Conceptual Correlates of Counting: Children's Spontaneous Matching and Tracking of Large Sets Reflects Their Knowledge…

7 authors, including:

Anna Shusterman
Wesleyan University

18 PUBLICATIONS   293 CITATIONS

SEE PROFILE

Jessica Taggart
University of Virginia

4 PUBLICATIONS   5 CITATIONS

SEE PROFILE

Talia Berkowitz
University of Chicago

5 PUBLICATIONS   44 CITATIONS

SEE PROFILE

Research Reports

Conceptual Correlates of Counting: Children's Spontaneous Matching and Tracking of Large Sets Reflects Their Knowledge of the Cardinal Principle

Anna Shusterman*[a], Pierina Cheung[a], Jessica Taggart[b], Ilona Bass[c], Talia Berkowitz[d],

Julia A. Leonard[e], Ariel Schwartz[f]

[a] Wesleyan University, Middletown, CT, USA. **[b]** University of Virginia, Charlottesville, VA, USA. **[c]** Rutgers University, Newark, NJ, USA. **[d]** University of Chicago, Chicago, IL, USA. **[e]** Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. **[f]** Boston University, Boston, MA, USA.

Abstract

The acquisition of counting is a major milestone for children. A central question is how children's non-verbal number concepts change as they learn to count. We assessed children's verbal counting knowledge using the Give-N task and identified children who had acquired the cardinal principle (Cardinal Principle Knowers, or CP-knowers) and those who had not (Subset-Knowers, or SS-knowers). We compared their performance on two tests of nonverbal numerical cognition. We report comparable performance between SS- and CP-knowers for matching and tracking small sets of objects up to four, but disparate performance for sets between five and nine, with CP-knowers outperforming SS-knowers. These results indicate that the difference between CP- and SS-knowers extends beyond their knowledge of the verbal number system to their non-verbal quantitative reasoning. The findings provide support for the claim that children's induction of cardinality represents a conceptual transition with concurrent, qualitative changes in numerical representation.

Keywords: cognitive development, numerical cognition, number development, cardinality, object tracking, approximate number system

The acquisition of counting, though it may seem to be a mundane skill, is a major milestone for children. In one sense, counting is clearly a feat of word learning: Children need to learn words for highly abstract number concepts and grasp how the structure of the count list gives meaning to the number words (i.e., going up one word in the list represents adding one item to the set; Gallistel & Gelman, 1992; Gelman & Gallistel, 1978). In another sense, counting is a feat of conceptual development: Children build an explicit symbolic representation that enables them to determine, track, and remember exact quantities (Carey, 2009; Carey & Sarnecka, 2006; Frank, Everett, Fedorenko, & Gibson, 2008; Gordon, 2004). A question central to this subject is how acquisition of the verbal count list interacts with non-verbal conceptual development in numerical reasoning.

Previous studies of children's verbal counting abilities have documented that children begin to recite a count list long before they develop stable meanings for the number words in this list. It then takes children two to three

years from the time they learn the count list to acquire the cardinal principle—the idea that the last number in the count represents the cardinality of the set (e.g., Fuson, 1988; Gelman & Gallistel, 1978; Schaeffer, Eggleston, & Scott, 1974; Wynn, 1992). The cardinal principle is a foundational concept that serves as a building block for later mathematical thinking. Using the Give-N task, researchers have consistently shown that English-speaking children start off learning the meanings of number words in a piecemeal fashion. First, around the age of two, children develop an understanding of the meaning of "one" ("one"-knowers; i.e., they can give one object when asked for one, and avoid giving one when asked for other quantities). They then make stage-like jumps approximately every six months to become "two"-, "three"-, and then "four"-knowers—collectively termed "Subset Knowers" (henceforth SS-knowers) because during these stages children know the exact quantities referred to by numerals for a *subset* of the count list they are able to recite (Le Corre, Van de Walle, Brannon, & Carey, 2006; Lee & Sarnecka, 2010, 2011; Sarnecka & Lee, 2009; Wynn, 1990, 1992). Children who can accurately provide an experimenter with any requested number (usually tested up to 6) are labeled "Cardinal-Principle Knowers" (henceforth CP-knowers). Becoming a CP-Knower is often regarded as a categorical shift distinct from the subset knower levels. While SS-Knowers might understand cardinalities of small numbers, they have not generalized the cardinality principle such that they can flexibly apply it to a range of set sizes.

Using a wide range of tasks, several studies have found that SS-knowers and CP-knowers differ qualitatively on their interpretations of number word meanings (e.g., Condry & Spelke, 2008; Le Corre et al., 2006; Le Corre & Carey, 2007; Sarnecka & Carey, 2008; Slusser & Sarnecka, 2011; Wynn, 1990, 1992). For example, only CP-knowers—but *not* SS-knowers— have exact numerical meanings for quantities larger than three or four. On the Give-N task, only CP-knowers can generate a large set of objects upon request (e.g., *Can you give me six fish?*), and this ability defines their status as CP-knowers. Further, CP-knowers, but not SS-knowers, appear to understand that moving forward one number word on the count list means adding one item to the set—e.g., when an object is added to a set of five objects, the set now has six objects and not seven (Sarnecka & Carey, 2008; see Davidson, Eng, & Barner, 2012 for evidence that this generalization extends only to a limited count list). Moreover, only CP-knowers can correctly match pictures that are labeled with large number words (e.g., *This picture has eight turtles. Find another picture with eight turtles.*), while SS-knowers are equally likely to choose the foil that matches on continuous extent such as total surface area, or non-numerical attributes such as color (green turtles) or mood (happy turtles; Slusser & Sarnecka, 2011). Finally, only CP-knowers, but not SS-knowers, show evidence that they understand equinumerosity—if two sets have the same cardinal value, they hold the same quantity (Sarnecka & Wright, 2013). Together, these studies show that CP-knowers differ from SS-knowers with regard to their knowledge of how number words relate to cardinalities of sets of objects, particularly for sets greater than four.

Several studies on populations that lack a verbal count list suggest that the ability to represent large exact quantities is dependent on having exact meanings for large number words. For example, studies with two Amazonian tribes show that verbal counting may allow humans to represent, match, and track large exact quantities (Frank et al., 2008; Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004). In one study of the Pirahã, a tribe that lacks a natural number system in their language, performance on set-matching tasks was excellent for small numbers (one, two, three) but became more variable as the set size increased (Gordon, 2004). A follow-up study by Frank and colleagues (2008) replicated Gordon's results, showing near-perfect performance on small set sizes and a linear decline in performance for quantities greater than four. In particular, they found that performance on larger sets (> 4) was especially impaired for tasks that required participants to remember

particular quantities for some short period of time or to re-create an array in a different spatial orientation than the target array. This pattern of results has also been found in another population. Flaherty and Senghas (2011) showed that older deaf signers in Nicaragua who lacked a stable count list also faced difficulties re-creating exact set sizes larger than six, but they were capable of re-creating small sets. Collectively, these studies show that speakers of languages that do not have a verbal count list have difficulty representing and tracking large but not small quantities, suggesting that the acquisition of verbal number is central to solving nonverbal numerical problems that require the representation of large exact quantities.

These findings are consistent with previous research showing that small sets of objects can be represented with parallel individuation (PI)—an object tracking system that is present in infants, adults, and other animal species (for review, see Feigenson, Dehaene, & Spelke, 2004). Using PI representations, individual objects are represented as mental symbols in working memory, but the system's working memory capacity is limited to representations of set sizes of up to three or four items in human adults (e.g., Feigenson & Carey, 2003, 2005; Feigenson, Carey, & Hauser, 2002; Hauser & Carey, 2003; Trick & Pylyshyn, 1994). This explains how people from the Amazonian tribes and Deaf Nicaraguan signers can still represent small exact sets by deploying the PI system, despite not having a verbal count list. However, given its set size limit, the PI system cannot support the representation of large sets beyond four. In fact, much previous research has pointed towards another innate non-verbal representation of sets—the approximate number system (ANS)—for representing sets of all sizes (Dehaene, 2011; Piazza, Pinel, LeBihan, & Dehaene, 2007; Whalen, Gallistel, & Gelman, 1999; Xu & Spelke, 2000; see Feigenson et al., 2004 and Cantrell & Smith, 2013 for reviews). Unlike the PI system, the ANS represents number as continuous mental magnitudes and is limited in precision: The ability to discriminate any two numbers is determined by their ratio (i.e., Weber's law). For example, it is easier to discriminate 5 from 6 dots than 55 from 56 dots, even though the absolute difference is one dot in both comparisons. Therefore, although the ANS can represent large sets, it cannot generate an exact representation of large quantities, leading to the less precise performance on number tasks seen amongst people from Amazonian tribes and the Deaf Nicaraguan signers. Consistent with this cognitive science literature, in this paper we refer to sets within the PI range as "small" sets and beyond this range as "large" sets.

If neither PI nor the ANS can support the representation of large exact quantities, how do humans represent, for example, a set of ten objects exactly? As studies on populations that lack a verbal count list suggest, the process of acquiring number word meanings may allow us to represent and track large exact quantities (Carey, 2004, 2009). Thus, one possibility is that acquiring the cardinal principle coincides with gaining the ability to track large exact quantities. Drawing on the cross-linguistic evidence, Frank et al. (2008) argued that the verbal counting system is a "cognitive technology" that makes existing exact number representations explicit, allowing a person to note and remember exact quantities larger than four via counting. It is possible that the acquisition of verbal counting influences even more fundamental processes than memory for ephemeral quantities: Learning number language may actually change perceptual, attentional, or conceptual aspects of number representation.

Notably, both of the nonverbal number systems undergo significant developmental change over the same time period during which most children acquire meanings for number words. The acuity of the ANS improves in a protracted fashion throughout early childhood, with a discriminability ratio of 1:2 in 6-month-old infants (Xu & Spelke, 2000) and 2:3 in 9- month-old infants (Lipton & Spelke, 2003, 2004). Numerical acuity dramatically increases in children between 3 and 5 years—the same period during which children acquire verbal counting

(Halberda & Feigenson, 2008)—and eventually reaches a discriminability ratio of about 7:8 in Western adults (Barth, Kanwisher, & Spelke, 2003). Developmental change is also evident in the PI system: Newborns can track only two objects (Coubart, Izard, Spelke, Marie, & Streri, 2014); 12-month-olds fail tasks requiring them to track more than three objects simultaneously (Feigenson & Carey, 2003, 2005); between the ages of 3 and 6, this limit increases to sets of four or even five (O'Hearn, Hoffman, & Landau, 2011; Ross-Sheehy, Oakes, & Luck, 2003; Starkey & Cooper, 1995).

Previous studies studying the relationship between children's counting skills and their ability to solve non-verbal numerical tasks (i.e., tasks that do not require the interpretation of numerical language like "two," "six," "more") have yielded mixed findings. Importantly, many of the studies on number language and concepts do not provide convincing evidence regarding the relationship between acquiring the cardinal principle and representing large sets in typical development, because the methods used do not evaluate both of these abilities simultaneously (Brannon & Van de Walle, 2001; Huntley-Fenner & Cannon, 2000; Mix, 1999a, 1999b, 2008a, 2008b; Mix, Huttenlocher, & Levine, 1996; Negen & Sarnecka, 2010). To test the hypothesis that acquiring a count list allows us to represent and track large exact quantities, two criteria need to be met: First, it is necessary to compare preschoolers who have not yet acquired the cardinal principle (i.e., SS-knowers) to those who have (i.e., CP-knowers). The SS-knowers serve as a meaningful comparison group: Although they often have a verbal count list up to "eight" or higher, they only have meanings for the small numbers. Second, it is important to use tasks that require children to represent large exact sets greater than four.

