Title
Inference and Association in Children's Early Numerical Estimation

Authors
Sullivan, Jessica
Barner, David

Publication Date
2014

DOI
10.1111/cdev.12211

CHILD DEVELOPMENT

Child Development, xxxx 2014, Volume 00, Number 0, Pages 1–16

Inference and Association in Children's Early Numerical Estimation

Jessica Sullivan and David Barner
University of California, San Diego

How do children map number words to the numerical magnitudes they represent? Recent work in adults has shown that two distinct mechanisms—structure mapping and associative mapping—connect number words to nonlinguistic numerical representations (Sullivan & Barner, 2012). This study investigated the development of number word mappings, and the roles of inference and association in children's estimation. Fifty-eight 5- to 7-year-olds participated, and results showed that at both ages, children possess strong item-based associative mappings for numbers up to around six, but rely primarily on structure mapping—an inferential process—for larger quantities. These findings suggest that children rely primarily on an inferential mechanism to construct and deploy mappings between number words and large approximate magnitudes.

How do number words like *seven* and *thirty-five* get linked to sets of things in the world? When children make these mappings, do they form item-by-item associations between numerals and quantities? Or do they use inferential processes based on their knowledge of how numbers are related to one another? Although philosophical discussions of mathematical knowledge provide strong reasons to doubt that perception alone could supply the logical meanings encoded by number words (e.g., Frege, 1884/1953; Kant, 1781/1929), experimental psychologists have established that once children acquire such meanings, number words do eventually get mapped to perceptual representations of quantity (e.g., Carey, 2009; Gelman & Gallistel, 1978; Le Corre & Carey, 2007; Siegler & Opfer, 2003; Wynn, 1990, 1992). Surprisingly little is known, however, about the mechanisms by which linguistic and nonlinguistic representations of number become linked during development, and thus what roles inference and association play in this process. Here, we explored this question, and asked how young children begin to map number words onto perceptual representations of quantity.

Humans and nonhumans alike have access to an approximate number system (ANS) for representing numerical content (for review, see Dehaene, 1997). Consistent with Weber's law, this system supports the nonverbal comparison of sets on the basis of their numerical ratio. For example, sets that stand in a 1:2 ratio are equally easy to discriminate as whether the sets are relatively small (e.g., 5 vs. 10) or relatively large (e.g., 500 vs. 1,000; Barth, Kanwisher, & Spelke, 2003; Brannon & Terrace, 2000; Feigenson, Dehaene, & Spelke, 2004; Whalen, Gallistel, & Gelman, 1999). The ANS is used to represent numerical content in early infancy, and by at least 6 months of age, infants can reliably discriminate quantities at a 1:2 ratio (e.g., Xu & Spelke, 2000). The acuity of this system grows slowly over development well into the teenage years, and typically converges on a ratio of about 7:8 or higher in adults (Halberda & Feigenson, 2008; Halberda, Mazzocco, & Feigenson, 2008).

The ANS eventually becomes linked to the verbal number word system, probably sometime after children begin to acquire the logical meanings of these words (Le Corre & Carey, 2007). Evidence for this comes primarily from studies of estimation. In typical dot array estimation experiments, participants see a series of rapidly flashed dot arrays, and then label each array with a number word (Atkinson, Francis, & Campbell, 1976; Barth, Starr, & Sullivan, 2009; Frank & Barner, 2012; Huntley-Fenner, 2001; Izard & Dehaene, 2008; Le Corre & Carey, 2007; Lipton & Spelke, 2005; Mundy & Gilmore, 2009). As the number of items in an array grows, the degree of error in the participant's

Thanks to members of the LAD Lab for comments on earlier versions of this manuscript and to Karen Bejar, Amanda Chamberlain, Michael Sim, and Shirlene Wade for substantial help in gathering and coding these data. This work was supported by a Jacobs Graduate Fellowship and an NSF Graduate Research Fellowship to J.S., and by a grant to D.B. from the James S. McDonnell Foundation.

Correspondence concerning this article should be addressed to Jessica Sullivan, Department of Psychology #0109, University of California, San Diego, La Jolla, CA 92093. Electronic mail may be sent to jsulliva@ucsd.edu

 0009-3920/2014/xxxx-xxxx
DOI: 10.1111/cdev.12211

estimate grows too (e.g., Whalen et al., 1999). Critically, this pattern of error, like that found in studies of nonverbal numerical comparison, can be described by Weber's law, suggesting that numerical estimation requires linking number language to the ANS (for related evidence, see Dehaene, 1989; Moyer & Landauer, 1967; see also Vul, Barner, & Sullivan, 2013). Evidence for such a link is found not only in adults but also in children as young as 4 years of age (Holloway & Ansari, 2009; Huntley-Fenner, 2001; Le Corre & Carey, 2007). By this age, but probably not before, children have constructed some sort of rudimentary mapping between number words and ANS representations. Once children have begun to form these mappings, their estimates remain relatively inaccurate for several years, but improve slowly over development (Barth et al., 2009; Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Booth & Siegler, 2008; Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008; Lipton & Spelke, 2005; Mundy & Gilmore, 2009; Siegler & Opfer, 2003). Little is known, however, about the learning mechanisms that underlie the formation and refinement of mappings between the ANS and the verbal number system, and thus what causes these changes in estimation ability.

In studies of adults, two mechanisms have been proposed to explain how the verbal and nonverbal number systems might become linked: associative mapping (AM), an associative mechanism (see Lipton & Spelke, 2005), and structure mapping (SM), a mechanism that relies on structurally mediated inference (Carey, 2009; Gentner, 1983; Gentner & Namy, 2006; Sullivan & Barner, 2012). AM involves the creation of item-by-item associations between particular number words and the magnitudes they represent. For example, for a word like *twenty,* the creation of AMs involves associating the word *twenty* with a nonverbal (ANS) representation of approximately 20, via experience in the world (e.g., "20 students," "20 crackers," "20 minutes," etc.) and by *not* associating it with sets of discriminably different magnitudes (e.g., "twenty" does not apply to "10 students," "40 crackers," or "60 minutes").

Structure mapping, in contrast, involves creating a single link between the count list and numerical representations in the ANS. This link is formed on the basis of the similarity of the structures of these two systems, rather than on associations between particular numbers and sets of things in the world. For example, according to the SM hypothesis, knowledge that the number word *forty* comes after the word *twenty* in the count list guarantees that *forty* will always be mapped to larger quantities than *twenty*. This is because the count list, like the ANS, is an ordinal system of representation, and is structured such that the ordering of its symbols is predictive of the ordering of approximate magnitude representations in the ANS. Furthermore, not only are representations in each system ordered but also the distance between symbols in the count list is predictive of the distance between the magnitudes that they encode. Thus, a mature SM might reflect not only the relative ordering of numbers in both systems but also their relative distance (e.g., that *forty* is twice as far into the count list as *twenty* and therefore should be mapped onto a set that is twice as large as the one labeled by *twenty*). Given these properties, according to SM, the mappings for number words are defined in relation to one another, and do not require a veridical link between, for example, *forty* and "40 things" (for a related discussion of anchoring and adjustment, see Tversky & Kahneman, 1974). Crucially, the content of each number word in a system created through SM is dependent on the content of all other number words.

To understand the predictions that AM and SM make, consider how these mechanisms might be recruited during estimation. By the AM hypothesis, each number word is associated with a particular state of the ANS, such that mappings are independent of one another. This predicts that adjustments to one mapping (e.g., via feedback) should not automatically influence other mappings: Misleading feedback about number word mappings should cause local—but not global—shifts in estimation performance. Also, on average, estimates of discriminably different magnitudes should be reliably different because these magnitudes will almost always be mapped to different number words (e.g., estimates for sets of 50 and 100 should typically differ).

