what we have termed the ANB view. On the Compare-Sets task, children who did not know the meanings of *five* and *six* applied the same word to unequal sets, and different words to identical sets.

The Compare-Sets findings are consistent with those of Condry et al. (2002). In that study, children at Level II did not judge that the words *four* and *eight* should change with numerosity. However, that study also found that Level II children could not recall the initial number word even in the absence of any transformation, a question equivalent to the memory-check question used in our Transform-Sets task. Level II children in the present study answered the question correctly on 109 of 120 trials (89%). In other words, they showed no evidence of the difficulty described by Condry et al. This leads us to believe that the Condry et al. task may simply have been more difficult than ours. For example, 2- and 3-

year-olds may have found the forced-choice format of the present study easier than the yes/no format of the Condry et al. task.

Returning to the present study, there appear to be two possibilities:

1. the Transform-Sets task is a valid measure of children's SUN knowledge, whereas the Compare-Sets task underestimates it, or
2. the Compare-Sets task is a valid measure of children's SUN knowledge, whereas the Transform-Sets tasks overestimates it.

Let us consider each of these possibilities separately. The first possibility is that children at Levels II and III do understand the SUN principle as it is described here, but nevertheless responded at chance on the high-number Compare-Sets trials. This is really two questions: (1) If children know the SUN principle, why didn't they apply it on the high-number trials; and (2) If they didn't apply it on the high-number trials, why did they apply it on the low-number trials?

Question (2) is easy to answer. The children did not need to apply the SUN principle on the low-number trials. It was possible to succeed on these trials just by recalling the numerosity of each animal's snack. As we know from infant research, the numerosities 2 and 3 can be identified without counting. Thus, the question

Frog had three peaches – do you think Lion had two or three? could be interpreted as, simply

Do you think Lion had two or three peaches? The child could answer this question by recalling what Lion's snack looked like, without thinking about Frog's snack at all.

Question (1) is more puzzling. Of course a task may simply be too difficult, but the memory-check criterion verified that the children were paying attention and remembered whether the snacks were the same or not. The Level III children, who performed at chance on the high-number Compare-Sets trials, were not much younger than the Level IV children, who performed above chance (average age 3–7 versus 3–8). Nor were they much less proficient at counting, at least at the numerosities tested here (Level III children were able to count sets of 5.6 on average, versus 5.8 for the Level IV children). Neither of these differences even approached statistical significance. So the results probably don't reflect greater ability of the Level IV children overall, but rather some specific developmental change related to number. Is that change the acquisition of the SUN principle?

If so, we have arrived at possibility (b) – that Level II–III children do not apply a SUN principle to higher number words, but were somehow able to succeed on the Transform-Sets task anyway. In fact, there is a way they might do so. The task contrasted quantitative changes (adding or subtracting 1 object) with non-quantitative changes (shaking or rotating). Perhaps children could succeed on this task just by knowing that the words *five* and *six* have something to do with quantity. They could reason that when quantity changed, the number word changed. When quantity stayed the same, the number word stayed the same. Perhaps children would expect *any* quantifier to change under these circumstances, not just number words. What is needed is a test that determines whether children distinguish number words from other quantifiers like *some* or *a lot*. Although both types of words relate to quantity, the SUN principle only applies to number words.

Study 2 looked at children's use of *six* and *a lot*. We reasoned that if they treat these words as synonymous, then the findings from the Transform-Sets task could be reinterpreted and it would be reasonable to conclude that children do not have a SUN assumption. If, on the

other hand, children do not treat *six* and *a lot* as synonymous, then it seems reasonable to conclude that children do have a SUN assumption about number words, but that the Compare-Sets task for some reason failed to elicit it.

3. Study 2

A new task directly assessed whether children who do not yet know the exact meaning of *six* consider it synonymous with *a lot*. We reasoned that if children view *six* as specific and unique, then a set of *six* should no longer have *six* after more items are added. If (as we speculate above) the Transform-Sets task taps into children's treatment of quantifiers in general, rather than number words in particular, then children should also expect the label *a lot* to change with the addition of more items to a set. On the other hand, if children expect only *six* and not *a lot* to change when quantity changes, then it would seem that they really do treat number words differently. This would allay our concerns about the Transform-Sets task.