Previous developmental studies exploring the relationship between counting and non-verbal numerical representation have not met these two criteria. Using a non-verbal triad task, Mix (1999a, 1999b, 2008a, 2008b) showed children a target set of objects (2, 3, or 4), and asked them to match it from two to three alternative arrays that differed from the target set in various dimensions (e.g., the number of objects, density between individual objects, length, and object type). Mix tested children with a range of counting abilities (i.e., ages consistent with SS- and CP-knowers) and found that children with minimal counting proficiency (i.e., those who could give at least two objects in Give-N) could recognize numerical equivalence between sets despite surface dissimilarities, but children who completely lacked counting ability could not (see Negen & Sarnecka, 2010, for similar findings). These findings reveal that verbal counting knowledge is correlated with non-verbal numerical cognition, in that minimal counting ability corresponded with above-chance performance on a matching task. In these studies, further counting development, including performance on the Give-N task consistent with understanding cardinality, was not related to performance on the non-verbal tasks. However, the stimuli used in the study entailed small numbers only. If cardinality is specifically related to non-verbal representation of *large* exact sets, the relationship could not have been observed in this study.

In another study exploring the relationship between counting and non-verbal number representation, Brannon and Van de Walle (2001) used a non-verbal numerical comparison task and found an overall effect of verbal counting knowledge similar to Mix's (1999a, 1999b, 2008a, 2008b) studies. Specifically, children who knew at least some number word meanings were better at numerical comparison than those who did not know any number words. However, while this study used both small and large sets, the children were 2- and 3-year-olds and all in the pre-counting or SS-knower range. Thus, these findings again indicate a relationship between the onset of verbal counting and non-verbal numerical cognition, but the question remains whether there is an additional relationship between acquiring cardinality and non-verbal representation of large sets.

Finally, numerous recent studies have reported a more general correlation between symbolic number knowledge, typically performance on a math achievement test, and nonverbal numerical acuity in the ANS, in infants through adults (e.g., Libertus, Feigenson, & Halberda, 2011; Libertus, Odic, & Halberda, 2012; Mazzocco, Feigenson, & Halberda, 2011; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013; Starr, Libertus, & Brannon, 2013). Most of the studies in this area do not focus on how counting ability and the acquisition of the cardinal principle relates to the development of nonverbal numerical representations. Several recent studies explicitly test for and find a relationship between cardinality and ANS acuity (Chu & Geary, 2015; Shusterman, Slusser, Halberda, & Odic, 2016; Wagner & Johnson, 2011), indicating a qualitative change in magnitude representations when children acquire cardinality. These studies used a dot comparison task in which participants were shown two sets of dots and asked to select the more numerous set (e.g., "Point to the side with more dots"). While this task is often used to investigate the ANS, this paradigm reveals how finely participants perceive more-less relations, but not how accurately participants represent any particular quantity. Thus, the dot-comparison studies do not explicitly test for a difference between CP- and SS-knowers in the exact representation of quantities specifically larger than four.

To summarize, previous developmental studies have not found a relationship between the acquisition for number words beyond four (i.e., beyond the PI range) and the representation of exact quantities above four. The studies conclude only a minimal relationship between number words and performance on non-verbal tasks beyond the onset of counting, in contrast with the data from atypical populations, which suggest a strong relationship between number language and number concepts. However, those developmental studies did not explicitly hypothesize or test for a relationship between language and thought in the representation of large numbers. Additionally, there is a need in the field to develop non-verbal tasks other than dot comparison to assess the nature of children's cognitive representations of number. The current study addresses these challenges.

The Present Study

The goal of the present research is to investigate whether children who have grasped the counting system in language (i.e., CP-knowers) show better performance in matching and tracking large quantities in non-verbal numerical tasks than those who have not (i.e., SS-knowers). We hypothesized that children should be able to solve nonverbal problems involving small quantities (one, two, and three) regardless of their verbal number knowledge, drawing on the PI system (Chi & Klahr, 1975; Feigenson et al., 2004; Starkey & Cooper, 1980). In contrast, for quantities larger than three, children who have not fully grasped the verbal counting system should show more variable and less accurate performance than children who have mastered verbal counting.

To test this, we examined preschool children's performance on verbal and nonverbal number tasks using both small (1-3) and large (5-20) sets. In the present research, we developed two non-verbal numerical tasks, inspired by Hannula and colleagues' research on 'spontaneously focusing on numerosity' or SFON (e.g., Edens & Potter, 2013; Hannula & Lehtinen, 2005; McMullen, Hannula-Sormunen, & Lehtinen, 2013, 2014; see Rathé et al., 2016, for a recent review). SFON is defined by Hannula and Lehtinen (2005) as a child's self-generated tendency to pay attention to and engage with quantities and number without prompting or any explicit cues in the environment. SFON is measured with tasks that include the exact numerosity of a set as one of several dimensions to which a child might attend (e.g., imitating the experimenter's action of feeding carrots to a rabbit),

without explicit guidance or instruction regarding counting or numbers (i.e., without statements like "How many are there?" or "Count the carrots.").

In Experiments 1 and 2, we adapted the Cardinality task from Hannula and Lehtinen (2005) to create the Caterpillar Game: children were presented with stuffed toy 'caterpillars' with some number of 'feet' sewn on, and were asked to retrieve socks for each caterpillar. The major difference between our method and the original was that children in our study completed trials with all set sizes presented, whereas Hannula and Lehtinen started with one truncated the procedure if children did not provide an exactly correct response for a certain set size. Thus, not all children in their study received trials with large set sizes, precluding systematic analyses of children's performance on those set sizes. We presented children with a variety of set sizes that ranged from 1 to 10. Critically, following the spirit of other SFON research, the instructions did not involve number words or any numerical language (e.g., "more"). Children were also never explicitly prompted to count (e.g., by asking "how many"), but they were not discouraged from doing so.

We predicted that if acquiring the cardinal principle makes it easier for children to encode large quantities, CP-knowers, but not SS-knowers, should be more accurate in retrieving socks for the caterpillars. However, if any difference were found between SS- and CP-knowers' performance on the non-verbal matching task, it could be alternately explained by children's ability to generate verbal estimates corresponding to the number of items in the set. Therefore, we also assessed children's estimation knowledge, and the quality of children's mapping between verbal numerals and approximate representations of quantity in the ANS, using an estimation task— "Fast Cards"—used in previous studies (Le Corre & Carey, 2007; Odic, Le Corre, & Halberda, 2015; Shusterman et al., 2016). In this task, children were shown sets of items flashed quickly on the screen and were asked to guess the number of dots. Previous studies have used this task to identify children who provide larger estimates for larger sets (Mappers) from those whose estimates do not differentiate among large sets (Non-Mappers). We analyzed children's performance on the Caterpillar task as a function of the quality of their mappings on Fast Cards.

While Experiments 1 and 2 tested children's ability to *match* large numerosities, in Experiment 3 we developed another non-verbal numerical task testing children's ability to *track* exact large numerosities. Children were shown an apparatus called "Mr. Elephant." They saw *N* balls (between 2 and 7) go into the chute on the top of the box, and then Mr. Elephant "blew" either *N* or *N* – 1 (between 1 and 7) balls out of his trunk. Children were then asked if all the balls came out. Similar to the non-verbal numerical matching task in Experiments 1 and 2 (the Caterpillar Game), children were never explicitly prompted to count; however, they were also not discouraged from doing so. We predicted that, if induction of the cardinal principle allows children to encode and track large quantities, CP-knowers would demonstrate better performance on tracking the larger numbers of objects that were put into the chute (i.e., more than 3 or 4) compared to SS-knowers.

Experiment 1

In Experiment 1, children were tested on two assessments of verbal number knowledge (Give-N and Fast Cards) and one assessment of nonverbal number reasoning (Caterpillar Game). In the Caterpillar Game, three small set sizes (1, 2, and 3) and three large set sizes (6, 7, and 9) were used to test the hypothesis that CP-knowers would outperform SS-knowers in the high but not in the low number range.

Methods

Participants

Forty-nine children (*M* = 50 months, range = 36–63 months; 31 females) were tested at a child development laboratory or at nearby preschools. One additional child refused to follow the rules of the games during testing and was excluded from analyses. Participants were drawn from a socio-economically diverse area and were primarily white and from middle-class backgrounds. Children received small prizes for their participation and parents who traveled to the lab received a $5 travel reimbursement.

Testing Session

Children were run in a single session on the Caterpillar Game, Elicited Counting, Give-N, and Fast Cards, in that order. The Caterpillar Game was run first so that its performance would not be affected by exposure to the explicit counting tasks.

Caterpillar Game

Seven 19 in. long caterpillars were created from dark green soccer socks that were stuffed with batting and sewn shut (Figure 1). Each caterpillar was uniquely decorated with a distinct face and features and had a different number of light green feet, three small numbers (1, 2, 3) and three large numbers (6, 7, 9). The feet were distributed along the two sides of the caterpillar body. Seventeen of the children were also tested with a five-footed caterpillar, which was introduced later in the experiment. Thirty-six identical white infants' socks were arrayed on a table on the other side of the room from the experimenter and child. The positions of the experimental stimuli were adjusted so that children could not easily see the caterpillar when standing near the socks.
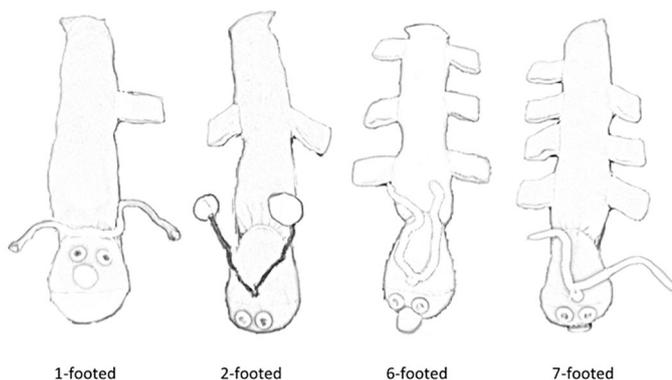

Figure 1. Representative schematic drawings of caterpillars (from photographs) with different numbers of 'feet'.

On each trial, children were introduced to a caterpillar and were told the following:

> Sammy wants to go for a walk, but he needs socks. See the socks over there? Could you get just enough socks for Sammy? Be careful though! If you don't bring enough socks, his feet will be cold. But if you bring too many socks, it will make a mess. Sammy's parents really do not like a messy room, so we don't want to have extra socks lying around. Can you go there and get *just enough socks*?"

