Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Sullivan, Jessica
Barner, David

Publication Date
2010

Peer reviewed

Mapping number words to approximate magnitudes: associative learning or structure mapping?

Jessica Sullivan (jsulliva@ucsd.edu)
David Barner (barner@ucsd.edu)
Department of Psychology
University of California, San Diego

Abstract

How do we link number words to the magnitudes they represent? We investigated the roles of associative learning and structure mapping in linking the Approximate Number System to number words. Four tasks demonstrated that individuals have strong associative links between magnitudes and number words for relatively small sets, but have weak associative links for larger sets. These results point to multiple mechanisms for the mapping of number words to magnitudes.

Keywords: Language acquisition, approximate magnitudes, word learning, number words

Introduction

How does language represent human numerical knowledge? Are the referents of numerals determined primarily by inference and logical relations between words? Or do we identify the referents of words like *twelve* and *fifty-seven* by associating them, item-by-item, with sets in the world? Humans can represent the approximate numerical magnitude of a set nonverbally using the Approximate Number System (ANS), and previous research has shown that our system of number language is deeply linked to the ANS: adults recruit the ANS when estimating the cardinality of rapidly presented arrays, and judgments about verbally presented numerals show many of the same biases as nonverbal judgments about quantities (Barth, Kanwisher, & Spelke, 2003; Whalen, Gallistel, & Gelman, 1999, Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; Holloway & Ansari, 2008; Duncan & McFarland, 1980). We know that mature mathematical thinkers can relate their verbal number system to the nonverbal ANS—but we don't know *how* these systems are mapped onto each other.

By some accounts, number words gain their numerical content in part via a mapping to the ANS (Verguts & Fias, 2006, Piazza, Pinel, Le Bihan, & Dehaene, 2007; Mundy & Gilmore, 2009). However, surprisingly little is known about the nature of this mapping, and descriptions of possible mechanisms are rare in the literature. It is therefore not well understood what roles associative and inferential processes play in relating number words to ANS representations. One possibility is that as humans accumulate experience with number words, they form item-specific associations between individual words (e.g., *ten*) and corresponding magnitudes (e.g., approximately 10 objects). This view, which we will call the associative mapping hypothesis (AM), predicts that the strength of mappings should vary according to how frequently words are used to refer to perceptually available quantities. Also, it predicts that mappings should be relatively independent of one another, such that a change in one mapping does not automatically impact another mapping. The AM hypothesis is supported by evidence from children's early mappings of small number words: children learn the associations between their first number words and the magnitudes they refer to one at a time, taking nearly two full years to learn the associations between the number words "one" through "four" and the magnitudes they denote (Wynn, 1992).

A problem with the AM hypothesis is that humans may get only limited experience with the denotations of some number words, and no experience at all with others. It seems implausible, for example, that experience with 1 million things would be required to support estimates for sets of this size. Instead, it appears that inferential abilities are required to give meaning to unfamiliar quantities, perhaps on the basis of more familiar amounts. One possibility, for example, is that associative learning about small quantities (e.g., 1 – 10) supports a structure mapping (SM): a linking of representations in one domain to those in another on the basis of their common structure (Izard & Dehaene, 2008; Thompson & Opfer, in press).

One signature of structure mapping is that when a subject's response for a given quantity is changed via feedback (i.e., calibrated), responses for other quantities should shift accordingly. Evidence for this comes from Izard & Dehaene (2008), who showed that mislabeling a visually presented set (e.g., calling a set of 30 dots "twenty-five") led participants to shift their estimates not only for the calibrated quantity, but for all quantities tested. However, this study provided only small amounts of miscalibration, resulting in tiny differences between conditions. For example, even for the most extremely miscalibrated trials, participants mapped the number word "thirty" onto arrays with an average set size between 31.5 and 33.7 (neither of which is perceptually discriminable from 30 for normal adults). As a result, although it appears that SM plays some role in estimation, it remains unclear how malleable estimation behaviors are, and thus what the relative roles of AM and SM are in the mapping of number words to the ANS.