Besides helping us interpret the results of earlier tasks, we hoped that this task would allow us to include Level I children. (Level I children had been dropped from Part 1 after early testing showed that most of them failed the memory-check questions on the Transform-Sets and Compare-Sets tasks.) In Part 2, children were first sorted into performance levels, and then given a novel task contrasting the terms *six* and *a lot*.

3.1. Method

3.1.1. Participants—Participants included 17 children (ten girls, seven boys) ages 2–7 to 3–6 (mean 3-1), recruited from the same preschools as in Part 1. Based on the Abbreviated Give-a-Number task, these children were sorted into performance levels: Level I children ($n = 7$, mean age 2–10, range 2–7 to 3-3), Level II children ($n = 7$, mean age 3-2, range 2–10 to 3–5), and Level III children ($n = 3$, mean age 3–4, range 3-1 to 3–6). Another two children tested were at Level IV; their data were dropped from the analysis.

3.1.2. Tasks

3.1.2.1. Abbreviated Give-A-Number task: The purpose of this task was to sort children into performance levels. Materials included the monkey puppet and erasers from Part 1. Children were asked for *one*, *three*, and *six* erasers; often this was enough to determine their performance level. (For example, a child who gave 1 for *one*, 3 for *three*, and 6 for *six* (by counting) was placed in Level IV. On the other hand, a child who gave 1 for *one*, 2 for *three*, and 2 for *six* was placed in Level I.) If necessary, children were also asked for *two* and *five* erasers (e.g. a child who succeeded at *one* and failed at *three* would be asked for *two*; a child who grabbed 6 for *six* without counting would be asked for *five*). Children might also be asked for any or all of the number words a second time (e.g. a child who gave 1 for *one*, 3 for *three*, 3 for *six*, 2 for *two*, and grabbed 5 for *five* without counting was asked for *three*, *five* and *six* again). In order to succeed at a given number word, children had to give the correct numerosity for that number word each time they were asked (once or twice) and also refrain from ever giving that numerosity for another number word.

3.1.2.2. Six-Versus-A-Lot task: The purpose of this task was to determine whether children at Levels I–III treat *six* and *a lot* as synonymous. Materials included two plastic bowls, approximately 13 cm in diameter and 7 cm deep, and ten clear plastic containers of small objects (e.g. pennies, see Appendix A). There were two warm-up trials, followed by eight test trials. For the warm-up trials, the empty bowls were placed in front of the child, who was told

The way we play this game is, I'm going to put *one* penny in here (experimenter places 1 penny in a bowl) and *one* penny in here (experimenter places 1 penny in the other bowl). All right. So this bowl has one penny in it, and this bowl has one penny in it. And here are some more pennies (experimenter pours all the remaining pennies into one of the bowls). Okay, now I'm going to ask you a question about *one* penny. Which bowl has *one* penny?

The other warm-up trial was the same except for the final question, which was "Okay, now I'm going to ask you a question about *some* pennies. Which bowl has *some* pennies?" The order of warm-up trials was counterbalanced across participants. The purposes of these warm-up trials were (a) to show that either bowl (the one containing additional objects or the one that remained unchanged) could be the correct choice, and (b) to check that the child understood basic quantifiers. Children who did not contrast *one* with *some* on the warm-up trials were excluded from the analysis (five children were excluded for this reason).

The test trials were similar to the warm-up trials, except that the experimenter started by placing six objects in each bowl. The test trials were of two types: *six* trials and *a lot* trials. On the *six* trials, the child was told (for example)

I'm going to put *six* pennies in here (experimenter places 6 pennies in a bowl)... and *six* pennies in here (experimenter places 6 pennies in the other bowl). All right. So this bowl has six pennies, and this bowl has six pennies. And here are some more pennies (experimenter pours all the remaining pennies into one of the bowls). OK, now I'm going to ask you a question about *six* pennies. Which bowl has *six* pennies?