In contrast, according to the SM hypothesis, number words become related to magnitudes via a global mapping between two structures (the verbal and nonverbal number systems), and estimates are based on inferences about the ordering and distance between number words and magnitudes. Thus, number word mappings are nonindependent and mutually constraining, such that an estimate for one set should constrain estimates for all other numbers. When a participant's estimate for a given quantity is changed via feedback (i.e., calibrated), other estimates should also change correspondingly. Also, depending on contextual factors and feedback (Izard & Dehaene, 2008; Sullivan & Barner, 2012; Sullivan, Juhasz, Slattery, & Barth, 2011), discriminably different magnitudes may not always be mapped onto different number words (e.g., a participant might provide the same estimate for a set of 50 in one context as they do for 100 in another context).

There is evidence that adults use both AM and SM to support estimation. In one recent study (Sullivan & Barner, 2012), adults were asked to match a number word to one of two discriminably different dot arrays (e.g., map the word *fifty* to either 50 dots or 100 dots). Participants succeeded when the magnitude of the arrays was relatively small (10 vs. 20), but performed substantially worse for larger comparisons (e.g., 50 vs. 100) although these comparisons differed by the same numerical ratio. Importantly, these same participants showed no effect of set size when asked to perform a numerical discrimination task, in which they judged which of these two dot arrays was larger. This pattern of findings resulted in an interaction between magnitude and task: Subjects easily discriminated sets regardless of their magnitude on the discrimination task, but their success at matching number words to one of these two sets was mediated by magnitude (i.e., they were much better at mapping words to sets for smaller comparisons relative to larger ones). This result suggests that participants relied heavily on AM for small number words, but much less for larger numbers.

In a second set of tasks (Sullivan & Barner, 2012), adults made estimates of dot arrays after receiving misleading information about the largest quantity they would see in the experiment (e.g., being told that the largest set they would see is 750 when it was actually 375). Relative to a baseline estimation task, estimates made after this miscalibration were shifted for all but the smallest numbers, resulting in an interaction between magnitude and task (calibrated vs. uncalibrated estimation). Together, these two sets of tasks provided strong evidence that adults (a) relied primarily on AM for smaller number word mappings, (b) had weaker AMs for larger magnitudes, and (c) relied more heavily on SM for larger magnitudes. These findings are consistent with other studies of estimation, which have also found that information about one mapping can affect estimates for other number words (Izard & Dehaene, 2008; Shuman Sullivan et al., 2011).

Although these adult studies establish that both inference and association are deployed by mature estimators, they leave open how such mappings are constructed, and how AM and SM interact in early development. As a result, it is currently unknown how the mapping process begins, and whether children use inference to guide estimation from the beginning, or are initially restricted to item-based associative learning.

One previous study found evidence that children can recruit SM when estimating, but only tested older children who were already competent estimators (Thompson & Opfer, 2010). In their study, Thompson and Opfer (2010) trained elementary-school-aged children to recognize the structural relation between a familiar number range (e.g., 0–100) and an unfamiliar number range (e.g., 0–10,000). This training improved their estimation performance for the unfamiliar number range, suggesting that children used their structural knowledge of the smaller number range to guide their interpretation of the larger numbers. This suggests that when children understand how a particular part of the count list is structured (e.g., the ordering and distance between numerals), they can use SM to guide their estimates. However, the study did not test the mechanism children use to make spontaneous, untrained estimates, leaving open the possibility that even elementary-school-aged children do not typically recruit SM to guide estimates. Also, it did not test how SM emerges earlier in development, and how it relates to AM—that is, whether children initially rely more on AM or use SM from early in development.

In this study, we investigated how children construct early mappings between number language and the ANS, by testing whether they begin by relying primarily on item-based associations, or instead use structure-based inferences. We also investigated whether children's reliance on these two mechanisms changes during development, as their estimation abilities improve (Siegler & Booth, 2004). We see several possible ways that these systems could interact to support estimation. On the one hand, it is possible, as argued by Lipton and Spelke (2005), that children initially rely heavily on AMs (e.g., perhaps by forming AMs for magnitudes up to 50 or 100). By this view, early mappings might be formed exclusively via associative item-specific learning, but slowly give way to SM as children learn the global relation between counting and numerical magnitude. Alternatively, it is possible that children initially rely primarily on SM to guide estimation, and acquire AMs slowly, as they gain experience with estimation (see Izard & Dehaene, 2008, for a model that posits only SM). A final possibility is that children acquire a small set of strong AMs—for example, up to 10 or 12, like adults—before beginning to use SM for estimation.

Although many previous studies have investigated the mechanisms by which children acquire the logical meanings of number words (beginning with *one*, *two*, *three*, and then learning the principles that govern counting; see Carey, 2009, for review), no previous study has distinguished between the mechanisms that might guide children's mappings between number words and perception. Understanding the mechanisms that guide the formation of these mappings is important for several reasons. First, estimation performance is known to be linked to a host of educational outcomes (Booth & Siegler, 2008; Siegler & Booth, 2004; Siegler & Ramani, 2009), so understanding the learning mechanisms that allow children to become successful estimators is vital to understanding the factors that drive math success. Second, estimation performance is known to improve dramatically over development (e.g., Barth & Paladino, 2011; Booth & Siegler, 2008; Siegler & Opfer, 2003; Slusser, Santiago, & Barth, 2013), yet it remains unknown whether such improvements are guided primarily by improvements to nonverbal number knowledge, verbal number knowledge, or to the mappings between them. Finally, and perhaps most importantly, testing the roles of AM and SM in supporting estimation provides a test case for investigating the fundamental question of how inference and association combine to link language to the content of the world.

In this study, we used four within-subjects computerized measures to assess children's knowledge of number word mappings. These measures explicitly tested the roles of AM and SM in guiding verbal estimation and were adapted from previous adult studies of these learning mechanisms (Sullivan & Barner, 2012). When considered together, these tasks address the nature of children's early mappings by testing the roles of AM and SM in estimation. We also used two noncomputerized assessments of verbal number knowledge.

First, we assessed whether mappings for particular number words were formed via strong AMs. To do this, participants completed a pair of tasks requiring judgments about sequentially presented arrays of dots. In the discrimination task, participants decided which of two sets contained more dots. Performance on this task, which only required nonverbal number knowledge, served as a baseline for our critical measure, the number matching task. In this task, participants decided which of the two presented sets matched a particular number word (a judgment that requires mapping the verbal number system to ANS representations of number). By comparing performance on the number matching task to the baseline performance on the discrimination task, we assessed whether participants reliably map particular number words (e.g., *fifty*) to one of two discriminably different sets (e.g., 50 vs. 100). This allowed us to test the degree to which number words had strong AMs. For magnitudes that are mapped via strong AMs, performance should be relatively high on the number matching task, as visual arrays should activate their appropriate verbal labels directly. However, to the degree that children lack strong AMs, they should be less accurate at these judgments, even for highly discriminable quantities. Critically, if children—like adults—rely more on AM for small numbers than for larger numbers, we should find that performance on the number matching task remains relatively high for small number words (relative to the discrimination task), and declines as a function of numerical magnitude for larger number words, resulting in a significant interaction between task and magnitude (for discussion, and similar data from adults, see Sullivan & Barner, 2012).

In a second set of tasks, we more explicitly assessed the role of SM in guiding early estimation performance by testing whether misleading feedback influenced estimation performance. Participants completed two estimation tasks. In the first task, they were given no feedback about the range of numbers being estimated (uncalibrated estimation); in the second task, an experimenter provided a single instance of calibration by telling participants the magnitude of the largest set that they would see during the experiment (calibrated estimation). We asked whether children's estimates were influenced by misleading feedback and therefore whether they recruited SM to support their estimates. We did this by comparing performance on the uncalibrated estimation task to performance on the calibrated estimation task. If participants rely primarily on AM for a given quantity, then calibration should have no effect, as AMs are independent from one another. However, calibration should have a significant effect on estimation performance if mappings rely strongly on SM, as this would indicate the use of a single data point to realign the entire SM. Again, if AMs are relatively strong for small numbers but decline in strength and give way to SMs for larger numbers, we should see no effect of calibration for the smallest numbers estimated, and a larger effect of calibration for larger numbers.