Critically, the experimenter never explicitly suggested to children that they should count the socks, and they avoided using phrases like "how many socks" or "the right number of socks." Children were then allowed to go

to the sock table, which was located between 2m and 5m away from the testing area, and bring socks, which the experimenter and the child then put onto the caterpillar's feet. Once all the socks that the child had brought were used (or all the feet covered), the experimenter asked the child if there were "just enough socks." Children were encouraged to retrieve more socks or return extra socks to the pile as many times as needed. If they brought too many and did not spontaneously correct their error, the experimenter pointed out that the area was now messy and asked the children to return the extra socks; if they did not bring enough, the experimenter pointed out that the caterpillar's feet would be cold, and asked the child to retrieve additional socks. The child thus received feedback on every trial. We only present analyses of children's responses on the first attempt to retrieve socks. This is because for most trials, the number of socks that children needed to bring was small after the first retrieval, and therefore children were very accurate with the second retrieval.

Each session began with a one-footed caterpillar to help children understand the task[i], and ended with a two-footed caterpillar to ensure that children were attentive throughout the session (i.e., if children were consistently correct on this trial, we assumed that they understood the constraints of the task through all trials). Counting was neither encouraged nor forbidden, but we noted whether the child counted caterpillars' feet or socks. The one-footed caterpillar was used as the practice trial for each participant. The remaining trials (three, five, six, seven, or nine feet) were administered in one of three pseudo-random orders. A different caterpillar was used on every trial.

Elicited Counting

To establish if the child had a stable count list, the experimenter placed a line of 12 rubber ducks on a table and asked the child to count them. Each child's highest count was recorded.

Give-N

This task was adapted from Wynn (1992) following the method of Le Corre and Carey (2007). Children were presented with twelve small yellow toy ducks and a large green bowl (the "duck pond"). On each trial, children were asked to put a specific number of ducks in the pond (e.g., "Can you put *one* in the pond?"). Before coding each trial, the experimenter asked, "Is that *N*?" and gave the children an opportunity to spread out the ducks in a line, check if they had the correct number, and fix the number of ducks if they identified an error. The first trial asked for one duck, followed by a request for two ducks, three ducks, and so forth until they made an error. The order of trials was $N = 1, 2, 3, 4, 5, 6, 8, 7$. If children made an error, the experimenter requested one fewer items on the next trial. To be classified as an *N*-knower, children had to (1) give *N* on at least two out of three trials, (2) fail to give *N*+1 on at least two out of three trials where *N*+1 was requested, and (3) avoid giving *N* on at least two-thirds of the trials asking for more than *N*. Children who were classified as a '1-knower', '2-knower', '3-knower' and '4-knower' were collectively classified as 'Subset-Knowers' (SS-knowers). Children who passed the three trials (6, 8, 7) were classified as cardinal-principle knowers, or CP-knowers.

Fast Cards

Following Le Corre and Carey (2007), we used a verbal estimation task to identify whether children had made a mapping between number words in the non-subitizing range (i.e., between 4 and 10) and approximate quantities. The purpose of this task was to investigate whether children's performance on the Caterpillar Game could be explained generating verbal estimates for the target number of feet. In this task, children were told "This is a game where your job is to say the number word that goes with each picture. What do you see? One fish. So for this picture, you say 'one.'" During a demonstration set, children saw 1 to 15 fish presented in

sequential order to orient them to the task. The experimenter provided the correct answer on each demonstration trial. Children then received four test blocks (trains, hats, monkeys, and snowflakes). Each block used a different fixed random order of 1, 2, 3, 4, 6, 8, and 10 items. In two blocks, total picture area and envelope size were held constant and item size varied; in the other two blocks, item size was held constant while picture area and envelope size varied. Stimuli were displayed on a laptop with a PowerPoint presentation for 1 s. Two children gave responses above 20, which were replaced with a score of 20 to reduce the impact of outlier guesses (such as 100) while giving children credit for guessing a large number. Following previous studies (Le Corre & Carey, 2007), children were classified as Mappers if the linear slope of the mean responses for 6-, 8-, and 10-item trials was above 0.3; otherwise, the child was classified as a Non-Mapper.

Results

Elicited Counting, Give-N, and Fast Cards

Thirty-eight of the 49 participants were able to count to 10 with no errors, and all could count to 8 without error. Using the Give-N task, children were classified into knower levels. We found 19 SS-knowers (M = 47 months) and 30 CP-knowers (M = 51 months). SS-knowers were further classified as one-knowers (N = 3), two-knowers (N = 9), three-knowers (N = 3), four-knowers (N = 3), and five-knowers (N = 1). Finally, responses from Fast Cards were used to sort children into Mappers and Non-Mappers. We found that no SS-knowers were able to map large number words onto large numerosities (i.e., zero SS-knowers met the criterion for Mappers). For CP-knowers, we identified 16 Non-Mappers (M = 52 months) and 14 Mappers (M = 51 months). Using the 6-, 8-, and 10-item trials, the mean slope for Non-mappers was 0.12 (SD = .28) and the mean slope for Mappers was 0.96 (SD = .69).

Caterpillar Game

For the practice trial, children were shown a one-footed caterpillar. We found that SS-knowers brought a mean of 2.16 socks, while CP-knowers brought a mean of 2.47 socks. CP-knowers performed perfectly when the caterpillar had two feet, and SS-knowers were also highly accurate (M = 2.05). Children's relatively worse performance on the one-footed caterpillar is likely because children expected the caterpillar to have at least two feet, and did not fully understand the constraint that they should not bring back too many socks. The first trial thus provided children the opportunity to understand and reinforce the rule of bringing just enough socks for the caterpillar. Means and standard deviations for the number of socks retrieved at each target size are shown in Table 1.

To examine if SS-knowers and CP-knowers performed differently on the Caterpillar Game, we first analyzed the number of socks children retrieved for the caterpillars, followed by an analysis of the mean absolute errors on each retrieval.

Mean Retrieval

On low-number trials, SS-knowers and CP-knowers showed similar performance: On their first attempt, both groups brought more socks for the three-footed than the two-footed caterpillar, SS: $t(18)$ = 2.19, p = .042, d = 1.03; CP: $t(29)$ = 6.53, p < .001, d = 2.43. However, performance differed between groups on the high-number trials: Only CP-knowers demonstrated an understanding that they should bring more socks for more feet. Focusing on six and nine – the smallest and largest of the high-number trials – CP-knowers brought more socks for the nine-footed than for the six-footed caterpillar, $t(29)$ = 4.74, p < .001, d = .89, while SS-knowers did

Table 1

Means and SDs for the Number of Socks Retrieved for Caterpillars With X Feet

Group		Set Size (X)							
		1	2	3	5[a]	6	7	9	20
Expt. 1									
SS ($N = 19$)	M	2.16	2.05	2.89	4.71[a]	5.68	5.37	5.58	-
	SD	1.64	0.23	1.66	2.93	2.54	2.61	3.06	
CP ($N = 30$)	M	2.47	2.00	2.83	4.90[a]	5.80	6.10	7.40	-
	SD	2.00	0.00	0.70	1.52	1.32	1.77	1.81	
Expt. 2									
SS ($N = 7$)	M	2.14	2.43	-	4.88	-	-	-	7.00
	SD	1.35	1.13		2.53				4.36
CP ($N = 16$)	M	2.50	2.00	-	5.25	-	-	-	9.75
	SD	2.22	0.00		1.06				4.74

Note. SS = Subset-knowers; CP = Cardinal-principle knowers.
[a]Only 7 SS- and 10 CP-knowers were tested with the 5-footed caterpillar in Experiment 1.

not, $t(18) = 0.16$, $p = .878$, $d = .034$.[ii] Additionally, for the nine-footed caterpillar, CP-knowers brought significantly more socks than did the SS-knowers, $t(47) = 2.62$, $p = .012$, $d = .686$. This difference between SS- and CP-knowers was not significant at lower set sizes. Finally, the number of socks that CP-knowers brought for the six-, seven-, and nine-footed caterpillars increased linearly with the increasing number of feet ($M_{slope} = 0.55$, $SD = .58$). This slope was significantly and robustly different from a chance slope of zero, $t(29) = 5.151$, $p < .001$, $d = 5.15$. In contrast, SS-knowers exhibited a flat slope over these three high-number trials ($M_{slope} = -0.02$, $SD = .91$) that was not significantly different from zero, $t(18) = 0.072$, $p = .943$, $d = .27$. The mean slope of CP-knowers was significantly higher than SS-knowers, $t(47) = 2.66$, $p = .01$, $d = .776$, reflecting CP-knowers' tendency (and SS-knowers' inability) to bring more socks for caterpillars with more feet on the high-number trials (6-to-9 range).

Mean Error

SS-knowers and CP-knowers also differed in the magnitude of their errors in retrieving socks. Error rate was defined as the absolute difference between the number of feet on the caterpillar (target) and the number of socks brought on the first attempt (response). For example, for a seven-footed caterpillar, children who brought back either five socks or nine socks would have an error of 2.

We conducted a 2x2 mixed ANOVA with Knower-Level (SS vs. CP) as a between-subjects factor and Set Size (Small vs. Large) as a within-subjects factor. The dependent variable is summed errors across trials. The analysis revealed an overall main effect of Knower Level, $F(1,47) = 14.89$, $p < .001$, $\eta_p^2 = .241$, showing that the magnitude of errors of CP-knowers ($M = 4.07$, $SD = 3.83$) was smaller than that of SS-knowers ($M = 8.89$, $SD = 4.89$). Not surprisingly, we also found a main effect of Set Size, $F(1,47) = 92.18$, $p < .001$, $\eta_p^2 = .662$, showing that the magnitude of errors on large number trials ($M = 5.45$, $SD = 4.43$) was larger than that on small number trials ($M = 0.49$, $SD = 1.06$). We also found the hypothesized Knower-Level x Set Size interaction, $F(1,47) = 11.75$, $p = .001$, $\eta_p^2 = .200$. For the small-number trials, the magnitude of errors was similar for SS-knowers, $M = 0.79$, $SD = 0.15$, and CP-knowers, $M = 0.30$, $SD = 0.65$, $t(47) = 1.60$, $p = .117$, $d = .57$. However, for the large-
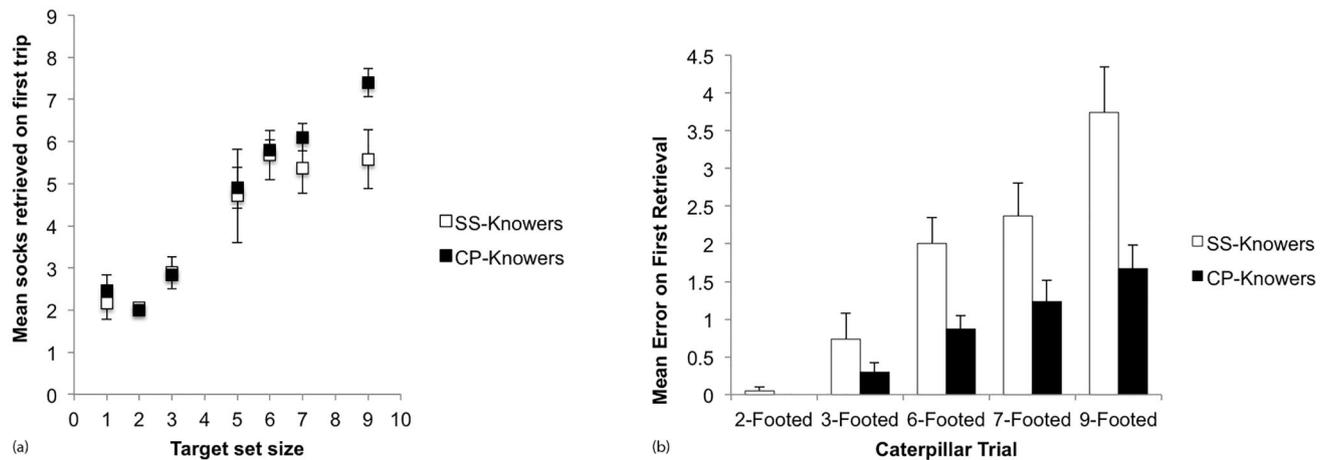
Figure 2. (a) SS- and CP-knowers' mean number of socks retrieved at each set size. (b) SS- and CP-knowers' mean error on small and large number trials. Error bars represent 1 SEM.

number trials, the magnitude of errors for SS-knowers, $M = 8.11$, $SD = 4.63$, was significantly greater than that for CP-knowers, with a large effect size, $M = 3.77$, $SD = 3.40$, $t(30) = 3.53$, $p = .001$, $d = 1.29$. We also analyzed errors for the six-, seven-, and nine-footed caterpillars individually, and found that SS-knowers had significantly higher error rates than CP-knowers for each of the large-number footed caterpillars (six-footed: $t(47) = 3.14$, $p = .003$, $d = .92$; seven-footed: $t(47) = 2.28$, $p = .027$, $d = .67$; nine-footed: $t(47) = 3.31$, $p = .002$, $d = .97$; see Figure 2).