On the *a lot* trials, the child was told

I'm going to put *a lot* of pennies in here (experimenter places 6 pennies in a bowl) ... and *a lot* of pennies in here (experimenter places 6 pennies in the other bowl). All right. So this bowl has *a lot* of pennies, and this bowl has *a lot* of pennies. And here are some more pennies (experimenter pours all the remaining pennies into one of the bowls). OK, now I'm going to ask you a question about *a lot* of pennies. Which bowl has *a lot* of pennies?

Trials were blocked so that each child completed four of one type and then four of the other. The order of blocks was counterbalanced across participants. To maintain their interest, children were allowed to pour the objects back into their original container after each trial, and to choose the objects for the next trial.

3.2. Results and discussion

Because there were too few children at each performance level to compare levels systematically, all participants' data were analyzed together. Each trial received a score of 1 if the child pointed to the bowl that had more objects added, or 0 if the child pointed to the bowl that remained unchanged (i.e. still contained 6 objects). Fig. 3 illustrates the results. Overall, children did not apply the word *six* to a set of *six-plus-more* items (black bar). However, they did apply the phrase *a lot* to a set of *a-lot-plus-more* items (white bar). These response patterns were significantly different from each other ($t(16) = -7.03, P < 0.001$), and significantly different from chance (2.0) ($P_s < 0.05$). A separate analysis of Level I children alone showed that even in this youngest group, the response patterns were still significantly different – both from each other (paired $t(6) = -20.14, P < 0.001$) and from chance (2.0) ($P_s < 0.001$).

One concern was that children's answers might be influenced by a pragmatic contrast between *six* and *a lot*. For example, a child who received the *a lot* trials first might, when

presented with the *six* trials, choose the less-full bowl – not because she saw *six* itself as specific, but simply because a different word would prompt a different response. To control for this possibility, a separate analysis was conducted of the first block of trials only. This analysis found even stronger results in the predicted direction, independent samples ($t(16) = -9.62, P < 0.001$). Children’s application of *six* to the unchanged numerosity was clearly not prompted by a pragmatic contrast with *a lot*.

Thus, it appears that children do not treat *six* and *a lot* as synonymous, even at Level I. A possible objection to the Six-Versus-A-Lot task is that since a large number of objects are added, it leaves open the possibility that *six* could still be considered an approximate quantifier of relatively small sets. For example, if children considered *six* to mean approximately 4–8 objects (but not 50), they might say that the full bowl did not contain *six*, without thinking that *six* is a specific number. However, this seems unlikely because the Transform-Sets task described in Part 1 of this study found the same results when sets changed by only 1 item.

4. General discussion

4.1. On balance, results favor the SUN view

We now have results from three tasks designed to find out whether children expect unmapped number words to refer to specific, unique numerosities. On the Transform-Sets task, children who knew the exact meanings of the words *one* and *two* only (Level II) or of *one*, *two*, and *three* only (Level III) judged that the application of *five* and *six* changes when a set gains or loses even a single item. They also judged that the application of *five* and *six* does not change when a set is merely shaken up or turned around. On the Compare-Sets task however, Level II–III children were willing to apply the same unmapped number word (*five* or *six*) to two unequal sets, and to apply different words to identical sets.

The Six-Versus-A-Lot task attempted to resolve this conflict by testing a possible flaw in the Transform-Sets task, namely that it might tap into children’s knowledge of quantifiers in general, rather than number words in particular. On this task, however, children did not treat the non-numerical quantifier *a lot* as specific – only the number word *six*. This leads us to believe that children do indeed view unmapped number words (unlike other quantifiers) as referring to specific, unique numerosities, even before they know exactly which numerosity each word refers to. In terms of the views of number-word learning described in the introduction to this paper (Section 1), we believe that, despite the contradictory findings of the Compare-Sets task, our results tend to support the SUN view, rather than the ANB view.