In a third set of tasks, we assessed children's verbal number knowledge to determine how the use of different estimation strategies in the first tasks was related to children's knowledge of the count list. This was of interest as, according to past studies, counting ability predicts certain aspects of estimation performance (Barth et al., 2009; Davidson, Eng, & Barner, 2012; Lipton & Spelke, 2005). There were two tasks. First, participants were given two counting assessments—a free count exercise ("count to 100") and a scaffolded counting exercise ("finish this counting sequence"). The second task tested children's understanding of the relative ordering of number words. Participants decided which of two boxes contained more stickers after hearing, for example, that one box contained *twenty* stickers and one box contained *forty* stickers. Children who possess a strong understanding of the order and structure of the count list should succeed at this task, whereas those who do not will likely fail (e.g., Davidson et al., 2012; Le Corre, 2013).

Method

Participants

Thirty-two 5-year-olds (range = 5;0–5;11) and twenty-six 7-year-olds (range = 7;0–7;11) participated. Participants were recruited from a database of interested families maintained by the Psychology Department at UCSD, and were compensated for travel expenses and given a small prize. Participants lived in the greater San Diego area, and were primarily Caucasian and upper middle class. Of the 58 participants tested, 5 were excluded: for failure to complete at least 10 trials on any task in the experiment ($n = 2$), due to inattention ($n = 2$), or due to computer error ($n = 1$). Data from the 7 participants who successfully completed at least one critical pair of tasks (e.g., number matching and discrimination; estimation and calibrated estimation; counting and verbal ordering) but failed to complete all tasks were included in the relevant groupwise analyses. The remaining 46 participants (21 boys) contributed a full data set (twenty-three 5-year-olds ranging in age from 5;0 to 5;11 and twenty-three 7-year-olds ranging in age from 7;1 to 7;11) and were included in all analyses.

Procedure

Participants were tested in a quiet laboratory setting, after written parental consent and the child's verbal assent were secured. Participants were seated approximately 40 cm from a 27-in. Mac OSX computer screen while they completed four computerized tasks and two noncomputerized tasks. The total testing time was approximately 1 hr, and participants were offered breaks between each game. Participants completed the discrimination task first, the number matching task second, the uncalibrated estimation task third, and the calibrated estimation task fourth. For all computer tasks, the presentation time for stimuli was brief (400 ms), preventing the use of counting. The majority of participants completed both noncomputerized tasks (counting and verbal ordering) after completing all four computer tasks; however, some participants completed one or both of these tasks as part of a break between computer tasks. Because task order was fixed (to prevent, e.g., performance on the calibrated estimation task from influencing performance on the other computerized tasks), it was not possible to test whether task order influenced performance. However, performance on the discrimination and number matching tasks could not have been influenced by the noncomputerized tasks (because they always came before these tasks), and there was no qualitative evidence of an influence of noncomputerized task order on estimation (e.g., children who did the counting task before estimating were not more likely to provide sequential estimates on the estimation task than children who did the counting task after).

Numerical Discrimination

Participants saw two sets of dots and had to decide which set had more dots. The arrays were presented on a black background and were flashed sequentially for 400 ms, and each set was backward masked with random noise for 100 ms. On each trial, sets differed in numerical magnitude by a 1:2 ratio. Trials were presented in a fixed random order. There were 20 possible comparisons (see Figure 1), and each participant saw each comparison up to four times. Sets were matched for item size on half of the trials and for total area on the other half (MatLab code; Dehaene, Izard, & Piazza, 2005), and comparisons ranged from small (3 vs. 6) to large (300 vs. 600). The task ended after the participant saw all 80 trials, or after 10 min. The color of the dots was constant within a given trial, but varied across trials (colors included red, blue, yellow, white, cyan, green, and magenta). On half of the trials, the first set contained more dots. Participants chose the set they thought was larger either by saying "first set" or "second set" or, for older children, by entering "1" or "2" onto the numerical keypad. This task functioned as a within-subjects control for the number matching task to ensure that participants could discriminate the quantities presented.

Number Matching

The stimuli and procedures were identical to those used in the numerical discrimination task, except that participants were instructed to match a given number word to one of two visually presented arrays. A numeral appeared in black font on a gray background before the onset of the trial, and the experimenter read the numeral out loud if the child did not spontaneously do so. Only after the child heard the number word did the experimenter show the arrays. On half of the trials, the numeral matched the larger of the two sets; also, on half of the trials it matched the first set presented.

Uncalibrated Estimation

On both this and the calibrated estimation task, participants saw a single dot array on each trial and were asked to estimate how many dots they saw. All participants provided their responses verbally, and an experimenter entered them using a numeric keypad. Stimuli were sets of dots on a black screen and were taken from the pool of arrays presented in the discrimination and number matching tasks. Twenty-six numerosities were presented up to three times each: 3, 4, 5, 6, 8, 10, 12, 16, 20, 24, 30, 32, 40, 48, 50, 60, 75, 80, 100, 120, 150, 200, 240, 300, 480, and 600. Arrays were presented in one of two fixed random orders. The task ended after the participant saw all 78 trials, or after 10 min.

Calibrated Estimation

Stimuli and instructions were identical to those in the uncalibrated estimation task, except that participants were first told the size of the largest set they would see (Sullivan & Barner, 2012). Although the largest set that participants saw was 600 in all conditions, they were randomly assigned to be told that the largest set they would see was 25, 75, or 750 (with approximately one third of the children in each age group assigned to each calibration condition). Although the degree of miscalibration here may seem extreme, our previous work suggests that it goes unnoticed by naïve subjects. In our adult study of calibrated estimation (Sullivan & Barner, 2012), participants who were asked whether they thought that calibration was misleading rarely reported that it was (despite sometimes being calibrated to expect a maximum of 75, when the actual maximum was 350). In this study, only one participant (a 7-year-old) thought that the calibration was "silly."

Counting Assessment

Participants were asked to guess how high they could count. For the free count assessment, they then counted as high as they could. Each child was encouraged to continue counting until they reached 100 (e.g., Barth et al., 2009; Davidson et al., 2012), or until they made eight errors, all of which were recorded by the experimenter. Participants were then given an additional scaffolded counting assessment (e.g., Lipton & Spelke, 2005). For this task, the child was asked to finish counting a sequence that an experimenter started—for example, "7, 8, 9—what comes next?" All participants successfully said "ten," and then completed the remainder of the counting assessment, which included: 16, 17, 18, ___, ___; 48, 49, ___, ___; 97, 98, ___, ___; 247, 248, ___, ___; 296, 297, ___, ___, ___; 447, 448, ___, ___, ___; 498, 499, ___, ___; 997, 998, 999, ___.

Verbal Ordering Assessment

Children's knowledge of the ordering of the number words was assessed in this task. Participants were shown two toy gift boxes and were told that there were stickers inside each box. The experimenter then said, "This box [pointing to box on left] has X stickers in it, and this box [pointing to box on right] has Y stickers in it. Which box has more stickers?" where X and Y were replaced with two number words that differed by a 1:2 or 3:4 ratio. The child was then asked to point to the box that contained more stickers. Participants completed 16 trials involving numbers ranging from 4 to 700—on half of the trials, the box on the left had more stickers. Trials were presented in one of two fixed random orders.

Results

Number Matching and Discrimination

Here, we asked three main questions about the role of AM in children's estimation. First, is there evidence that children have AMs for at least some number words? Second, how high do strong AMs extend? Finally, how does children's use of AM change over time (e.g., what changes between 5 and 7 years of age, as children become competent estimators)?