The present study explored the nature of the relationship between the ANS and the count list, and the relative contributions of associative and structural mapping. We hypothesized that neither mechanism, in isolation, could explain how number words are mapped to the ANS.

Whereas AM is limited by the experiences that individuals have with particular magnitudes, in order for a SM to have content, it must be supported by reliable mappings between small number words and magnitudes. Without associative mappings for at least some number words, we submit, no structure mapping could take place. To our knowledge, no previous study has directly tested the contribution of these two mechanisms, and as a result, little is known about their relative contribution to number word mappings. To test our hypothesis, we conducted an experiment with four within-subjects measures that probed for evidence of associative and structural mappings at different numerosities.

In the Calibrated Estimation task, we measured participants' accuracy at labeling sets after they were provided misleading information about the range of set sizes to be presented. We predicted that quantities that have strong associative mappings should be less susceptible to calibration than those with weaker mappings. As in previous studies, we expected that participants' estimates of magnitudes would be influenced by the feedback they received. However, in the present task, we made two critical methodological changes. First, we provided only verbal calibration. While in past studies participants were shown an array and then told that it contained *x* dots (where *x* was either an accurate or inaccurate number word label), in the present study we did not mislabel arrays. Instead, we simply stated that "the largest set you will see is *x"*. In this way, we ensured that any influence of feedback was not because participants constructed new associative mappings, but was due purely to an inference about the structure mapping relation. A second difference was that we provided much more extreme calibration than in previous studies, in order to test the strength of associative mappings throughout the number line. We reasoned that misleading feedback should not influence estimation performance for any magnitude with a strong AM link to the number system, whereas structurally linked magnitudes should be quite susceptible to calibration.

An assumption in our analysis of the estimation task is that a participant's individual estimates act as inputs to a structure mapping, and that each act of estimation therefore constrains later estimates in an experiment. On this view, we predicted that a very similar task in which participants were asked to match a label to one of two visually presented sets would disrupt the structure mapping process. We reasoned that presenting two sets to participants would cause them to experience uncertainty, and thus prevent them from calibrating their mappings across trials. As such, we predicted that performance on this task should suffer for number words that have weak associative mappings to magnitudes. Where stronger AMs exist (e.g., for smaller numbers), performance should not suffer as much, since by our hypothesis the forced choice task relies on the associative strength between the number word and its corresponding magnitude representation. Although past studies have used a forced choice method to test mappings in young children (Lipton & Spelke, 2005; Gilmore & Mundy, 2009), these studies provided calibration before the study in the form of a familiarization phase. Our study, in contrast, sought to remove all forms of feedback, whether from the experimenter or from trial-to-trial self-calibration, in order to test the strength of associative mappings at different magnitudes.

We conducted two additional tasks as within-subject controls for the Calibrated Estimation and Number Matching tasks. The first was an Uncalibrated Estimation task, which served as a within-subject baseline for the Calibrated Estimation task. The second, a Numerical Discrimination task, used stimuli identical to those in the Number Matching task but asked subjects to judge which of the two sets on each trial was more numerous. This ensured that participants were able to discriminate the quantities used in the Number Matching task, and that any difficulties with the forced choice task were due to their number word mappings and not other stimulus properties.

For both the Number Matching task and the Calibrated Estimation task, we predicted that participants would exhibit strong associative mappings for some number words, resulting in smaller effects of calibration and higher levels of success on the forced choice task. In particular, we predicted that the strength of associative mappings would be strongest for the smallest number words, due to relatively greater experience with these words and their corresponding quantities. Corresponding to this, we also predicted that larger magnitudes would be more susceptible to miscalibration in estimation, and be more difficult to map to number words in the forced choice task.

Materials and Methods

Participants

Thirty adults from the UCSD community participated for course credit. One additional participant was excluded from analyses for failure to complete all tasks.