This interpretation also accords well with our anecdotal experiences of testing. A great many children (perhaps half of all children tested) initially responded to requests like *Can you give six apples to the monkey* by saying “I don’t know how to do six.” (The experimenter’s reply in these cases was “That’s okay, you can just do your best.”) After such an exchange, at least one child responded to the follow-up question (*Is that six?*) with an optimistic “Maybe!” Such behavior is consistent with a view of number words as specific.

4.2. How to explain the Compare-Sets results?

We are thus left with the question of how to interpret the Compare-Sets results. If children do have a SUN assumption, why did they not use it to perform this task? The Compare-Sets memory-check criterion ensured that the children were paying attention and remembered whether the sets were ‘the same’ or not. The Level IV children (who did very well on the task) were not significantly older or more proficient at counting than the Level III children (who performed at chance), so we assume that the Level IV children’s success does somehow derive from having mapped the exact number-word meanings. But if we want to

argue in favor of the SUN view that Level II–III children *already* view unmapped number words as referring to exact numerosities, then how does mapping the words to their meanings help children succeed on the Compare-Sets task? We discuss several possible answers below.

4.2.1. Unmapped number words might tax short-term memory—Even if children understand that unmapped number words refer to unique numerosities, mapping each word to its numerosity could make the whole package easier to think about. A similar phenomenon has been observed with regard to language development at the community level. Hunt and Banaji (1988) observed that over a period of decades, Californian surfers have developed a vocabulary for describing waves, including the terms *hollow* and *flat*. Presumably, an earlier generation of surfers would have described the same waves using sentences. Recognizing word meanings mainly uses long-term memory; analyzing sentence structure mainly uses short-term memory. Thus, the development of specialized terms allows speakers to shift the burden from short-term to long-term memory. It could be that mapping the meaning of a number word confers a similar gain in short-term memory. This suggestion is consistent with theories on the development of expertise in other domains (see e.g. Ericsson & Charness, 1994; Kintsch, Patel, & Ericsson, 1999).

4.2.2. Approximation—Another possibility is that Level II–III children chose to give approximate answers on the Compare-Sets task.¹ It is not unusual for adults to use number words in approximate ways. Consider the following hypothetical conversation:

Colleague A: “How big was Intro Psych when you taught it last year?”

Colleague B: “300 students. Why? Did they split it up this year?”

Colleague A: “No, it’s still 300.”

Of course, both speakers know that the exact number of students enrolled is probably not 300. But in this context, 300 is close enough. Why would Level II–III children use approximation on the Compare-Sets task if Level IV children don’t? And why only on the high-number trials? Perhaps because, as discussed above, unmapped number words are unwieldy concepts, making the whole task more difficult. Adults might also use approximation in cases when accurate answers are difficult to come by – imagine Colleagues A and B again, chatting in the hallway.

Colleague B: “So, how many students do you have in Intro Psych?”

Colleague A: “300.”

On the other hand, imagine that B asks A the same question in A’s office, where the latest enrollment figures happen to be sitting on the desk right in front of A.

Colleague B: “So, how many students do you have in Intro Psych?”

Colleague A: (glancing down at desk) “297.”

The simplest explanation of the difference in A’s answers would be that when the enrollment information was sitting right in front of her, an exact answer was easy to give. In the hallway, it would have been difficult.

For children at Level II–III, the fact that unmapped number words are placeholders (albeit placeholders for specific, unique numerosities) might make them awkward enough, and the Compare-Sets chain of inferences cumbersome enough, to be avoided in favor of an

¹We are grateful to Stanislas Dehaene and an anonymous reviewer for suggesting this possibility.

approximate answer. For Level IV children, however, calculating the correct answer is easy (as it is for Level II–III children on the low-number trials, either through inference or recall) and so the exact answer is offered.

4.2.3. Continuous versus discrete quantities—Still another possibility is that the children may have analyzed the Compare-Sets stimuli as continuous quantities rather than as sets of discrete objects.² In fact, the stimuli were designed in a way that encouraged this – they were pictures of slightly overlapping items (e.g. peaches), placed in a row. This was done deliberately, so that children could use length cues to determine easily whether the rows were equal or not. (Remember that we were most interested in children at Levels II and III, who could not yet use counting to make this determination.) Even if, as we have argued in this paper, children view unmapped number words as referring to specific numerosities, they may not apply this reasoning to their thinking about continuous quantities.