To test the role of AM in supporting children's estimation, we asked whether performance on the number matching task declined significantly as a function of the numerical magnitude being tested, relative to performance on the discrimination task. Although it was possible that the number matching task would be generally more difficult than the discrimination task due to the increased demands it places on subjects, the key test of our hypothesis was whether performance on the number matching task interacted with magnitude, as found in our previous study of adults (Sullivan & Barner, 2012). In this study, the ratio between magnitudes on both tasks was always held constant at a 1:2 ratio. Therefore, any interaction due to magnitude can only be explained by the relative strength of AMs for numbers of different sizes.

For these analyses, we constructed a binomial logit model using the lmer package in R (Bates & Sarkar, 2007; R Development Core Team, 2010), predicting task accuracy from task (number matching or discrimination), the smaller magnitude in any given comparison (e.g., "3" for the comparison 3 vs. 6), and their interaction. For this and all other linear mixed models (LMMs), subject was considered a random factor. In 5-year-olds, we found an effect of task ($\beta = 1.1$, $SE = .09$, $p < .0001$), an effect of magnitude ($\beta = -0.001$, $SE = .0007$, $p < .01$), and no interaction ($\beta = -0.0008$, $SE = .0009$, $p > .05$). Overall, participants performed worse on the number matching task than on the discrimination task, and performance on both tasks declined as numerical magnitude increased. In 7-year-olds, we also found an effect of task ($\beta = 1.94$, $SE = .12$, $p < .0001$), and an effect of magnitude ($\beta = -0.004$, $SE = .0007$, $p < .0001$). Critically, in these older children, an interaction of task and magnitude also emerged ($\beta = -0.0023$, $SE = .001$, $p < .05$). Thus, like the 5-year-olds, 7-year-olds performed worse on the number matching task than the discrimination task, and worse on larger magnitudes than smaller magnitudes. Also, like in adult populations (Sullivan & Barner, 2012), the effect of magnitude was mediated by task: Performance declined steeply as a function of numerical magnitude for the number matching task, whereas performance on the discrimination task did not (see Figure 1).

From these data, we can conclude that, at least by the age of 7, children, like adults, rely strongly on AM for relatively small numbers and rely much less on AM for larger numbers. The results for the 5-year-olds are somewhat more difficult to interpret. Without a significant interaction between magnitude and task, it is difficult to determine whether 5-year-olds recruit AM when estimating. The first possible explanation of this finding is that 5-year-olds—like adults and 7-year-olds—displayed a magnitude-based decline in accuracy on the number matching task (relative to the discrimination task), but that this trend was not detected by our statistical test. This would predict that—despite a lack of interaction—a direct comparison of performance on the number matching task to the discrimination task for each comparison presented (e.g., 3 vs. 6, 4 vs. 8, … 300 vs. 600) should reveal no difference in task performance for smaller comparisons, and large differences in performance for larger comparisons. A second possibility is that children relied strongly on AM for both small and large numbers, such that analyses of individual comparisons should reveal no difference in performance on the discrimination and the number matching task for all comparisons tested. A third possibility is that children lacked AMs for any of the magnitudes we tested. If this were the case, large differences between performance on the discrimination and number matching tasks should emerge for all comparisons tested.

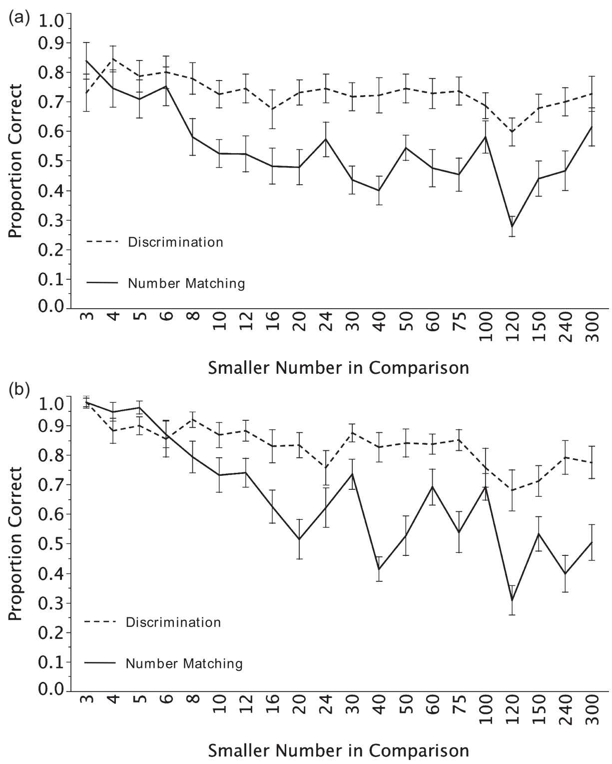

Figure 1. Proportion correct on the number matching task (solid line) and discrimination task (dashed line) for (a) 5-year-olds and (b) 7-year-olds. All comparisons were presented at a 1:2 ratio. The smaller number in each comparison is represented on the *x*-axis (e.g., "3" denotes the comparison of 3 vs. 6). Error bars denote the SEM.

To assess these possibilities, we used Dunnet's mean comparison—an analysis that corrects for multiple comparisons—to test whether accuracy differed on the number matching task relative to the discrimination task for each comparison tested. For 5-year-olds, accuracy on these two tasks did not differ for any comparison containing six or fewer items (3 vs. 6, 4 vs. 8, 5 vs. 10, 6 vs. 12; all *p*s > .1), but did differ for 13/16 of the larger comparisons (all *p*s < .05; the trials on which there was no difference were as follows: 24 vs. 48, $p = .052$; 100 vs. 200, $p = .23$; 300 vs. 600, $p = .24$). This provides some evidence that 5-year-olds have strong AMs for numbers up to about six. For 7-year-olds, the pattern of performance was very similar. They showed no difference in accuracy on the number matching task relative to the discrimination task for the four smallest comparisons tested (all *p*s > .05), but they performed significantly worse on the number matching task than on the discrimination task on 13/16 of the larger comparisons (all *p*s < .05 except for: 24 vs. 48, $p = .091$; 60 vs. 120, $p = .078$; 100 vs. 200, $p = .28$). Taken together, these data suggest that 5- and 7-year-olds have strong AMs for numbers up to around six, and have much weaker AMs for larger numbers.

An alternative explanation of these data is that children have AMs for many numbers, but that these AMs are highly inaccurate for all but the smallest numbers. This would cause children to perform poorly on the number matching task, especially on large numbers, where a bias to underestimate has been found in both children and adults (Izard & Dehaene, 2008; Siegler & Opfer, 2003). In past studies, such alternative explanations have been ruled out for adults (Sullivan & Barner, 2012). However, to test for this possibility in children, we asked whether subjects showed evidence of over- or underestimation during the task. If children underestimated the quantities presented, then they should have performed better on trials where they saw 20 versus 40 dots and were asked to find *forty* (foil magnitude is smaller) than those where they are asked to find *twenty* (foil magnitude is larger). This is because if the participant erroneously thinks that each set contains fewer items than it actually does, then when forced to find *forty*, they should strongly prefer the larger quantity (i.e., the correct response, as the smaller quantity will appear even smaller than it actually is, making it a very implausible choice). In contrast, when asked to find *twenty*, this same participant should often get the trial wrong, and map *twenty* to the larger of the two sets.

To test this, we compared performance on trials where the foil magnitude was larger than the target to those where it was smaller for each child, to determine whether they had an underestimation bias. Next, we classified each child as either showing an underestimation bias or not (there were no overestimators in our population, unlike in adults; Sullivan & Barner, 2012): A child was considered an underestimator if they provided the correct answer significantly more often when the foil magnitude was larger than the target, relative to when it was smaller than the target. Finally, we asked whether our main pattern of findings differed across these two groups. For 5-year-olds, there was a difference in performance between underestimators and those who did not underestimate. Whereas underestimators showed an effect of task and of magnitude (all *p*s < .0001), those who did not showed neither effect (all *p*s > .1). However, by age 7 there was no such difference: Children showed an effect of task (all *p*s < .0001), of magnitude (all *p*s < .0001), and showed an interaction both if they underestimated ($p < .001$) and if they did not underestimate ($p < .05$). Thus, whereas 7-year-olds clearly did not possess strong AMs for larger number words, the evidence for 5-year-olds' use of AM remains equivocal. To further test the mechanisms guiding children's number word mappings, we next turn to the estimation task.