Procedure

Participants were seated approximately 40 cm from a 27" Mac OSX computer screen and completed 4 computerized tasks. Half of the participants completed the Number Matching task first, and half completed the Discrimination task first. All participants then completed first the Uncalibrated and then the Calibrated Estimation task.

Number Matching: This task evaluated participants' ability to match number words with one of two visually presented sets. Participants heard a number word, saw two arrays of dots flash sequentially on a computer screen, and judged which array best matched the word. Stimuli were sets of red dots on a black screen, and were presented for 400 ms. Trials compared sets that differed in numerical magnitude by either a 1:2 ratio or 3:4 ratio. Sets were matched for density on half of the trials and for total surface area on the other half, and comparisons ranged from small (4 vs. 8) to large (370 vs. 740). For 1/3 of the trials, the smaller of the two sets presented contained fewer than 30 items (Small Number Trials), for 1/3 it contained more than 30 and fewer than 110 (Medium Number Trails), and for the remaining 1/3, it more than 110 items (Large Number Trials).

Numerical Discrimination*:* This task served as a within-subjects control for the Number Matching task to ensure that participants could discriminate the quantities presented. The stimuli and procedure were identical to the Number Matching task, except that participants indicated which set contained more dots (instead of matching a word with a set).

Calibrated Estimation: This task tested the malleability of participants' numerical estimates. Participants saw sets of dots and estimated their numerosities, recording their estimates using the numeric keypad on a computer keyboard. Although the largest set that participants saw was 350 in all conditions, participants were told that the largest set they would see was either 75 (N=10), 375 (N=10), or 750 (N=10). Critically, this misleading feedback could not be used to alter associative mappings since, because unlike in previous calibration studies, the incorrect number word anchor was not paired with an array, and thus participants could not form new associative mappings between magnitudes and number word labels (e.g., Izard & Dehaene, 2008; Shuman, unpublished thesis). Instead, this feedback could only have influenced participants' notions about the structure and range of magnitudes.

Stimuli were sets of red dots on a black screen. Fifteen numerosities ranging from 8-350 were presented 36 times each. Each numerosity was matched for both density (15 trials) and total occupied area (15 trials) with each other numerosity presented, and non-numerical properties of the sets were otherwise varied for the remaining 6 trials. Participants received 270 of the possible 540 trials in the Calibrated condition, and 270 in the Uncalibrated condition. Trials were presented in random order.

Uncalibrated Estimation: This task served as a within-subjects control for the Calibrated Estimation task to provide a baseline of the participant's Uncalibrated estimates. Stimuli and instructions were identical to those in the Estimation task, except participants were given no information about the largest set they would see.

Results

Number Matching and Discrimination

If participants have associative mappings between individual number words and approximate magnitudes, then they should be able to use these mappings to guide the labeling of sets in the Number Matching task. For example, if the number word *twenty* is associatively mapped to a mental representation of 'about 20 things', then participants should never match the word *twenty* to an array that is discriminably different from 'about 20' (e.g., 40). In other words, for all magnitudes that have associative mappings, participants should perform equally well on the Discrimination and Number Matching tasks, because discriminably different magnitudes should be mapped to unique number words.

We first explored whether performance differed on the Number Matching task, as compared to the Discrimination task. Qualitatively, every participant performed worse on the Number Matching task than on the Discrimination task, indicating that matching a number word to the correct array is more difficult than judging which array is more numerous. A paired-samples t-test revealed that this trend reached significance (all $p<.05$) for 21/30 participants (binomial probability $p<.01$). This effect was consistent across the range of comparisons presented: participants were significantly less accurate on the Number Matching task than the Discrimination task for Small, Medium, and Large Number Trials (all $p<.01$).