What is needed to test this possibility is two new tasks. One task should be as similar as possible to the Compare-Sets task but use stimuli that children analyze as sets of discrete objects. The other task should be as similar as possible to the Six-Versus-A-Lot task (or the Transform-Sets task) but use stimuli that children analyze as continuous quantities. If children applied the SUN principle in the case of discrete objects but not continuous quantities, the discrepancy in findings of the present study would be explained.

4.2.4. Food versus non-food items—A related explanation hinges on the fact that the Compare-Sets stimuli were pictures of food, whereas the stimuli in the other tasks were not.³ Could this have prompted children to think about the stimuli in a different way? Specifically, in light of the continuous-versus-discrete question mentioned above, food arrays may be particularly prone to analysis as continuous quantities. In her work with chimpanzees, Boysen has reported a response bias when candy arrays are used as stimuli (Boysen, Berntson, Hannan, & Cacioppo, 1996; Boysen, Mukobi, & Berntson, 1999). The bias disappears when symbolic representations of number are used instead. It is possible that children, like chimps, think about food differently than they think about buttons and batteries.

What is needed to test this possibility is, again, two new tasks. One should be similar to the Compare-Sets task but use non-food arrays. The other should be similar to the Six-Versus-A-Lot (or Transform-Sets) task, but use food. If children applied the SUN principle to the non-food items but not the food items, then the discrepancy in findings of the present study would be explained.

4.3. Number words in different contexts are not analyzed as homonyms

Some theories (including, but not limited to the ANB theory described in this paper) have suggested that children first learn the number words as homonyms. That is, they assume that the counting word *five* and the determiner *five* have unrelated meanings, despite sounding the same. If the SUN view is correct, then the homonym idea probably is not. In order to apply the SUN principle to cardinal number words and not overextend it to all quantifiers, children must have some way of figuring out which words are the cardinal number words in the first place. (The fact that children do not overextend the SUN principle to cover all quantifiers is demonstrated by the Six-Versus-A-Lot task reported here.)

²We are grateful to Alan Leslie for suggesting this possibility.

³We thank Alan Leslie for suggesting this possibility as well.

How do children identify the cardinal number words as a special subset of quantifiers? According to the homonym view, children cannot use the fact of these words appearing in the counting sequence as evidence about their properties, since these two contexts are not connected in the child's mind. Perhaps children use contextual cues to identify the cardinal numbers? Bloom and Wynn (1997) have identified four properties of cardinal number words. Number words (a) are used only with count nouns – *two dogs* but not *two water*, (b) do not appear with modifiers – *very much* but not *very three*, (c) precede adjectives in the noun phrase – *four yellow cars* but not *yellow four cars*, and (d) can occur in the partitive construction – *five of the boys*. Could it be that children apply the SUN principle to any word with these properties?

We think not. Of the four properties, (c) and (d) are true of all quantificational determiners, and (a) and (b) are true of subsets including (but not limited to) number words. There are even a few non-number words (*another, several*) that fit all of these criteria. But there is no evidence that children mistakenly believe that *another* or *several* are number words.

On the other hand, children do identify the counting words as a set, a point made by Fuson (1988, p. 35) :

Children seem to learn very early the distinction between words that are in the number-word sequence and words that are not in the sequence. In our experiments through the years with 3-, 4-, and 5-year-olds, not one of the more than 500 subjects has ever used anything but number words when asked to say the sequence or to count entities. With more than 40 2-year-olds, three children have used letters from the alphabet (either alone or mixed in with number words) on one trial each. Gelman and Gallistel (1978) also reported very infrequent use of nonnumber words in counting by 2- to 5-year-olds. Baroody (1986b) reported for a large sample of moderately and mildly retarded children aged 6 to 14 that only one ever used nonnumber words when counting (the alphabet was used). (p. 35)

It seems likely that children are able to distinguish cardinal number words from other quantifiers by noticing that only these words (and not other quantifiers, like *a lot*) are also used in counting. The discovery that children apply a SUN principle to unmapped number words supports this assumption.