Estimation and Calibrated Estimation

Analyses

Before conducting analyses, we excluded all responses of 0 and 1 ($N = 13/3{,}768$), as well as all responses more than 10 times larger or smaller than the presented numerosity ($N = 538/3{,}768$). In addition, we removed outliers by excluding all data points more than 3 *SD* from the mean of each participant's estimate of each presented set size ($N = 36/3{,}768$). The frequency of these aberrant responses was comparable in the uncalibrated (14% of data) and calibrated (17% of data) conditions, suggesting that they were more likely the result of creativity than of (a) fatigue (which we would expect to increase in the calibrated estimation task, as it comes after the uncalibrated task) or (b) idiosyncratic responding as a result of our calibration manipulation. Analyses were carried out on the remaining 3,181 responses, except for the ordinality analyses, which included all data (because the direction—but not magnitude—of estimates is all that is measured in the ordinality analyses, outliers do not have disproportionate influence on this measure and therefore need not be removed).

We analyzed two measures of estimation performance. The first was the linear relation between the participant's estimate and the size of the target set. The second was a measure of estimation ordinality. A response was labeled as ordinal if its estimate changed in the correct direction relative to the previous trial. For example, if a larger set was presented on trial n than on trial $n-1$, then the participant's estimate was considered to be ordinal if it was larger on trial n than on trial $n-1$. In this way, we were able to measure children's structural knowledge of the relative ordering of mappings on a trial-to-trial basis, even in cases where estimates were not yet accurate.

Results

First, we conducted a series of analyses to confirm that (a) participants were attending to the task and (b) 7-year-olds outperformed 5-year-olds on the estimation task. Unsurprisingly, a LMM predicting estimates from magnitude, age, and their interaction revealed a significant effect of magnitude ($\beta = 0.27$, $SE = .02$, $p < .001$) and a significant interaction of age and magnitude ($\beta = 0.11$, $SE = .02$, $p < .0001$). Because there was no main effect of age ($\beta = -5.43$, $SE = 4.85$, $p > .05$), this analysis suggests that both 5- and 7-year-olds provided estimates that were linearly related to the presented magnitude, but that their estimates differed for some (but not all) of the magnitudes presented. This is consistent with previous accounts of the development of estimation performance, which have found that 5-year-olds' estimates of relatively large numbers differ from those of older children, and from their own estimates of relatively small numbers (Ebersbach et al., 2008; Siegler & Opfer, 2003).

To test whether children of different ages also differed in their ordinality score (a measure relevant to assessing SM), we constructed a LMM predicting ordinality from age, magnitude, and their interaction. Again, magnitude was a significant predictor of ordinality ($\beta = 0.002$, $SE = .0008$, $p < .01$), suggesting that children of all ages were less likely to provide ordinal estimates for relatively large sets. Age was also a significant predictor of ordinality ($\beta = 0.316$, $SE = .141$, $p < .05$), such that older children were more likely to provide ordinal estimates (79% of trials) than younger children (75% of trials). There was no interaction ($\beta = -0.002$, $SE = .001$, $p > .05$). Because of the substantial effect of age on estimation performance, we analyzed 5-year-olds' data separately from 7-year-olds' data for all subsequent analyses.

In previous studies of adults, changes in estimation behavior caused by misleading feedback have been interpreted to indicate a reliance on SM (Izard & Dehaene, 2008; Shuman, 2007; Sullivan & Barner, 2012). Thus, our first two analyses tested the effects of calibration to determine whether children recruited SM when making estimates. We also tested whether, as in adults, estimates of relatively small numbers were less influenced by misleading feedback than relatively large numbers.

To answer these questions, we constructed an LMM predicting estimates from the presented magnitude, calibration condition (calibrated vs. uncalibrated), and their interaction for each age group separately. Note that these analyses not only test whether feedback influenced estimates for the miscalibrated number (e.g., the largest set estimated) but also for all numbers. For 5-year-olds, we found a significant effect of magnitude ($\beta = 0.238$, $SE = .02$, $p < .0001$), and a significant interaction of magnitude and calibration type ($\beta = 0.069$, $SE = .03$, $p < .01$). For 7-year-olds, we found a similar pattern of results: a significant effect of magnitude ($\beta = 0.352$, $SE = .02$, $p < .0001$) and a marginal interaction of magnitude and calibration type ($\beta = 0.059$, $SE = .03$, $p = .051$). This suggests that 5- and 7-year-olds' estimates were predicted by the target magnitude across all calibration conditions, but that the nature of this relation differed across calibration conditions (see Figure 2). Critically, the interaction between magnitude and calibration type indicates that estimates of some numerical magnitudes were less affected by calibration than others. This is consistent with the view that some numbers are mapped via SM—and thus are subject to calibration—whereas others are mapped via AM and not subject to calibration.

Models—like the ones reported above—that test the effect of calibration across all magnitudes and all calibration conditions are the clearest way to detect overall the effects of SM. This is because these models allow us to measure the effect of calibration at many different magnitudes, and allow us to compare performance across calibration conditions. Still, one might wonder at which point in the count list calibrated estimates differed from uncalibrated estimates. Although this type of analysis does not allow us to make predictions about the direction of calibration (e.g., if children underestimate throughout the uncalibrated task, then even participants in the calibrated to 25 and calibrated to

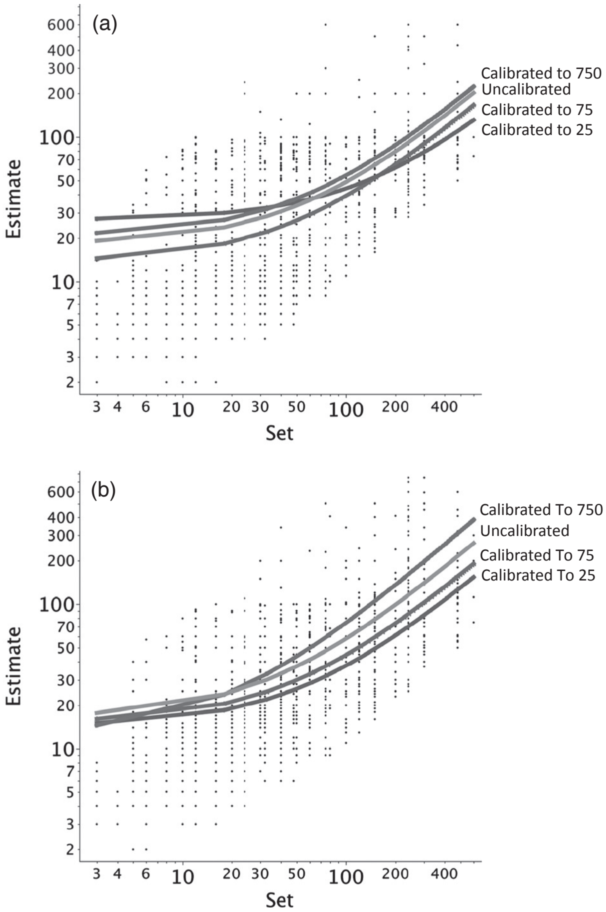

Figure 2. Estimates in log–log space by calibration type for (a) 5-year-olds and (b) 7-year-olds. Data points are individual estimates; lines are model best fits. Calibration condition is indicated next to each best fit line. Note that the degree of separation between calibration conditions increases with magnitude for both 5- and 7-year-olds, even though (because the plots are in log–log space) the separation between conditions appears constant for 5-year-olds.