Performance by Knower-Level

Although the sample size did not permit analysis by each knower-level, we were interested in whether children's performance increased across knower-levels prior to the CP transition. We separated the SS-knower group into two groups, 1/2-knowers ($N = 12$) and 3/4-Knowers ($N = 7$). We compared these two groups to each other and found no significant differences on any of the critical dependent measures: mean retrieval for large set sizes, total error for large set sizes, and slope from 6 to 9, even after controlling for age (retrieval: $F(1,16) = .076$, $p = .786$, $\eta_p^2 = .005$; error: $F(1,16) = .297$, $p = .593$, $\eta_p^2 = .018$; slope: $F(1,16) = .768$, $p = .394$, $\eta_p^2 = .046$). Finally, to confirm that the difference between SS-Knowers and CP-Knowers would hold for both groups of SS-Knowers, not just the younger or less knowledgeable ones, we ran a series of One-Way ANOVAs with Group as the independent variable (3 levels: 1/2-Knowers, 3/4-Knowers, and CP-Knowers) and Total Error for the large sets, Slope for the large sets, and Mean Retrieval for the large sets as the dependent variables. All three ANOVAs were statistically significant, $F(2,46) > 4.15$, $p < .03$, $\eta_p^2 > .22$. Importantly, post-hoc LSD tests indicated that the differences between the two groups of SS-Knowers were not statistically significant, but the differences between 3/4-Knowers and CP-Knowers were (Total Error: 1/2-Knowers ($M = 8.67$) vs 3/4-Knowers ($M = 7.14$), $p = .419$, $d = .33$; 3/4-Knowers vs CP-Knowers ($M = 3.77$), $p = 0.46$, $d = 1.00$; Slope: 1/2-Knowers ($M = .125$) vs 3/4-Knowers ($M = -.26$), $p = .275$, $d = .416$; 3/4-Knowers vs CP-Knowers, ($M = .55$), $p = 0.011$, $d = 1.37$; Retrieval: 1/2-Knowers ($M = 5.33$) vs 3/4-Knowers ($M = 5.90$), $p = .487$, $d = .24$; 3/4-Knowers vs CP-Knowers ($M = 6.43$), $p = .067$, $d = .38$, the last finding significant in a one-tailed t-test.)

Effects of Counting

One explanation for why CP-knowers outperformed SS-knowers is that CP-knowers more readily engaged counting as a problem-solving strategy. To address this, we analyzed effects of counting on task performance. We loosely defined children as "Counting" if they showed evidence of engaging in overt counting on any trial (N = 26) and "Not Counting" if they did not count on any of the trials (N = 23). Only three participants chose to count on every trial, and nobody counted both the feet *and* the socks on a single trial. Of the 19 SS-knowers, 7 counted on at least one trial, while 12 never counted. Of the 30 CP-knowers, 19 counted on at least one trial, while 11 never counted. Counting was marginally associated with Knower-Level, $\chi^2(1)$ = 3.28, p = .07.

Of the 26 children who counted, only four, all CP-Knowers, brought the correct number of socks for each trial on which they counted, implying that counting does not guarantee success on this task. However, on large number trials (6, 7, and 9), children who counted made fewer total errors (3.27) on average than children who did not count (7.91), $t(29)$ = 4.10, p < .001, d = 1.53.

Next we explored whether Counting and Knower-Level contributed independently to success on the task, or whether a propensity to count was behind the better performance seen in the CP-knowers. An ANOVA with total error on large sets as the dependent measure, counting behavior (Counting, Not Counting) and Knower-Level (SS, CP) as fixed factors, and age in months as a covariate, showed main effects of counting, $F(1,44)$ = 12.95, p .001, η_p^2 = .215, and knower-level, $F(1,44)$ = 7.00, p = .011, η_p^2 = .137, no interaction, and no effect of nor interaction with age. Looking only at the sub-group of children who did not count, CP-Knowers still made fewer total errors in the large-number range than did SS-knowers, $t(21)$ = 2.38, p = .027, d = 1.039, a large effect. This CP-knower advantage was also observed for children who counted, though with a smaller effect, $t(24)$ = 2.18, p = .039, d = .636. Thus, the effects of knower-level on performance in the caterpillar game appear to be independent of whether children choose to count during the task and independent of age.

Effects of Verbal Estimation Skills

We also tested the possibility that CP-knowers outperformed SS-knowers because of superior abilities in generating verbal estimates. The Caterpillar Game is, in some sense, a non-verbal estimation task. Our group of CP-knowers included both Mappers, who could generate fairly accurate verbal estimates for set sizes between 6 and 10, and Non-Mappers, whose estimates were far less accurate. We reasoned that if verbal estimation ability was beneficial to performance on this task, then CP-Mappers should outperform CP-Non-Mappers. Contrary to this prediction, CP-Mappers and CP-Non-Mappers made an equal number of total errors on the large-number trials (mean total error 3.94 vs. 3.97, respectively, $t(28)$ = -0.29, p = .770, d = -.11) and exhibited identical slopes on these trials (0.56 for Non-Mappers, 0.54 for Mappers, $t(28)$ = 0.05, p = .958, d = .019). There was no correlation between the slope of children's verbal estimates and the slope of their responses on the Caterpillar Game for numbers above 3 (r = -.07, p = .64). Thus, although Mappers and Non-Mappers differ in the accuracy of their verbal estimates for higher numbers, their non-verbal estimation abilities appear identical.

Set Size Five

Because of the marked differences between SS-and CP-knowers on the high number set sizes, it was of interest to know whether SS-knowers' ability to solve the task was poor for any set size outside the PI limit, or alternatively, whether their ability declined incrementally with increasing set size. Consequently, the last seventeen children tested received an additional trial with a five-footed caterpillar intermixed with the other

trials. SS- and CP-knowers brought similar numbers of socks, 4.71 (SS) vs. 4.90 (CP), $t(15) = 0.17$, $p = .87$, $d = .088$. CP-knowers' errors were smaller in magnitude than SS-knowers', 0.70 vs. 2.00, though this difference did not reach significance, $t(15) = 1.61$, $p = .13$, $d = .831$. Though these findings suggest that SS- and CP-knowers performed similarly on the 5-footed caterpillar, CP-knowers brought significantly fewer socks for the five- than the six-footed caterpillar, $t(9) = 2.25$, $p = .05$, $d = 2.1$, while this difference was not statistically significant in SS-knowers, $t(6) = 0.79$, $p = .46$, $d = .51$. To further explore the question of whether SS-knowers' performance in the large number range drops linearly, we analyzed their slopes for the five-, six-, seven-, and nine-footed caterpillars. Only children who were tested on the five-footed caterpillar were included in this analysis. We found that SS-knowers exhibited a flat slope over these four high-number trials ($M_{slope} = 0.14$) that was not significantly different from zero, $t(16) = 0.37$, $p = .72$, $d = .61$. These results suggest that SS-knowers treated five as they did the other large sets: less precisely than their own performance with small sets, and less precisely than did CP-knowers.

Effects of Age, Sex, and Fall vs. Spring Testing

As a final test, separate ANCOVAs were run with Knower-Level as the independent factor and total error as the dependent variable; counting and one other factor (age, sex, or spring vs. fall testing) were included as covariates in each analysis. Time of testing was identified as a possible covariate because children who were tested in the spring had several more months of schooling than those tested in the fall, and thus may have demonstrated better performance on the Caterpillar Game. Knower-Level was significantly associated with total error in all three analyses. None of the three covariates yielded significant effects (Age: $F(1,45) = 2.43$, $p = .13$, $\eta_p^2 = .051$; Sex: $F(1,45) = 0.015$, $p = .903$, $\eta_p^2 = .037$; Time of testing: $F(1,45) = 0.103$, $p = .750$, $\eta_p^2 = .096$). These results indicate that having exact meanings for number words higher than four is robustly related to the ability to solve problems with quantities higher than three in a non-verbal task, even after controlling for the effects of spontaneous counting, age, sex, and amount of schooling.

Discussion

Consistent with our hypothesis, Experiment 1 demonstrated comparable performance between SS- and CP-knowers for small numbers up to four, and contrasting performance between SS- and CP-knowers for large numbers between six and nine. CP-knowers gave increasing responses for larger targets in the high-number range, while SS-knowers did not show differential responses for caterpillars that had a large number of feet. These differences were statistically significant and robust with large effect sizes.

Although there were not significant differences for two of the large set sizes (six and seven), this is likely because SS-knowers brought roughly 5.5 socks no matter how many were required. Thus, while their responses appeared accurate for set sizes 6 and 7, this was likely just an artifact of 5.5 being SS-knowers' typical response for all set sizes. In the more informative comparisons on set size nine, error rates for 6-9, and slopes in the 6-9 range, SS- and CP-knowers' responses were clearly different.

These effects could not be explained by overt counting or by estimation skills. Rather, the results suggested that children's knowledge of verbal counting and cardinality were related to more refined approximate performance on the non-verbal numerical task. The results implied a sharp cutoff in performance beyond the PI range, but this needed to be tested further as Experiment 1 did not systematically include, for all children, a set size of 5, just beyond the PI boundary. Experiment 1 also showed a stark contrast between small and large sets

for subset-knowers, in that they were sensitive to quantities within the small-number range, but completely insensitive to quantities beyond it. Given that even subset-knowers have access to approximate number representations, which should support reasoning about larger quantities, this result raises the question of why such representations were not engaged on trials with quantities above four. Experiment 2 was designed to explore these two findings further by introducing a 5-footed caterpillar, just beyond the PI range, and a 20-footed caterpillar, to test whether subset-knowers would express more sensitivity to a more extreme numerical difference.