Table 1: Mean accuracy on the Discrimination and Number Matching tasks by magnitude of smaller set

Set Size	Discrimination	Number Matching
Small (<30)	92%	85%
Medium	85%	63%
Large (>110)	82%	70%

To explore in greater detail whether magnitude influenced accuracy on these two tasks, we compared accuracy on the Number Matching task to accuracy on the Discrimination task for each magnitude presented. Interestingly, participants did *not* perform significantly worse on the Number Matching Task than the Discrimination Task for any of the comparisons containing magnitudes smaller than 15 (Dunnett's mean comparison, all $p>.05$)[1]. This suggests that participants may have associative mappings for relatively small number words. Consistent with this, accuracy on the Number Matching task was not constant for all magnitudes tested: a regression of accuracy onto set magnitude by ratio revealed an effect of ratio ($F(1,1954)=22.4$, $p<.01$) and an effect of set magnitude ($F(1,1954)=70.5$, $p<.01$), but no interaction ($F(1,1954)=1.4$, *ns*). This pattern of performance indicates that participants had greater difficult matching labels to larger sets relative to smaller ones, and is consistent with the hypothesis that small, but not large, magnitudes are associatively linked to number words.

To explore this trend further, we compared accuracy on the Number Matching task for Small, Medium, and Large Number Trials. Accuracy was significantly different as a function of set magnitude ($F(2,987)=54.3$, $p<.01$), and a post-hoc comparison of mean accuracy revealed that

[1] Accuracy also did not differ between the Number Matching and Discrimination tasks for the four largest comparisons presented. This effect appears to be driven by trials where the larger set was also the correct set: participants may have developed a simple response heuristic for these trials like "when I hear an unusually large number word, I select the larger of the two sets". This is unlikely to be evidence of associative mapping for large sets.

participants performed significantly better on Small Number Trials than either Medium or Large Number Trials (both *p*<.01), but that accuracy on the Medium and Large Number Trials did not differ significantly from each other (t=-.57, *ns*). Participants did not struggle to match number words to magnitudes when the words presented were relatively small, yet accuracy declined rapidly as a function of set magnitude, and remained low for the largest sets.

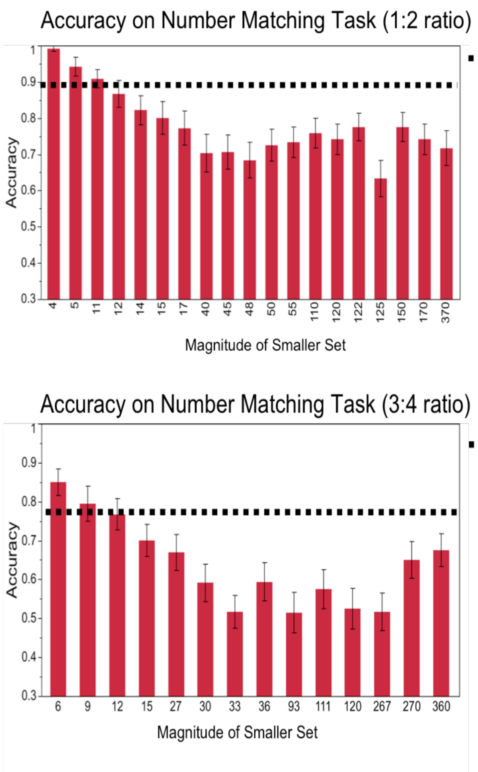

Figure 1: Accuracy on Number Matching task as a function of magnitude of the smaller set being compared

Estimation and Calibrated Estimation

Before completing analyses, we excluded all responses smaller than 2, and all responses more than a factor of 10 larger or smaller than the presented numerosity, as these were likely to be typos. Additionally, we removed outliers on a participant-by-participant basis by excluding all data points more than 3 SD from the mean of each participant's estimate of each presented magnitude (total exclusions: 313/15,450 data points)[2].