75 conditions might provide larger estimates when calibrated than uncalibrated), it allows us to further probe the interaction of calibration and magnitude found above. The relatively small amount of data gathered for each set size precludes a detailed analysis of the effect of calibration for each set size, for each calibration type, at each age. However, 5-year-olds showed mean differences (with nonoverlapping standard errors) in estimation performance between the calibrated and uncalibrated conditions for sets starting as small as 12–32, depending on calibration condition, whereas 7-year-olds showed effects for sets as small as 6–10 (see Table 1). Thus, the effect of calibration, while at times small, emerged across the entire number line, and not just for the very largest numbers tested.

Table 1
The Smallest Magnitudes at Which There Was an Effect of Calibration

Effect of calibration	Set	Uncalibrated *M* (*SE*)	Calibrated *M* (*SE*)
5-year-olds			
Calibrated to 25	32	11.5 (1.96)	25.8 (10.3)
Calibrated to 75	12	29.1 (2.7)	13.1 (7.6)
Calibrated to 750	16	15.5 (4.0)	25.4 (5.3)
7-year-olds			
Calibrated to 25	10	6.2 (0.66)	17.6 (4.2)
Calibrated to 75	8	6.5 (0.87)	12.4 (1.6)
Calibrated to 750	6	5.0 (0.73)	13.5 (5.3)

Note. Each calibration condition represents an independent population of participants. Because of this, direct comparisons of uncalibrated estimates across calibration conditions are not meaningful.

To explore the influence of calibration at the subject level, we asked whether each participant's estimates differed by calibration type. Overall, seventeen of the twenty-eight 5-year-olds (60.7%) who contributed full estimation data sets showed either a main effect or interaction involving calibration. We classified each participant as showing an effect of calibration if the model predicting their estimates showed an effect of calibration at $p < .05$, and we classified each participant as showing an interaction of calibration and magnitude of the model showed a significant interaction term at $p < .05$. This means that by chance alone, we would expect 5% of our population to display a significant effect of calibration, and an additional 5% to display an interaction (as our threshold for each analysis was $p < .05$). So, we would expect an effect or interaction of calibration to emerge in our models for 10% (e.g., 5% + 5%) of participants by chance alone. However, many more participants showed an effect of calibration than could be accounted for by chance (binomial $p < .01$; chance = 0.1). Also, nine 5-year-olds showed a significant interaction of magnitude and calibration, suggesting that these participants incorporated misleading feedback into their estimates in the same way that adults do.

For 7-year-olds, 11/23 participants (47.8%) showed an effect or interaction of calibration, also far more than would be expected by chance alone (binomial $p < .01$). Of these, eight showed an interaction, indicating that calibration influenced estimation differently as a function of magnitude. Specifically, these participants were less influenced by misleading feedback for smaller sets, and were more influenced for larger sets. Given the relatively small number of estimates contributed by each participant, such large effects of calibration at both the group and individual levels suggest that participants at all ages relied on SM for most magnitudes tested. More importantly, for the 17 participants who demonstrated a significant interaction of magnitude and calibration, we see evidence that children recruit strong AM for estimates of small numbers, but rely more on SM for estimates of larger numbers.

Next, we asked whether participants' ability to provide ordinal estimates was mediated by calibration condition. If children rely primarily on SM to construct estimates, then there should be no difference in ordinality (despite there being substantial differences in the magnitudes of the actual estimates) across calibration conditions. However, if calibration does not induce a global shift in number word mapping, and instead causes children to deploy idiosyncratic estimation strategies, we might expect ordinality to suffer in the calibrated condition. We found three main results. First, participants in both age groups provided ordinal estimates significantly more often (binomial $p < .0001$) than would be expected by chance in the calibration condition (here, chance was 0.5 because any given estimate is likely to be in the correct direction relative to a previous estimate half of the time; Figure 3). Second, there was no difference in the likelihood of providing an ordinal estimate across calibration conditions (5-year-olds: mean 0.750 ordinal when calibrated, 0.754 uncalibrated, $\chi^2 = 0.08$, $p > .05$; 7-year-olds: mean 0.799 ordinal when calibrated, 0.778 when uncalibrated, $\chi^2 = 1.1$, $p > .05$). Finally, the high rate of ordinality was not restricted only to small sets: By age 5, mean ordinality was significantly above chance for 7/12 comparisons > 50, and was numerically higher than chance for 11/12 comparisons. This suggests that calibration had a global effect on how number words were mapped to the ANS, consistent with the predictions of SM.

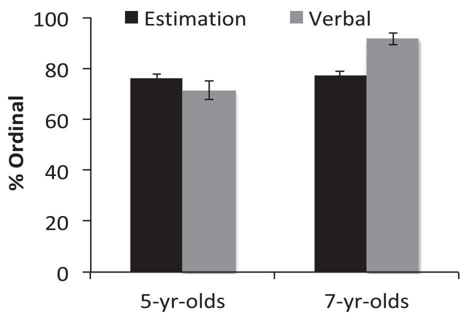

Figure 3. Children's ordinality scores on the estimation task and verbal ordering task.

Counting and Verbal Ordering

Counting

Previous reports have linked counting ability to estimation ability (Davidson et al., 2012; Lipton & Spelke, 2005). However, this relation is complex, and sometimes even poor counters demonstrate strong estimation ability (Barth et al., 2009). Counting performance cannot be used to adjudicate between the use of AM and SM to support estimation, as better counting should be correlated both with more item-based experience and better structural knowledge of number. Still, for relating this work to the existing literature, we tested whether better counters (a) provided more accurate estimates, (b) provided more ordinal estimates, and (c) were more likely to be influenced by calibration. Highest count for 5-year-old averaged 56.9 on the free count test (range = 8–100; mode = 100); for 7-year-olds the average was 80.8 (range = 28–100; mode = 100). For the scaffolded counting assessment, highest count averaged 162 for 5-year-olds (range = 10–1,000; mode = 249) and 643.4 for 7-year-olds (range = 100–1,000; mode = 1,000). The discrepancy in performance between the free count test and the scaffolded counting assessment shows that our participants often possessed knowledge of how relatively large number words are related, even in cases where they were unable to produce those words when reciting the count list. For example, consider the 23 participants who could not count to 100 on the free count test. Of these, 10 counted to 100 or higher on the scaffolded counting assessment, and the mean highest scaffolded count for this subset of participants was 182. Clearly, even children who lack total proficiency with the routine of counting to 100 still know something about very large number words. One conclusion to draw from this finding is that children may be able to produce sensible estimates for magnitudes that are outside of their productive count range, especially if they possess knowledge of a handful of large number words (see Barth et al., 2009, for evidence that this is sometimes the case). Thus, the highest number a child can count to need not be the biggest number they can use to estimate (and in fact, even our scaffolded task may underestimate participants' knowledge).

Next, we asked whether counting performance predicted any of our estimation measures. To do this, we considered counting ability to be a continuous predictor of a variety of estimation outcomes. We analyzed each counting task separately. Neither counting measure predicted the linear slope of participants' uncalibrated estimation performance: free count, $F(1, 45) = .07$, $p = .79$, and scaffolded count, $F(1, 45) = .87$, $p = .37$. Also, better counting ability did not predict higher rates of ordinal responding: free count, $F(1, 45) = .51$, $p = .48$, and scaffolded count, $F(1, 45) = .42$, $p = .52$). Finally, better counting ability did not predict a participant's likelihood that a participant showed a significant effect of calibration: free count, $\chi^2 = 1.90$, $p = .17$, and scaffolded count, $\chi^2 = 2.35$, $p = .13$. Thus, counting was unrelated to our critical measures of estimation performance, unlike in previous studies (Barth et al., 2009; Davidson et al., 2012; Lipton & Spelke, 2005). This is possibly because many of our children were substantially older than those in previous studies (i.e., such correlations may only exist early in the development of estimation abilities), or because children's highest count was not indicative of the actual range of numbers they were familiar with. Future studies should investigate this by testing a wide range of children using a single set of tests.