4.4. If the idea of large, exact numerosities is not acquired with the cardinal principle, then when?

The question of whether children assume that sets have specific numerosities is different from the question of whether children assume that number words label these numerosities. The latter, however, entails the former. Because we have argued here that children treat number words as referring to specific numerosities, we have implicitly argued that children conceive of such things as specific numerosities.

This point is not uncontroversial. Hurford (1987), who has written in detail about number systems across languages, suggests that the culturally earliest systems of representing number do not allow for the expression of large, exact numerosities at all. Peter Gordon (2003) reports on a contemporary Amazonian people who use a *one/two/many* quantifier system and have trouble establishing numerical equivalence among large sets. Bilingual studies (e.g. Spelke & Tsivkin, 2001) suggest that exact arithmetic facts are represented in a language-specific format, and transfer poorly to other languages and facts (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). Such findings have led scholars to wonder whether the very idea of large, exact numerosities might in itself be acquired along with the linguistic number system – the essence of the ANB view described above.

On the other hand, the historian of numbers Georges Ifrah discusses several methods of calculating large, exact numbers in the absence of number words. These include carving notches on bones or sticks, tying knots in string, and putting pebbles in a pile (the word *calculation* derives in fact from the Latin *calculus*, “small stone”). Ifrah offers the example of a shepherd tending a herd of 55 sheep. Lacking any word for the number 55, the shepherd can nonetheless keep track of his flock on a piece of wood or bone, by carving one notch into it for each sheep. Each morning and evening, as the sheep move between shelter and pasture, the shepherd can “count” them off by moving his finger down the row of notches (Ifrah, 1985, p. 9). The use of such methods bespeaks an understanding that large, exact numerosities exist, even by people who lack number words to describe them. In that case, it would seem wrong to conclude that people who have no number words also have no idea of number.

After all, the reverse seems possible – to have a formal command of numbers without much intuitive grasp of the quantities they represent. The authors of this paper can add, subtract, multiply and divide the numbers one million, one billion, and one trillion. Given enough time and objects, we believe we could generate these numerosities for a puppet. But our sense of the quantities involved remains so poor that neither of us knew until recently which number most closely approximated her own age in seconds.⁴

These and other considerations (including the findings reported here) suggest that linking counting to cardinal number words is unlikely to be the way children learn to conceive of exact, large numerosities. Just how they do so remains unclear. To say that the idea of exact numerosities is present before Level IV is not to say that it is innate. All the children in the present study knew at least one, and usually two exact number words – language could still play an important role in the development of numerical concepts, just not the role described in the ANB view.

Then again, perhaps in attributing an integer list to adults, and then asking how children acquire it, scholars have misunderstood the difference between adult and child knowledge of numbers. Perhaps what *all* of us have, besides the number sense we share with other animals (Dehaene, 1997), is just the belief that large, exact numerosities exist, plus better or worse technologies for keeping track of them. The linguistic number system allows us to represent numerosities far beyond the range of our unaided number sense, but at the end of the day we may still have much in common with Teddy, angling for that magic number of M&Ms.

Acknowledgments

This research was supported by a fellowship from the University of Michigan Culture and Cognition Program to the first author, and by NICHD grant HD36043 to the second author. We are grateful to the teachers, staff, and children of the University of Michigan Children’s Center, University of Michigan Center for Working Families, University of Michigan Hospital Child Care Center, Jewish Community Center Early Childhood Center, and Gretchen’s House 6 for participating in this research. Thanks also to Susan Carey, Summerson Carr, Stanislas Dehaene, Samuel David Epstein, Nicola Knight, Brian Malley, Marilyn Shatz, Paulo Sousa, Henry Wellman, and two anonymous reviewers for their thoughtful comments on earlier drafts.