Experiment 2

Results from Experiment 1 indicated a clear distinction between SS-knowers and CP-knowers on their performance on large-number trials. Specifically, SS-knowers appeared to treat all large sets (between 5 and 9) similarly in the Caterpillar Game, while CP-knowers differentiated among the large set sizes. Experiment 2 was conducted to address two questions that were raised by the Experiment 1 results. First, it was not clear whether children's representation of set size 'five' followed the pattern of exact responses to small numbers, or poorly differentiated responses to large numbers, or an intermediate pattern. Many SS-knowers brought approximately five socks for a variety of large number trials, making it difficult to tell whether responses of five socks to five feet was an accurate response based on an exact representation, or a typical response to large numbers. The five-footed caterpillar trials in Experiment 1 provided preliminary evidence that children handled sets of five as a "large" number, suggesting a sharp drop-off in performance beyond the PI system's limit. However, because this set size was added partway into the study, only a small group of children was tested on 'five' and the comparison of error between SS-knowers and CP-knowers did not reach statistical significance. Experiment 2 therefore aimed to replicate the patterns on set size five with a different sample of children.

Second, it was not clear whether SS-knowers had entirely failed to notice the differences between the six, seven, and nine-footed caterpillar, or, alternatively, whether they had difficulty discriminating these quantities in order to solve a number-relevant problem. To address this, we included a caterpillar with 5 feet and a caterpillar with 20 feet, as children should be more likely to notice the difference between 5 and 20 feet than the difference between 6 and 9 feet. If SS-knowers used this very marked difference to retrieve more socks for the 20-footed caterpillar, it is likely that they simply did not notice the differences between the high-number caterpillars in Experiment 1. If, however, SS-knowers did not bring more socks for the 20-footed caterpillar, perhaps they noticed the difference but were unable to apply this information to solve the numerical problem.

Methods

Participants

Twenty-three children (*M* = 48 months, range = 37–61 months) participated in Experiment 2. Children were recruited similarly as in Experiment 1. They completed Give-N, Fast Cards, and the Caterpillar Game, in that order. Give-N and Fast Cards were identical to Experiment 1.

Caterpillar Game

The Caterpillar game consisted of four trials, with set sizes 1, 2, 5, and 20. As in Experiment 1, the first trial involved a one-footed caterpillar to make sure children understood the task and the two-footed caterpillar came

last. The order of the two middle trials (five- and twenty-footed caterpillars) was counterbalanced across children. Two additional children, both subset-knowers, did not complete the Caterpillar Game and were therefore not analyzed.

Results

We identified 7 SS-knowers (1 one-knower, 4 two-knowers, 1 three-knower, and 1 4-knower) and 16 CP-knowers using Give-N. CP-knowers were further sorted using Fast Cards into 8 Non-Mappers and 6 Mappers; 2 CP-knowers did not complete the task. A preliminary analysis revealed no differences at all between CP-Mappers and CP-Non-Mappers on both 5- and 20-foot trials (p > .4), so these groups were combined for analysis as CP-knowers.

We first analyzed children's performance on the five-footed and twenty-footed caterpillars separately to better understand children's responses to quantities outside the PI range (see Table 1). For the five-footed caterpillar, SS-knowers brought an average of 4.88 socks (*SD* = 2.53) while CP-knowers brought 5.25 (*SD* = 1.07). The difference in mean retrieval was not significant between groups, $t(21) = 0.05$, $p = .961$, $d = .02$, but the difference in error was large and significant: SS-knowers' error averaged 2.13 socks compared to 0.63 for CP-knowers, $t(21) = 3.13$, $p = .005$, $d = 1.37$.

For the twenty-footed caterpillar, both SS- and CP-knowers noticed and commented that there were "a lot" of feet, and indeed retrieved more socks, on average, for this caterpillar than children had in Experiment 1 for any other set size (7.00 for SS, 9.75 for CP), including the 9-footed caterpillar in Experiment 1; this difference between 9 and 20 was statistically significant only in CP-knowers, $t(44) = 2.42$, $p = .02$, $d = .73$.

Next, we asked whether children, particularly the SS-knowers, were sensitive to the difference between the 5-footed caterpillar and the 20-footed caterpillar. CP-knowers retrieved significantly more socks for the 20- than the 5-footed caterpillar, $t(15) = 3.67$, $p = .002$, $d = 1.12$. SS-knowers also retrieved more socks for the 20- than the 5-footed caterpillar, though this difference did not reach statistical significance, $t(6) = 1.77$, $p = .127$, $d = 1.18$. Analyzing individual children's data, we note that five SS-knowers brought more socks for the more numerous target, and only one child brought more for the less numerous one; this pattern was significant, Wilcoxon signed ranks test, $Z = 17$, $p = .05$ (one child brought the same number). These findings suggest that neither SS- and CP-knowers treat the 5-footed and 20-footed caterpillars as if they are the same. Although the pattern of results is similar in both groups, CP-knowers differentiated their responses more by bringing more socks for the 20-footed caterpillar.

Discussion

Experiment 2 extended the findings from Experiment 1 by examining children's performance with sets of five (just beyond the PI range) and sets of twenty (more than double the highest set size from Experiment 1). For the five-footed caterpillar, the variance in the responses was much higher in SS- than CP-knowers, just as it was for the six-, seven-, and nine-footed caterpillars in Experiment 1. This pattern suggests that five is treated as a 'large' number by subset-knowers, corroborating the conclusion from Experiment 1 that there is a divergence in the quality of SS- and CP-knowers' performance on this task with set sizes outside the PI range.

For the twenty-footed caterpillar, the pattern of responses did not exactly mirror those observed for large sets under ten, in two ways. First, SS-knowers showed some sensitivity (by Wilcoxon test, above) to the difference between five and twenty, indicating that with sufficiently different target quantities, they could differentiate their responses. Second, even CP-knowers underestimated how many socks to bring back, and the differences between SS-and CP-knowers were not as strong for the 20-footed caterpillar as for other set sizes. If CP-knowers were generally more attuned to quantity–for instance, if their performance on twenty could be predicted by extrapolating from their responses for six, seven, and nine–then the 20-footed caterpillar provided them with the best chance to demonstrate a large difference from SS-knowers, but they did not. Notably, "twenty" was beyond the comfortable estimation range for most of the children. Rather than revealing a robust difference between SS- and CP-knowers in their representation of twenty, it appears that the difference between groups diminished when children were presented with a larger quantity that was less familiar to both groups.

In short, Experiment 1 and 2 indicated that CP-knowers had a more refined response to a quantity-matching task, with more accurate and less noisy estimates of socks for a given number of feet. We note two results that could have occurred but did not: CP-knowers could have solved the task precisely, simply by counting the feet and the socks; or CP-knowers could have exhibited better exact matching of the numbers of socks and feet with a different non-verbal representation (e.g., chunking). However, the observed pattern suggests that CP-knowers had a more refined *approximate* representation of the number of feet and corresponding socks for larger sets beyond the PI range (i.e. above four).

Experiment 3

While Experiments 1 and 2 revealed an advantage in CP-knowers in their use of *approximate* representations of large sets, Experiment 3 explored whether CP-knowers show advantages in the representation and use of *exact* numerical quantities. To this end, we created a new non-verbal task that required children to track an exact number of objects.

We created a different non-verbal paradigm–the Mr. Elephant Game–to test children's ability to track large exact quantities beyond the PI range. In the Mr. Elephant Game, the experimenter placed balls inside a box named "Mr. Elephant." On half of the trials, the experimenter surreptitiously stopped one ball from coming out of Mr. Elephant (by toggling a small plastic disc), and on the other half of the trials, all of the balls came out. At the end of each trial, children were asked whether there were balls left in the box. While both the Mr. Elephant Game and the Caterpillar Game were non-verbal numerical tasks, one fundamental difference between these two tasks is the nature of the responses elicited from the children. The Caterpillar Game requires children to retrieve "just enough socks" from a very large pile, posing an essentially open-ended problem of how many socks to retrieve. In contrast, the Mr. Elephant Game asks children to distinguish *x* objects from *x* – 1 objects by answering a yes-or-no question; thus, on a given trial, children's responses were always correct or incorrect.

Methods

Participants

Nineteen children (*M* = 46.8 months, range = 38–56 months, 9 females) participated in Experiment 3. All participants were recruited as in Experiment 1.

Testing Session

Fifteen of the 19 children were run in a single testing session on Give-N and the Mr. Elephant game, in that order. The remaining four participants were run on Give-N in one testing session and on Mr. Elephant in a second testing session two days later.

Mr. Elephant

A hollow, wooden cube (length, width, and height = 27 cm) was painted dark blue, and paper eyes and felt ears were pasted on the front and sides of the box to create "Mr. Elephant" (Figure 3). There was one cylindrical chute on the top of the box and another chute coming out the front, connected by a tube inside the box. Two small plastic doors inside the tube, operated by levers on the exterior of Mr. Elephant, allowed the experimenter to stop the balls from passing all the way through the tube. One door was near the top, and one was near the trunk.
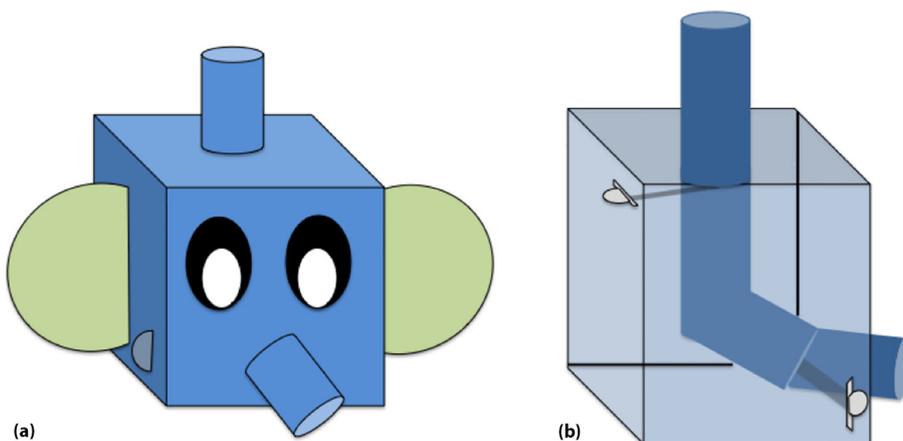

(a)   (b)

Figure 3. Schematic drawing of (a) exterior and (b) interior of Mr. Elephant apparatus with levers used by the experimenter to control the path of the balls through the tubes.

Testing Procedure

At the beginning of the experiment, the child was shown a bowl containing 7 green Styrofoam balls 4 cm in diameter. The experimenter explained to the child that Mr. Elephant liked to eat the "green peanuts" and then blow them out of his trunk. But sometimes, the child was told, the peanuts got stuck in his trunk, so Mr. Elephant needed the child's help to make sure all the peanuts came out.

On each trial, the experimenter placed either 2, 3, 5, or 7 balls on top of the box in a fixed pseudorandom order. Each number of balls was presented to the child twice—one trial releasing N balls, and the other trial releasing N-1 balls—yielding a total of eight trials per testing session. Each testing session began with the easier 2-ball trials, which introduced the procedure to the child, and ended with the 3-ball trials, to ensure that children understood the task and were attentive throughout the session. The remaining trials (5 and 7) were presented

in one pseudo-random order (5 in/4 out, 7-in/7-out, 5-in/5-out, 7-in/6-out). Feedback was available on every trial since one more ball either came out or did not on each trial.