Verbal Ordering

This task assessed children's knowledge of the relative ordering of number words in the absence of visual cues (e.g., which is more: *twenty* or *forty*?). Overall, participants performed well at this task, and both 5- and 7-year-olds provided correct responses more often than would be expected by chance alone (5-year-old mean = 71.4%; 7-year-old mean = 91.6%; binomial *p*s <.0001; see Figure 3). Children's accuracy improved substantially with age, $F(1, 45) = 22.40$, $p < .0001$, suggesting that knowledge of verbal ordering increases between the ages of 5 and 7 years. These data suggest that children in both age groups possess some understanding of the relative ordering of number words, a skill that is required for the adult-like deployment of SM. However, there was no relation between accuracy on the verbal ordering task and the likelihood that a participant would provide an ordinal response on the estimation task ($\beta = .01$, $SE = .07$, $p > .05$), and whereas verbal ordering performance improved substantially with age, estimation ordinality improved only a small amount (see Figure 3). Although possessing ordinal number word representations is a prerequisite for forming an SM, the presence of relatively good verbal ordering does not ensure that estimates will be ordinal.

Discussion

When children connect language to ANS representations of number, they rely heavily on inferential processes to do so. Based on data from six tasks, we found converging evidence that children, like adults, recruit both AM and SM to construct estimates, suggesting that inferential and associative processes are fundamental to the formation of number word mappings. However, although we found that children use both mechanisms to support estimation, children possessed fewer strong AMs than adults, and appeared to rely heavily on inferences about the structure of the number system. Taken together, our tasks converge to suggest that—from early in development—structural inference is fundamental to guiding connections between number language and number perception. This finding highlights not only the importance of SM in supporting numerical knowledge but also provides a window into understanding the mechanisms that guide children's learning about the relation between language and perception.

According to the AM hypothesis, each number word is mapped onto an ANS representation of numerical quantity on an item-by-item basis, as a result of experience with particular word–magnitude pairings (e.g., Lipton & Spelke, 2005). Converging evidence for children's use of AM comes from two sources. First, on the number matching task, participants tended to be accurate at matching a number word to one of two discriminably different sets when the sets contained a small number of dots: Specifically, both 5- and 7-year-olds showed no differences in performance on the discrimination and number matching tasks for sets containing six or fewer items, but showed large differences in all larger magnitudes. Second, on the calibrated estimation task, many children, like adults (Sullivan & Barner, 2012), were less influenced by misleading feedback for small numbers than for large numbers. Taken together, these findings suggest that our participants possessed strong, statistically reliable, AMs for numbers up to at least six. Of course, it is unlikely that there is anything special about the number six as a cutoff between AM and SM, as the strength of AMs appears to gradually decline as a function of magnitude (and thus the apparent cutoffs are only reflected in significance testing, and not in the pattern of effects themselves). In adult populations, there is large individual variability with respect to where individual subjects exhibit significant differences (Sullivan & Barner, 2012), and this would almost certainly be the case in child populations as well given a paradigm that allowed testing individual differences. Our data also do not differentiate between the possibilities that (a) children possess no AMs above six or (b) children possess AMs that are weaker for larger numbers. However, our data do clearly demonstrate that insofar as AM guides estimation, it plays the largest role in supporting estimates for relatively small numbers, and a much smaller role in supporting estimates for large numbers.

In contrast, the SM hypothesis posits that each number word mapping is constructed in relation to all other mappings, and that the verbal and nonverbal number systems become related to each other on the basis of similarities in their structures. Evidence for SM came from three sources. First, children's accuracy on the number matching task declined as a function of numerical magnitude relative to the discrimination task (resulting in a significant interaction for 7-year-olds), something that would not be predicted if all number words were mapped with equal strength on an item-by-item basis. Second, children's estimates were influenced by misleading feedback about the largest set on the calibrated estimation task, showing that most number word mappings are mutually constraining: Alterations to one mapping influenced many other mappings. This finding—that feedback about the largest set influenced estimates of other sets—has also been demonstrated in adults (Sullivan & Barner, 2012), suggesting that children and adults recruit similar mapping mechanisms. Third, during estimation tasks, children's responses tended to be ordinal (i.e., in the correct direction relative to previous estimates) regardless of estimation accuracy or calibration condition. Our data showed that even the worst estimators—the 5-year-olds—consistently provided ordinal responses more often than would be expected by chance. Taken together, these data suggest that even very young children make inferences based on the structure of the count list when estimating.

At the outset, we noted several ways in which AM and SM might combine to support estimation. One possibility was that children might rely heavily on AM to construct their first number word mappings between language and the ANS (Lipton & Spelke, 2005), and only begin to recruit SM later in development. Second, we noted that, early on, children might first acquire an adult-like set of AMs (e.g., strong up to around 12) before reliably using SM to guide estimation. A final possibility is that children might initially rely heavily on SM, taking advantage of structural knowledge of the count list to form an inferentially derived set of mappings, before acquiring most AMs. Of these three alternatives, our data are most consistent with the third: Both 5- and 7-year-olds were relatively unaffected by calibration for small numbers up to around six, and performed well for these numbers on the number matching task. However, our data are also potentially consistent with a version of the first hypothesis. It is possible that children initially have no AMs between number words and the ANS, but instead that AMs are mediated by other types of numerical representations. This view is possible under the hypothesis, controversial in some quarters, that estimation for the smallest quantities—for example, 1 to 4—is supported by "parallel individuation"—that is, a system for tracking multiple objects as they move through space (see Carey, 2009; Feigenson et al., 2004, for review; for details regarding multiple object tracking in adults, see Pylyshyn & Storm, 1988). In this view, children who estimated using parallel individuation, rather than the ANS, might have made robust estimates for numbers up to six by quickly subitizing a subset of an array, and then accurately extrapolating this estimate to the larger set (e.g., using an additive or multiplicative function; though see Cordes & Brannon, 2009; Negen & Sarnecka, 2010; vanMarle & Wynn, 2009, for evidence that the ANS could be used for small sets). In the context of this study, it is impossible to differentiate this possibility from the idea that estimation for all numbers involves the use of the ANS. More important, if AMs between language and the ANS exist at all, they are very limited in scope, and appear to be strongest for small, frequent, and thus familiar number words (for a discussion of the relative frequency of small vs. large number words, see Dehaene & Mehler, 1992). Thus, to the extent that AMs exist in children, they play a minimal role, whereas even our youngest participants showed evidence of SM in their estimation behavior, supporting the view that inference and analogy are essential learning mechanisms in forming number word mappings.

Importantly, 7-year-olds did not appear to have stronger AMs than 5-year-olds. Thus, children's AMs (however they are represented) did not change noticeably over a period of development in which significant improvements in estimation accuracy occur. This suggests that developmental improvements in estimation (e.g., Siegler & Booth, 2004; Siegler & Opfer, 2003) are not due to changes in AMs. Relatedly, although some have hypothesized that children must have a core set of AMs to support SM (Carey, 2009; Sullivan & Barner, 2012), these data suggest that this core set is very small. Children were able to recruit SM even when they only possessed robust AMs up to about six. This finding strongly suggests that the process of learning to estimate is one that is guided primarily by structural inference.

This finding—that AM plays a limited role in the development of estimation abilities, whereas SM plays a larger role—is relevant to several important questions previously debated in the estimation literature. First, this study is relevant to the observation that estimation ability in school-aged children predicts success in mathematics (e.g., Booth & Siegler, 2008; Siegler & Booth, 2004; Siegler & Ramani, 2009). From our data, it appears unlikely that those who are better at estimating (and thus better at math) have a relatively richer set of AMs. Although educators often focus on providing manipulatives (e.g., toy blocks) to help children visualize the quantities symbolized in math problems (e.g., Burns, 1996; see Uttal, Scudder, & DeLoache, 1997, for another view on the role of manipulatives), even our strongest estimators lacked strong AMs, suggesting that item-specific connections between language and visual representations of magnitudes may not drive early math success.