References

- Antell SE, Keating DP. Perception of numerical invariance in neonates. *Child Development*. 1983; 54:695–701. [PubMed: 6851716]
- Bloom P, Wynn K. Linguistic cues in the acquisition of number words. *Journal of Child Language*. 1997; 24:511–533. [PubMed: 9519584]

⁴HINT: 1 million seconds = about 11.5 days; 1 billion seconds = almost 32 years; 1 trillion seconds = the amount of time that has passed since the Neanderthals disappeared (Paulos, 1988, p. 10).

- Boysen ST, Berntson GG, Hannan MB, Cacioppo JT. Quantity-based interference and symbolic representations in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes*. 1996; 22(1):76–86. [PubMed: 8568498]
- Boysen ST, Mukobi KL, Berntson GG. Overcoming response bias using symbolic representations of number by chimpanzees (*Pan troglodytes*). *Animal Learning and Behavior*. 1999; 27(2):229–235.
- Carey S. Cognitive foundations of arithmetic: evolution and ontogenesis. *Mind and Language*. 2001; 16:37–55.
- Condry, K.; Cayton, G.; Spelke, E. Toddler counting: addition and subtraction by 3-year-old children. Paper presented at the 13th Biennial International Conference on Infant Studies; Toronto, Ontario. 2002 April.
- Dehaene, S. *The number sense: how the mind creates mathematics*. New York: Oxford University Press; 1997.
- Dehaene S, Spelke E, Pinel P, Stanescu R, Tsivkin S. Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science*. 1999; 284:970–974. [PubMed: 10320379]
- Ericsson KA, Charness N. Expert performance – its structure and acquisition. *American Psychologist*. 1994; 49(8):725–747.
- Freeman NH, Antonucci C, Lewis C. Representation of the cardinality principle: early conception of error in a counterfactual test. *Cognition*. 2000; 74:71–89. [PubMed: 10594310]
- Fuson, K. *Children's counting and concepts of number*. New York: Springer-Verlag; 1988.
- Gelman, R.; Gallistel, CR. *The child's understanding of number*. Cambridge, MA: Harvard University Press; 1978.
- Gordon, P. Numerical cognition without words: evidence from Amazonia. 2003. Manuscript submitted for publication
- Hunt, E.; Banaji, MR. The Whorfian hypothesis revisited: A cognitive science view of linguistic and cultural effects on thought. In: Berry, J.; Irvine, S.; Hunt, EB., editors. *Indigenous cognition: Functioning in cultural context*. Boston: Martinus Nijhoff Publishers; 1988. p. 57-84.
- Hurford, J. *Language and number: the emergence of a cognitive system*. New York: B. Blackwell; 1987.
- Huttenlocher J, Jordan N, Levine SC. A mental model for early arithmetic. *Journal of Experimental Psychology: General*. 1994; 123:284–296. [PubMed: 7931093]
- Ifrah, G. *From one to zero: a universal history of numbers*. New York: Viking; 1985.
- Karmiloff-Smith A. More about the same: children's understanding of post-articles. *Journal of Child Language*. 1977; 4:377–394.
- Kintsch W, Patel VL, Ericsson KA. The role of long-term working memory in text comprehension. *Psychologia*. 1999; 42(4):186–198.
- McGarrigle J, Donaldson M. Conservation accidents. *Cognition*. 1974; 3:341–350.
- Mehler J, Bever TG. Cognitive capacity of very young children. *Science*. 1967; 158:141–142. [PubMed: 6054816]
- Michie S. Why preschoolers are reluctant to count spontaneously. *British Journal of Developmental Psychology*. 1984; 2:347–358.
- Mix KS. Preschoolers' recognition of numerical equivalence: sequential sets. *Journal of Experimental Child Psychology*. 1999a; 74:309–332. [PubMed: 10552921]
- Mix KS. Similarity and numerical equivalence: appearances count. *Cognitive Development*. 1999b; 14:269–297.
- Mix KS, Huttenlocher J, Levine SC. Do preschool children recognize auditory-visual numerical correspondences? *Child Development*. 1996; 67:1592–1608. [PubMed: 8890501]
- Paulos, JA. *Innumeracy: mathematical illiteracy and its consequences*. New York: Hill and Wang; 1988.
- Schaeffer B, Eggleston VH, Scott JL. Number development in young children. *Cognitive Psychology*. 1974; 6:357–379.
- Sophian C. Early developments in children's use of counting to solve quantitative problems. *Cognition and Instruction*. 1987; 4:61–90.