The experimenter circumscribed the balls with her finger and said, "Look! I'm going to feed Mr. Elephant these peanuts!" At the beginning of each new trial, the experimenter said "Remember, let me know if you think a peanut is stuck." The experimenter then dropped the balls into the top chute one by one. The balls were blocked from immediately coming out of the front chute by the small plastic door near the trunk. On one of the two trials for each number, the experimenter surreptitiously toggled the second plastic disc to block the final ball from going down the top chute.

The experimenter then told the child that Mr. Elephant was going to blow out the 'peanuts,' lifted the disc blocking the front chute, and allowed the balls to come out. The child was then asked, "Did they all come out?" An affirmative response was correct for the 50% of the trials when all of the balls came out, while a negative response was correct for the 50% of trials when all but one of the balls came out. Once all of the balls were out, the experimenter said "Good job!" if the child was correct. If the child was incorrect, she would say "Let's check! Uh oh! I think a peanut is stuck! Can you make it come out? Thank you!" or "Oops! It doesn't look like any peanuts were stuck."

Results

We identified 9 SS-knowers (4 one-knowers, 1 two-knower, 3 three-knowers, 1 four-knower) and 8 CP-knowers using Give-N. Two five-knowers were also identified, but were excluded from subsequent analyses.[iii]

We tested whether CP-knowers would respond more accurately than SS-knowers on the large number trials in the Mr. Elephant Game, as they had on the Caterpillar Game. A 2x2 mixed ANCOVA was performed on the percentage of correct responses, with Set Size (Small [2- and 3- ball trials] vs. Large [5- and 7-ball trials]) as a within-subjects factor, Knower-level (SS vs. CP) as a between-subjects factor, and age (in months) was entered as a covariate. This analysis revealed a significant main effect of Knower-Level, $F(1,14) = 11.87$, $p = .004$, $\eta_p^2 = .46$, with CP-knowers ($M = .81$, $SD = .093$) correctly assessing whether all the balls had come out more frequently than SS-knowers ($M = .64$, $SD = .083$). In addition, as predicted, there was a significant interaction between Knower-level and Set Size, $F(1,14) = 8.49$, $p = .011$, $\eta_p^2 = .38$ (see Figure 4). There was no significant effect of age. Post-hoc pairwise comparisons revealed that in the Small trials, SS- and CP-knowers performed equally (SS: $M = .94$, $SD = .11$; CP: $M = .94$, $SD = .12$; $p = .93$, $d = 0$). In contrast, CP-knowers significantly outperformed SS-knowers on the Large trials (SS: $M = .34$, $SD = .18$; CP: $M = .69$, $SD = .13$; $F(1,14) = 15.24$, $p = .002$, $\eta_p^2 = .52$). Furthermore, both SS- and CP-knowers performed well above chance (50% accuracy) for small set sizes (p's $< .001$), while only CP-knowers performed above chance for large set sizes, $t(7) = 3.03$, $p = .019$, $d = 1.74$. In fact, surprisingly, SS-knowers fell significantly below chance on the large set size trials, $t(8) = -2.88$, $p = .021$, $d = 1.70$.
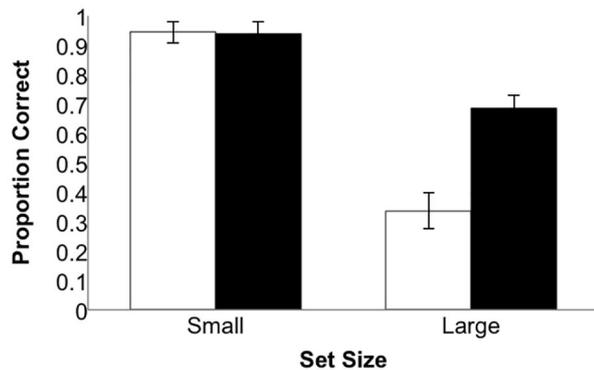
Figure 4. CP- and SS-knowers' performance on small (2 and 3 balls) and large (5 and 7 balls) set sizes in the Mr. Elephant task.

Looking at Table 2, it is clear that both SS- and CP-knowers were more accurate on *N*-1 trials, suggesting that children had an overall tendency to say that something was stuck in Mr. Elephant; nevertheless, CP knowers were more accurate than SS knowers on both trial types (*N* and *N*-1).

Table 2

Mean (SD) Proportion of Correct Responses for Large Sets, Separated by Empty (5 in/5 out, 7 in/7 out) and Non-Empty (5 in/4 out, 7 in/6 out) Trials

Trial Type	SS-Knowers	CP-knowers
Empty	.111 (.220)	.500 (.378)
One Inside	.611 (.417)	.813 (.372)
d^1 (d-prime)	-.947	.878

Effects of Counting

Counting behavior was not recorded for all participants, but was available for 4 of the 8 CP-knowers and 7 of the 9 SS-knowers in our sample. No participant counted on every trial, 4 children (2 SS-knowers and 2 CP-knowers) counted on at least one trial, and 7 children never counted. Of the two SS-knowers who counted, both counted on two trials, but responded inaccurately. Of the two CP-knowers who counted, one counted on one trial and responded accurately. The other CP-knower counted on two trials, and was accurate on one trial but inaccurate on the other trial.

Including only the children for whom counting behavior was recorded, an ANCOVA with SS/CP knower level as the independent variable, age and counting behavior as covariates, and proportion correct on small and large trials as a repeated measure, revealed a large and significant main effect of SS/CP, $F(1,7) = 10.66$, $p = .014$, $\eta_p^2 = 2.26$, and a marginal interaction between set size and SS/CP, $F(1,7) = 4.609$, $p = .069$, $\eta_p^2 = .40$, and no other significant effects or interactions. Specifically, there was no hint that counting behavior explained the SS/CP difference, because no effect of counting was detected, $F(1,7) = .034$, $p = .859$, $\eta_p^2 = .005$. Thus, knower-level, not counting behavior or age, predicted success on the Mr. Elephant task.

Discussion

Experiment 3 replicated and extended the pattern of results from Experiment 1 using a different task. SS- and CP-knowers performed similarly and were highly accurate on a non-verbal number task for smaller numbers (1 to 3), but CP-knowers significantly outperformed SS-knowers for larger numbers (> 4). Using a non-verbal tracking task, we found evidence that understanding the cardinal principle is related to better tracking and memory for large quantities. Specifically, in the Mr. Elephant Game, children were asked to *track one set of items* over a temporal and spatial gap, and to notice whether the complete set was re-established, whereas in the Caterpillar Game, children had to numerically *match* one set of objects (feet) to another set of objects (socks) in *one-to-one correspondence*. Nevertheless, we found the same pattern of results—namely, CP-knowers outperform SS-knowers on large number trials.

General Discussion

The current experiments test the relationship between preschool children's knowledge of cardinality and their responses on two non-verbal tasks: numerical matching (Caterpillar Game) and object-tracking (Mr. Elephant Game). The results provide strong support for the hypothesis that verbal counting and nonverbal quantitative reasoning are related in children, and help bring clarity to a conflicted body of findings about the relationship between these skills during development (e.g., Huntley-Fenner & Cannon, 2000; Mix, 1999a, 1999b; Rousselle, Palmers, & Noël, 2004). Just like adults whose language lacks an integer list (Flaherty & Senghas, 2011; Frank et al., 2008; Spaepen, Coppola, Spelke, Carey, & Goldin-Meadow, 2011), children who do not yet understand cardinality exhibit relatively coarse representations of numerical quantity when the quantity to be represented is greater than four—beyond the limit of the parallel individuation system. On small-number trials with set sizes of one, two, and three, SS-knowers performed no differently from CP-knowers, consistent with evidence for a primitive, language-independent cognitive system for representing small exact quantities (Agrillo, Piffer, Bisazza, & Butterworth, 2012; Feigenson, Carey, & Hauser, 2002; Starr, Libertus, & Brannon, 2013). However, on large-number trials with set sizes between five and nine, CP-knowers were more sensitive to the differences between quantities: In the Caterpillar Game, they brought more socks for larger sets of feet, and the number of socks they brought was closer to the target, and in the Mr. Elephant game, they more accurately identified whether the exact number of hidden objects re-appeared. Thus, mastery of verbal counting was correlated with performance on large, but not small, set sizes.

Three verbal mathematical capacities other than cardinality notably did not relate to performance on the non-verbal tasks. First, counting behavior during the non-verbal Caterpillar Game improved children's performance on the large number trials, but it did not explain the difference in success between SS- and CP-knowers. Second, verbal estimation skill within the CP-knower groups was also unrelated to performance on the Caterpillar Game (note that children were not tested on Fast Cards in Experiment 3). Third, no differences were observed across different levels of SS-knowers (e.g., "one"-knowers, "two"-knowers, etc.), with the caveat that the samples were small; as a group, SS-knowers' performance differed substantially from that of CP-knowers. Thus, after controlling for age, children's mastery of verbal counting—specifically, their understanding of cardinality—was the key predictor of accuracy on a non-verbal numerical matching task.

The current data establish the relationship between the acquisition of meanings for large numbers and non-verbal processing for quantities and numerals larger than "four." We provide developmental evidence for the cross-cultural finding that knowledge of a meaningful count list is related to more precise non-verbal representation of large numbers (Flaherty & Senghas, 2011; Frank et al., 2008; Gordon, 2004; Spaepen et al., 2011).

These findings raise two important questions: First, which cognitive systems underlie the observed change in performance on non-verbal tasks, and second, what is the causal mechanism underlying the relationship between verbal and non-verbal numerical knowledge?

Some researchers have argued that the role of number language is to provide a concept of exact number or a "tool for thought" enabling exact number representations (Frank et al., 2008), but this does not seem to be the right explanation for CP-knowers' more precise performance in this study. In the Caterpillar Game, even CP-knowers did not use exact representations of quantity to match numbers of socks to feet; rather, they tended to retrieve an approximately correct number of socks. Furthermore, both groups' near-perfect performance on the small set sizes indicates that they understood the task and its specifically numerical demands. Therefore, if CP-knowers paid more attention to numerosity, this was not a binary switch between attending or not attending to number: Both groups attended to numerosity in the small number trials, and even CP-knowers failed to engage exact number in the large number trials. Furthermore, when comparing adults from numerate cultures to adults from innumerate cultures, the reason for differences in performance seems clear: The adults in numerate cultures can count, and thus form a stable representation to assist memory for ephemeral events. The children in our study who understood counting, however, did not necessarily deploy it in the service of solving the non-verbal problems. Moreover, the advantage of CP-knowers over SS-knowers held even after taking counting behavior into account. Therefore, our findings suggest a different role of language than a "tool for thought": Acquiring language for numbers might be more deeply related to the cognitive representation of numerical quantity, or the use of these representations in computation, beyond providing an efficient tool for memory.

The observed data are consistent with several possible cognitive systems for non-verbal numerical representations that could support these patterns of performance. One explanation is that CP-knowers have better numerical acuity (e.g., Chu & Geary, 2015; Shusterman et al., 2016; Wagner & Johnson, 2011), which may help them more accurately encode and reason about the larger set sizes in the non-verbal tasks. This explanation accords with the data in some ways. CP-knowers' responses in the Caterpillar Game demonstrated two characteristics of representations in the ANS: They were approximately but not exactly correct, and their responses exhibited scalar variability (i.e., increasing standard deviations for larger targets). Thus, change in numerical acuity in the ANS could underlie changes in non-verbal numerical problem-solving in the Caterpillar Game.