A different explanation of the relation between estimation and education outcomes is that both draw on children's abilities to recruit SM. Our data showed no relation between counting and estimation ability, and this suggests that merely learning the routine of counting does not ensure adult-like knowledge of the structural relation between numbers within the count list. However, our findings do not rule out the possibility that math skill and estimation ability are related via their shared reliance on SM. Early math education often focuses heavily on teaching the structure of the verbal number system. For example, explicit instruction about the place-value system reinforces structural information about the relations between number words—by understanding place value, children might learn the relation between, say, 30 and 300 (or, conversely, a strong understanding of the relation between 30 and 300 might make it easier for some children to learn concepts like place value). Similarly, basic arithmetic processes involve relating symbolic representations of number to each other. Children who know that 20 + 20 = 40 may be better estimators because both estimation and early arithmetic draw on knowledge of the relation between number words. If this is the case, then focused instruction on the structure of the count list (and not just the routine of counting) may be the best way to improve both math and estimation outcomes. Future research will be required to explicitly test this claim, and to investigate the possible relation between SM and math success.

In this article, we have demonstrated how associative and inferential processes interact to guide children's estimates during development. We have shown that children as young as 5 recruit both AM and SM when estimating, and that they do so in remarkably similar ways to older children and adults. Both 5- and 7-year-olds possess strong AMs for numbers up to about six, and both recruit SM for larger numbers. Future work is required to model the particular ways in which these two learning mechanisms interact throughout development and into adulthood, with the specific goal of understanding the learning mechanisms guiding developmental changes in estimation. Characterizing these changes will help us to understand language and perception interact, and how inference and item-specific experiences combine to form shape conceptual knowledge of number.

References

Atkinson, J., Francis, M., & Campbell, F. (1976). The dependence of the visual luminosity limit on orientation, colour, and grouping in the stimulus. *Perception*, *5*, 335–342. doi:10.1068/p050335

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

Barth, H., & Paladino, A. (2011). The development of numerical estimation in children: Evidence against a representational shift. *Developmental Science*, *14*, 125–135. doi:10.1111/j.1467-7687.2010.00962.x

Barth, H., Starr, A., & Sullivan, J. (2009). Children's mappings of large number words to numerosities. *Cognitive Development*, *24*, 248–264. doi:10.1016/j.cogdev.2009.04.001

Bates, D., & Sarkar, D. (2007). *Lme4: Linear mixed-effects models using S4 classes*. R package version 0.9975-12. Retrieved from http://CRAN.R-project.org/

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*, 545–551. doi:10.1037/a0017887

Booth, J., & Siegler, R. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*, 1016–1031. doi:10.1111/j.1467-8624.2008.01173.x

Brannon, E., & Terrace, H. (2000). Representations of the numerosities 1-9 by Rhesus monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes*, *26*, 31–49. doi:10.1037/0097-7403.26.1.31

Burns, M. (1996). How to make the most of math manipulatives. *Instructor*, *105*, 45–51. Retrieved from http://www.scholastic.com/teachers/instructor

Carey, S. (2009). *The origin of concepts*. New York, NY: Oxford University Press.

Cordes, S., & Brannon, E. (2009). Crossing the divide: Infants discriminate small from large numerosities. *Developmental Psychology*, *45*, 1583–1594. doi:10.1037/a0015666

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. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Attention, Perception & Psychophysics*, *45*, 557–566. doi:10.3758/BF03208063

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.

Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments*. Unpublished manuscript. Retrieved from http://www.unicog.org

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*, 1–29. doi:10.1016/0010-0277(92)90030-L

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number-line and familiarity with numbers in 5- to 9-year old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology*, *99*, 1–17. doi:10.1016/j.jecp.2007.08.006

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

Frank, M., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, *141*, 134–149. doi:10.1037/a0024427

Frege, G. (1953). *The foundations of arithmetic* (J. L. Austin, Trans.). Oxford, UK: Oxford University Press. (Original work published 1884)

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

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170. doi:10.1016/S0364-0213(83)80009-3

Gentner, D., & Namy, L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science*, *15*, 297–301. doi:10.1111/j.1467-8721.2006.00456.x

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

Halberda, J., Mazzocco, M., & Feigenson, L. (2008). Individual differences in nonverbal number acuity predict maths achievement. *Nature*, *455*, 665–668. doi:10.1038/nature07246

Holloway, I., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's math achievement. *Journal of Experimental Child Psychology*, *103*, 17–29. doi:10.1016/j.jecp.2008.04.001

Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': Numerical estimation by 5-7 year olds. *Cognition*, *78*, B27–B40. doi:10.1016/S0010-0277(00)00122-0

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*, 1221–1247. doi:10.1016/j.cognition.2007.06.004

Kant, I. (1929). *Critique of pure reason* (N. K. Smith, Trans.). London, UK: Macmillan. (Original work published 1781)

Le Corre, M. (2013). *Size doesn't count: Evidence that knowledge of the cardinal principle does not grant knowledge of the size of the numbers picked out by number words*. Manuscript submitted for publication.

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*, 395–438. doi:10.1016/j.cognition.2006.10.005

Lipton, J., & Spelke, E. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development*, *76*, 978–988. doi:10.1111/j.1467-8624.2005.00891.x

Moyer, R., & Landauer, T. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519–1520. doi:10.1038/2151519a0

Mundy, E., & Gilmore, C. (2009). Children's mapping between symbolic and nonsymbolic representation of number. *Journal of Experimental Child Psychology*, *103*, 490–502. doi:10.1016/j.jecp.2009.02.003

Negen, J., & Sarnecka, B. (2010). *Analogue magnitudes and knower levels: Re-visiting the variability argument*. Proceedings of the 32nd Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society. Retrieved from http://mindmodeling.org

Pylyshyn, Z., & Storm, R. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 1–19. doi:10.1163/156856888X00122

R Development Core Team (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Shuman, M. (2007). *Computational characterization of numerosity perception and encoding*. (Doctoral dissertation, Harvard University)

Siegler, R., & Booth, J. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428–444. doi:10.1111/j.1467-8624.2004.00684.x

Siegler, R., & Opfer, J. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243. doi:10.1111/1467-9280.02438

Siegler, R., & Ramani, G. (2009). Playing linear number board games—but not circular ones improves lowincome preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*, 545–560. doi:10.1037/a0014239

Slusser, E., Santiago, R., & Barth, H. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology 142*, 193–208. doi:10.1037/a0028560

Sullivan, J., & Barner, D. (2012). How are number words mapped to approximate magnitudes? *Quarterly Journal of Experimental Psychology*, *66*, 389–402. doi:10(1080/17470218), 2012, 715655

Sullivan, J., Juhasz, B., Slattery, T., & Barth, H. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin and Review*, *18*, 557–563. doi:10.3758/s13423-011-0081-1

Thompson, C., & Opfer, J. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development*, *81*, 1768–1786. doi:10.1111/j.1467-8624.2010.01509.x

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. doi:10.1126/science.185.4157.1124

Uttal, D., Scudder, K., & DeLoache, J. (1997). Manipulatives as symbols: A new perspective on the use of concrete objects to teach mathematics. *Journal of Applied Developmental Psychology*, *18*, 37–54. doi:10.1016/S0193-3973(97)90013-7

vanMarle, K., & Wynn, K. (2009). Infants' auditory enumeration: Evidence for analog magnitudes in the small number range. *Cognition*, *111*, 302–316. doi:10.1016/j.cognition.2009.01.011

Vul, E., Barner, D., & Sullivan, J. (2013). *Slow drift of individuals' magnitude-to-number mapping*. Proceedings of the 35th Annual Conference of the Cognitive Science Society. Retrieved from http://mindmodeling.org

Whalen, J., Gallistel, C., & 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*, 155–193. doi:10.1016/0010-0277(90)90003-3

Wynn, K. (1992). Children's acquisition of number words and the counting system. *Cognitive Psychology*, *42*, 220–251. doi:10.1016/0010-0285(92)90008-P

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