- Sophian C, Adams N. Infants' understanding of numerical transformations. *British Journal of Developmental Psychology*. 1987; 5:257–264.
- Spelke ES, Tsivkin S. Language and number: a bilingual training study. *Cognition*. 2001; 78:45–88. [PubMed: 11062322]
- Starkey P, Cooper RG. Perception of numbers by human infants. *Science*. 1980; 210:1033–1035. [PubMed: 7434014]
- Strauss MS, Curtis LE. Infant perception of numerosity. *Child Development*. 1981; 52:1146–1152. [PubMed: 7318517]
- Van Loosbroek E, Smitsman AW. Visual perception of numerosity in infancy. *Developmental Psychology*. 1990; 26:916–922.
- Whalen J, Gallistel CR, Gelman R. Nonverbal counting in humans: the psychophysics of number representation. *Psychological Science*. 1999; 10:130–137.
- Wynn K. Children's understanding of counting. *Cognition*. 1990; 36:155–193. [PubMed: 2225756]
- Wynn K. Children's acquisition of number words and the counting system. *Cognitive Psychology*. 1992; 24:220–251.
- Xu F, Spelke E. Large number discrimination in 6-month-old infants. *Cognition*. 2000; 74:B1–B11. [PubMed: 10594312]

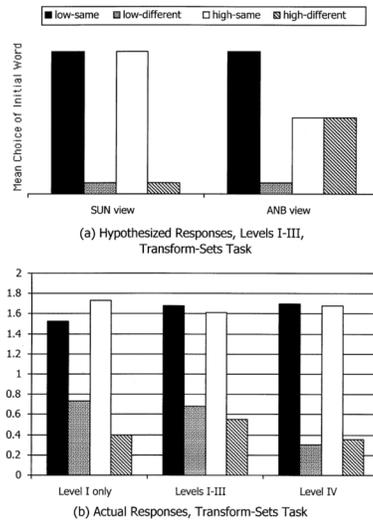


Fig. 1. Hypothesized and actual responses, Transform-Sets task.

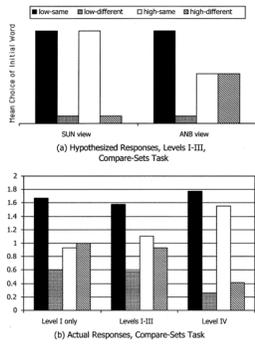


Fig. 2.
Hypothesized and actual responses, Compare-Sets task.

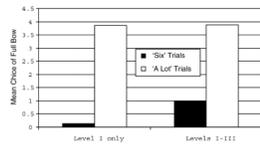


Fig. 3.
Results of Six-Versus-A-Lot task.

Appendix A

Six-plus-more task materials

Label	Description	Approximate number
Batteries	AA and AAA batteries, assorted brands	50
Beads	Wooden hair beads, 1.5 cm long × 1 cm diameter	100
Beans	Dried black beans, 1 cm × 0.5 cm	1500
Buttons	White and clear buttons, 1 cm to 2.5 cm diameter	500
Candles	Pink birthday candles, 6 cm long × 0.5 cm diameter	72
Fish	Foam cut-out fish, assorted colors, 3.5 × 3.3 cm × 2 mm	100
Jingle bells	Gold-colored jingle bells, 0.5 cm to 1.5 cm diameter	150
Paper clips	Vinyl coated paper clips, assorted colors, 2 cm × 0.5 cm	200
Pennies	US pennies	100
Washers	Hexagonal machine screw nuts, 1 cm diameter	500