However, ANS acuity is an unlikely explanation of the CP-knowers' superior performance on the Mr. Elephant Game: The most difficult trials required children to distinguish between outcomes of six and seven balls coming out on the seven-ball trials. This ratio of 1.16 is considered a very difficult discrimination ratio on other tests of numerical acuity in this age group (e.g., Halberda & Feigenson, 2008). It would therefore be surprising if differences in ANS acuity between CP- and SS-knowers drove differential performance on that task.

Differences in PI representations could also plausibly underlie the different performance between CP- and SS-knowers. In the Caterpillar Game, because the "feet" were distributed across the two sides of the caterpillar

"body," children could solve the problem *either* by noting the total set size *or* by noting the set size on each side (e.g., four on one side and five on the other on the nine-footed caterpillar). CP-knowers might have been more likely than SS-knowers to use a chunking strategy, noting the set on each side and combining them in their final response.

The distribution of the feet on either side of the caterpillar might have also affected SS-knowers ability to represent the total numerosity presented. Importantly, it is not the case that SS-knowers simply attended to one side and ignored the other; if they had ignored one side, their responses on the large number sets would have ranged from three (one side of the 6-footed caterpillar) to five (the largest number on one side on any trial, on the 9-footed caterpillar). This was not the case: SS-knowers responded with sets larger than five socks on many of the large number trials, suggesting that they were oriented to the totality of the presented set of feet.

In addition to potentially using the two sides to help break down the task into two 'chunks', CP-knowers might also have been better able to maintain exact representations of the set size on each side. Some research suggests that the set-size limit on PI increases through the preschool years from a set size limit of three in 3-year-olds to four or five in 5-year olds (Starkey & Cooper, 1995). This increased capacity could plausibly support children's performance with the seven- and nine-footed caterpillar, where children would have to remember a set of four or five on each side; a less mature system, with a limit of only three objects, would limit children's performance with these larger set sizes. Thus, a small increase in the set-size limit of the PI system could support the observed pattern in the Caterpillar Game.

The results from the Mr. Elephant Game are also consistent with this explanation: CP-knowers were much more likely to correctly assess both five-ball trials, demonstrating a true understanding of five discrete objects. SS-knowers, on the other hand, performed at chance for the sets of five. These observations strengthen the possibility that there is a relationship between an increase in the set-size limit of the PI system (to five) and the timing of children's acquisition of the cardinality principal. With much attention in recent years on the relationship between ANS acuity and symbolic mathematics, developmental change in the PI system has received little attention in recent years but may be important for children's emerging number concepts.

Although it is not considered a cognitive 'system' per se, another possible explanation for the CP-knowers' advantage is enhanced spontaneous focus on numerosity or SFON, a child's tendency to engage with quantities and number in her environment (Rathé et al., 2016). Children who intuitively pay more attention to numerosity in SFON tasks might enter more quickly into skillful counting or might be more motivated to practice counting (Hannula, Räsänen, & Lehtinen, 2007). Individual differences in SFON are consistently related to children's counting skills such as subitizing in the small number range and rote counting in the large number range (Batchelor, Inglis, & Gilmore, 2015; Edens & Potter, 2013; Hannula & Lehtinen, 2005; Hannula et al., 2007); furthermore, SFON in 3.5-year-old children is related to higher subsequent mathematical knowledge at ages 5 and 12 (Hannula & Lehtinen, 2005; Hannula-Sormunen, Lehtinen, & Räsänen, 2015). However, the current study differs from previous SFON studies in several important ways. Previous SFON studies have not distinguished performance between children who fully understand cardinality past the number "four" from those who do not. Counting ability is typically evaluated in SFON studies by procedural knowledge (e.g., Hannula et al., 2007) rather than a generalized understanding of cardinality. Additionally, the SFON tasks themselves typically focus on smaller set sizes (2, 3, or 4) than the ones used here (e.g., Hannula et al., 2007). In some studies, the participants are slightly older and may have already acquired cardinality (e.g., Batchelor et al.,

2015). Finally, few SFON studies take place in the cultural context of the U.S., where children's early number experiences may meaningfully differ. In short, development of theory related to SFON has enabled novel ways of thinking about the relationship between children's non-verbal enumeration and their verbal counting skills. The current study is the first to explore this relationship between language and thought using the SFON approach with numbers beyond the subitizing range, and to emphasize the conceptual rather than procedural development of counting and cardinality.

Notably, one recent study provides compelling evidence that the sequential-enumeration tasks used in some SFON studies likely draw on ANS representations (Sella, Berteletti, Lucangeli, & Zorzi, 2016). Nevertheless, SFON tasks should not be taken as a measure of ANS acuity: when guided, 'low-SFON' children can perform these non-verbal tasks at the same level as 'high-SFON' children (Hannula & Lehtinen, 2005), showing that SFON reflects children's self-guided attention to numerosity, not their underlying competence in numerical discrimination.

The current findings lend support to arguments for a qualitative shift, or conceptual change, between SS-Knowers and CP-Knowers. Previous research on this transition has focused on children's conceptual change in terms of their construction of a novel representation for numbers—namely, a meaningful count list (Carey, 2010; Sarnecka & Carey, 2008; Slusser & Sarnecka, 2011; Wynn, 1990, 1992). Our findings corroborate the qualitative change by showing that the acquisition of meaningful counting and cardinality is accompanied by additional changes in non-verbal reasoning.

Some previous researchers have argued that there is no semantic induction when children become CP-knowers based on the fact that CP-knowers fail to answer many questions about higher numbers within their count list (Davidson, Eng, & Barner, 2012). They posit that these limitations in CP-knowers' knowledge mean that there is no conceptual change, and no radical discontinuity between numerical representations in SS-knowers and those in CP-knowers. The current data argue against this perspective because they highlight a clear discontinuity in non-verbal representations that correlates with the discontinuity in number language. Additional evidence for a discontinuity between SS- and CP-knowers comes from Sarnecka and Wright (2013), who demonstrated that CP-knowers, but not SS-knowers, understand the principle of equinumerosity, and from Shusterman et al. (2016), who demonstrated in a longitudinal study that children's numerical acuity on a non-verbal dot-discrimination task increases right around the moment when they become CP-knowers. Drawing on all of these findings, we conclude that the transition to CP-knower status does indeed represent a major conceptual shift, and that the induction of cardinality is related to a broad suite of changes in children's representation of quantity, including a dramatic and abrupt increase in its precision.

Of course, there are limits to what children learn when they acquire the cardinal principle. Davidson et al. (2012) convincingly demonstrate that CP-knowers do not have stable number meanings 'as high as they can count,' contrary to the original suggestions of Carey, Sarnecka and others (Sarnecka & Carey, 2008) that children 'bootstrap' the meanings of all of the numbers within their count list when they become CP-knowers. Furthermore, children slowly refine the mappings between numerals and representations of quantities in the ANS (Le Corre & Carey, 2007). The acquisition of cardinality, then, might best be characterized as a 'limited semantic induction' but nevertheless a robust conceptual change. The CP-induction is important because it is in this moment that children acquire their first meanings for higher number words beyond the range of parallel individuation; because they recognize in a new way (i.e., in a way not available to SS-knowers) how the

structure of the count list imbues numbers with their meanings; and because, as we show here, they exhibit a corresponding shift in the precision of non-verbal representations for those quantities.

An additional parallel between verbal and non-verbal number concepts comes from the 20-footed caterpillar in Experiment 2. On the 20-footed caterpillar, neither group performed very accurately, and CP-knowers responded only marginally more accurately than SS-knowers. This pattern on the non-verbal task parallels the low verbal knowledge of "twenty" in both groups: "Twenty" is essentially an 'unknown' number for some CP-knowers and most SS-knowers: Although they may produce it, they often cannot reliably count to it using stable order (Shusterman & Berkowitz, 2011), and they often cannot generate sets of more than 10 objects in Give-N (Shusterman, Cheung, Sarbh, & Taggart, 2015). CP- and SS-knowers' limited verbal meaning for 20 (perhaps as something vague like *a lot*) may be related to their worse performance with this set size on the non-verbal task. Lower familiarity, distinctiveness, or acquired meaning of large number words like "twenty" (and like "five" for SS-knowers) therefore appears to be associated with less accurate performance on non-verbal tasks in which these quantities need to be represented in memory. Even after becoming CP-knowers, children clearly need to learn more about and become more familiar with higher number words. An open question is whether further development of number language, beyond cardinality, correlates with more precise numerical representations in non-verbal tasks.

Finally, we note that this study is correlational, and therefore cannot in itself address causality: Change in core non-verbal number representations may support number language; number language may induce change in cognitive systems involved in non-verbal numerical reasoning; or these changes may co-occur as part of an inter-related set of conceptual shifts. The cross-cultural findings with innumerate people imply that language causally changes the way in which quantities can be represented, and that in the absence of acquiring such language, conceptual representations are not pushed to change. Studies on the development of number concepts under conditions of accelerated language (e.g., with number training) or delayed language (e.g., due to limited access to a native language) will help to tease apart these possibilities.

To summarize, children's acquisition of exact cardinal meanings for large numbers (beyond the PI range) correlates with their performance on numerical problem-solving tasks that require remembering and matching large, exact quantities. In particular, responses to target sets between five and nine were less accurate and more variable in children who had not induced cardinality than in those who had. In contrast, smaller set sizes were handled easily by all children, regardless of knower-level. Performance on the non-verbal tasks did not vary for children at different SS-knower levels, as a function of developing better verbal estimation skills, nor as a function of counting behavior. Our findings accord with reports of close links between symbolic and non-symbolic mathematical competence in children (Libertus et al., 2011; Shusterman et al., 2016) and adults (Frank et al., 2008), and extend this conclusion by demonstrating the same result with two engaging tasks with low task demands.

These findings thus bring some clarity to a previously conflicted body of literature, by showing a distinct relationship between the acquisition of number meanings larger than "four" and non-verbal problem-solving with quantities larger than four. This pattern of results helps to explain why previous studies, which lacked the specific comparison between SS- and CP-knowers on large numbers, did not find hypothesized relationships between verbal and non-verbal number knowledge. These findings open up a new set of questions regarding

which non-verbal cognitive skills and causal mechanisms underlie the tight link observed between number language and number thought.

Notes

i) Many children brought two socks on the first trial, even though the caterpillar had just one foot, presumably because socks typically come in pairs. They were then asked to return the extra sock, which helped to reinforce the rules of the game.

ii) Given that all children could count up to 8 but not 9 on the Elicited Counting task (N = 11), we ran additional analyses excluding trials on the 9-footed caterpillar for children who could only count up to 8. The pattern of results remained almost identical. Thus, the results here include all children.

iii) Given that we had trials using '5' in the Mr. Elephant game, we took a conservative approach and did not analyze performance of the two five-knowers.

Funding

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

References

Agrillo, C., Piffer, L., Bisazza, A., & Butterworth, B. (2012). Evidence for two numerical systems that are similar in humans and guppies. *PLOS ONE, 7*(2), Article e31923. doi:10.1371/journal.pone.0031923

Bar-David, E., Compton, E., Drennan, L., Finder, B., Grogan, K., & Leonard, J. (2009). Nonverbal number knowledge in preschool-age children. *Mind Matters: The Wesleyan Journal of Psychology, 4*, 51-64.

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 86*(3), 201-221. doi:10.1016/S0010-0277(02)00178-6

Batchelor, S., Inglis, M., & Gilmore, C. (2015). Spontaneous focusing on numerosity and the arithmetic advantage. *Learning and Instruction, 40*, 79-88. doi:10.1016/j.learninstruc.2015.09.005

Brannon, E. M., & Van de Walle, G. A. (2001). The development of ordinal numerical competence in young children. *Cognitive Psychology, 43*(1), 53-81. doi:10.1006/cogp.2001.0756

Cantrell, L., & Smith, L. B. (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition, 128*(3), 331-352. doi:10.1016/j.cognition.2013.04.008

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus, 133*(1), 59-68. doi:10.1162/001152604772746701

Carey, S. (2009). Where our number concepts come from. *The Journal of Philosophy, 106*(4), 220-254. doi:10.5840/jphil2009106418

Carey, S. (2010). Beyond fast mapping. *Language Learning and Development, 6*(3), 184-205. doi:10.1080/15475441.2010.484379

Carey, S., & Sarnecka, B. W. (2006). The development of human conceptual representations: A case study. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 473-496). New York, NY, USA: Oxford University Press.

Chi, M. T., & Klahr, D. (1975). Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology, 19*(3), 434-439. doi:10.1016/0022-0965(75)90072-7

Chu, F. W., & Geary, D. C. (2015). Early numerical foundations of young children's mathematical development. *Journal of Experimental Child Psychology, 132*, 205-212. doi:10.1016/j.jecp.2015.01.006

Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General, 137*(1), 22-38. doi:10.1037/0096-3445.137.1.22

Coubart, A., Izard, V., Spelke, E. S., Marie, J., & Streri, A. (2014). Dissociation between small and large numerosities in newborn infants. *Developmental Science, 17*(1), 11-22. doi:10.1111/desc.12108

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition, 123*, 162-173. doi:10.1016/j.cognition.2011.12.013

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. New York, NY, USA: Oxford University Press.

Edens, K. M., & Potter, E. F. (2013). An exploratory look at the relationships among math skills, motivational factors and activity choice. *Early Childhood Education Journal, 41*(3), 235-243. doi:10.1007/s10643-012-0540-y

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6*(5), 568-584. doi:10.1111/1467-7687.00313

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*(3), 295-313. doi:10.1016/j.cognition.2004.09.010

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science, 13*(2), 150-156. doi:10.1111/1467-9280.00427

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307-314. doi:10.1016/j.tics.2004.05.002

Flaherty, M., & Senghas, A. (2011). Numerosity and number signs in deaf Nicaraguan adults. *Cognition, 121*(3), 427-436. doi:10.1016/j.cognition.2011.07.007

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Piraha language and cognition. *Cognition, 108*(3), 819-824. doi:10.1016/j.cognition.2008.04.007

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York, NY, USA: Springer.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1-2), 43-74. doi:10.1016/0010-0277(92)90050-R

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA, USA: Harvard University Press.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*(5695), 496-499. doi:10.1126/science.1094492

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457-1465. doi:10.1037/a0012682

Hannula, M. M., & Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learning and Instruction, 15*(3), 237-256. doi:10.1016/j.learninstruc.2005.04.005

Hannula-Sormunen, M. M., Lehtinen, E., & Räsänen, P. (2015). Preschool children's spontaneous focusing on numerosity, subitizing, and counting skills as predictors of their mathematical performance seven years later at school. *Mathematical Thinking and Learning, 17*(2-3), 155-177. doi:10.1080/10986065.2015.1016814

Hannula, M. M., Räsänen, P., & Lehtinen, E. (2007). Development of counting skills: Role of spontaneous focusing on numerosity and subitizing-based enumeration. *Mathematical Thinking and Learning, 9*(1), 51-57. doi:10.1080/10986060709336605

Hauser, M. D., & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: Examinations of content and format. *Cognitive Psychology, 47*, 367-401. doi:10.1016/S0010-0285(03)00050-1

Huntley-Fenner, G., & Cannon, E. (2000). Preschooler's magnitude comparisons are mediated by a preverbal analog mechanism. *American Psychological Science, 11*(2), 147-152. doi:10.1111/1467-9280.00230

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*(2), 395-438. doi:10.1016/j.cognition.2006.10.005

Le Corre, M., Van de Walle, G., Brannon, E. M., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology, 52*(2), 130-169. doi:10.1016/j.cogpsych.2005.07.002

Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behavior in number concept development. *Cognitive Science, 34*(1), 51-67. doi:10.1111/j.1551-6709.2009.01063.x

Lee, M. D., & Sarnecka, B. W. (2011). Number-knower levels in young children: Insights from Bayesian modeling. *Cognition, 120*(3), 391-402. doi:10.1016/j.cognition.2010.10.003

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*(6), 1292-1300. doi:10.1111/j.1467-7687.2011.01080.x

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica, 141*(3), 373-379. doi:10.1016/j.actpsy.2012.09.009

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense large-number discrimination in human infants. *Psychological Science, 14*(5), 396-401. doi:10.1111/1467-9280.01453

Lipton, J. S., & Spelke, E. S. (2004). Discrimination of large and small numerosities by human infants. *Infancy, 5*(3), 271-290. doi:10.1207/s15327078in0503_2

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLOS ONE, 6*(9), Article e23749. doi:10.1371/journal.pone.0023749

McMullen, J. A., Hannula-Sormunen, M. M., & Lehtinen, E. (2013). Young children's recognition of quantitative relations in mathematically unspecified settings. *The Journal of Mathematical Behavior, 32*(3), 450-460. doi:10.1016/j.jmathb.2013.06.001

McMullen, J. A., Hannula-Sormunen, M. M., & Lehtinen, E. (2014). Spontaneous focusing on quantitative relations in the development of children's fraction knowledge. *Cognition and Instruction, 32*(2), 198-218. doi:10.1080/07370008.2014.887085

Mix, K. S. (1999a). Similarity and numerical equivalence: Appearances count. *Cognitive Development, 14*(2), 269-297. doi:10.1016/S0885-2014(99)00005-2

Mix, K. S. (1999b). Preschoolers' recognition of numerical equivalence: Sequential sets. *Journal of Experimental Child Psychology, 74*(4), 309-332. doi:10.1006/jecp.1999.2533

Mix, K. S. (2008a). Children's equivalence judgments: Crossmapping effects. *Cognitive Development, 23*, 191-203. doi:10.1016/j.cogdev.2007.03.001

Mix, K. S. (2008b). Surface similarity and label knowledge impact early numerical comparisons. *British Journal of Developmental Psychology, 26*, 13-32. doi:10.1348/026151007X189109

Mix, K. S., Huttenlocher, J., & Levine, S. C. (1996). Do preschool children recognize auditory-visual numerical correspondences? *Child Development, 67*, 1592-1608. doi:10.2307/1131720

Negen, J., & Sarnecka, B. W. (2010). Analogue magnitudes and knower-levels: Re-visiting the variability argument. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1252-1257). Austin, TX, USA: Cognitive Science Society.

Odic, D., Le Corre, M., & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition, 138*, 102-121. doi:10.1016/j.cognition.2015.01.008

O'Hearn, K., Hoffman, J. E., & Landau, B. (2011). Small subitizing range in people with Williams syndrome. *Visual Cognition, 19*(3), 289-312. doi:10.1080/13506285.2010.535994

Piazza, M., Pinel, P., LeBihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron, 53*, 293-305. doi:10.1016/j.neuron.2006.11.022

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian Indigene group. *Science, 306*(5695), 499-503. doi:10.1126/science.1102085

Rathé, S., Torbeyns, J., Hannula-Sormunen, M. M., De Smedt, B., & Verschaffel, L. (2016). Spontaneous focusing on numerosity: A review of recent research. *Mediterranean Journal for Research in Mathematics Education, 15*, 1-25.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development, 74*(6), 1807-1822. doi:10.1046/j.1467-8624.2003.00639.x

PsychOpen
publishing psychology

Rousselle, L., Palmers, E., & Noël, M. P. (2004). Magnitude comparison in preschoolers: What counts? Influence of perceptual variables. *Journal of Experimental Child Psychology, 87*(1), 57-84. doi:10.1016/j.jecp.2003.10.005

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*(3), 662-674. doi:10.1016/j.cognition.2008.05.007

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology, 103*(3), 325-337. doi:10.1016/j.jecp.2009.02.007

Sarnecka, B. W., & Wright, C. E. (2013). The idea of an exact number: Children's understanding of cardinality and equinumerosity. *Cognitive Science, 37*(8), 1493-1506. doi:10.1111/cogs.12043

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114*(3), 418-431. doi:10.1016/j.jecp.2012.10.012

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology, 6*, 357-379. doi:10.1016/0010-0285(74)90017-6

Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2016). Spontaneous non-verbal counting in toddlers. *Developmental Science, 19*(2), 329-337. doi:10.1111/desc.12299

Shusterman, A., & Berkowitz, T. (2011). The development of number concepts in oral-deaf preschoolers. Poster presented at the Biennial Meeting of the Cognitive Development Society, Philadelphia, PA, USA.

Shusterman, A., Cheung, P., Sarbh, S., & Taggart, J. (2015). *Limitations in children's induction of the cardinality principle: Evidence from the Give-N task with larger quantities.* Poster presented at the Cognitive Development Society, Columbus, OH, USA.

Shusterman, A., Slusser, E., Halberda, J., & Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLOS ONE, 11*(4), Article e0153072. doi:10.1371/journal.pone.0153072

Slusser, E. B., & Sarnecka, B. W. (2011). Find the picture of eight turtles: A link between children's counting and their knowledge of number-word semantics. *Journal of Experimental Child Psychology, 110*, 38-51. doi:10.1016/j.jecp.2011.03.006

Spaepen, E., Coppola, M., Spelke, E. S., Carey, S. E., & Goldin-Meadow, S. (2011). Number without a language model. *Proceedings of the National Academy of Sciences of the United States of America, 108*(8), 3163-3168. doi:10.1073/pnas.1015975108

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science, 210*(4473), 1033-1035. doi:10.1126/science.7434014

Starkey, P., & Cooper, R. G. (1995). The development of subitizing in young children. *British Journal of Developmental Psychology, 13*(4), 399-420. doi:10.1111/j.2044-835X.1995.tb00688.x

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Infants show ratio-dependent number discrimination regardless of set size. *Infancy, 18*(6), 927-941. doi:10.1111/infa.12008

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*(1), 80-102. doi:10.1037/0033-295X.101.1.80

Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition, 119*(1), 10-22. doi:10.1016/j.cognition.2010.11.014

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*, 130-137. doi:10.1111/1467-9280.00120

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*(2), 155-193. doi:10.1016/0010-0277(90)90003-3

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology, 24*(2), 220-251. doi:10.1016/0010-0285(92)90008-P

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1-B11. doi:10.1016/S0010-0277(99)00066-9