As a group, participants provided estimates that were related to the presented magnitude, and that were influenced by misleading feedback (miscalibration): a regression of estimates onto magnitude by calibration type revealed a significant relationship between set magnitude and estimate (*F*(1,15129)=3912.5, *p*<.01), a significant effect of Calibration (*F*(1,15129)=605.3, *p*<.01), and a significant interaction of Calibration and magnitude (*F*(3,15129)=256.3, *p*<.01). Participants were influenced by misleading feedback, and the influence of miscalibration differed as a function of magnitude.

To explore the influence of feedback at an individual level, we performed the identical regression on each individual's data. 24/30 participants showed an effect of calibration: 9/10 participants who were calibrated to 75, 8/10 who were calibrated to 375, and 7/10 who were calibrated to 750 (binomial *p*<.01 for each calibration type). Of the 24 participants who were influenced by calibration, 21 demonstrated a significant interaction of Calibration and magnitude of set (binomial *p*<.01), indicating that calibration influenced estimation patterns differently as a function of magnitude. Specifically, participants were less influenced by misleading feedback for smaller magnitudes, and were more influenced for larger magnitudes.

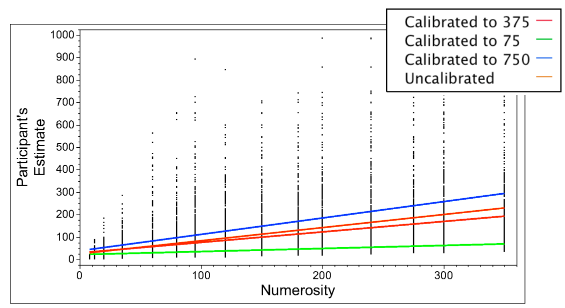

Figure 2: Estimation as a function of calibration

Next, we compared mean estimates for each presented numerosity in the Calibrated vs. Uncalibrated conditions. This isolated the magnitudes for which each participant[3] demonstrated the effects of miscalibration (Dunnett's mean, all *p*<.05). Overall, participants were influenced by calibration for relatively small magnitudes: one quarter of all participants provided different estimates for sets containing 8 items in the Calibrated vs. Uncalibrated conditions, and all but four participants were influenced by misleading feedback for at least one magnitude under 100. Participants were more resilient to misleading feedback for small magnitudes than large magnitudes, yet participants incorporated misleading feedback into the full range of estimates, and did not simply alter estimates for the largest sets presented.

Table 2: Smallest magnitude for which participants were influenced by calibration

Calibration	Mean	Median
75	35	61
375	36	42
750	20	45

[2] We also analyzed our data both excluding and including estimates of "75", "375", and "750" to ensure that any effect of calibration was not due to participants' repetition of the number word they had been miscalibrated to. There was no difference at either the group or individual level of analysis.

[3] Of the 24 who were influenced by calibration

Discussion

This study provides evidence for at least two mechanisms through which number words are mapped to approximate magnitudes. For the smallest numbers we tested, participants were accurate at matching arrays to number words, and they were resilient to misleading feedback. This suggests that, at least for relatively small and familiar magnitudes, adults may form associative mappings between number words and ANS representations. However, for larger magnitudes, adults struggled to correctly match number words with arrays, and were highly influenced by misleading feedback when estimating, making it unlikely that associative mappings play an important role in relating larger number words to magnitudes.

To our knowledge, this is the first study to demonstrate that adults do not directly link each number word in their count list to an ANS representation of approximately that magnitude. While past studies have shown that adults' estimates can be biased by misleading feedback (e.g., Izard & Dehaene, 2008; Shuman, unpublished thesis), in this previous research, the degree of miscalibration was minimal, and the sets that participants labeled as "thirty" before and after miscalibration were not discriminably different from each other using the ANS: this pattern of performance is consistent with either an associative mapping or structure mapping account. However, in the present study, participants failed to correctly match a number word to one of two discriminably different sets, and consistently provided different estimates of a set's magnitude when provided misleading information than when allowed to estimate without constraints. This suggests that adults do not possess associative mappings between large number words and magnitudes.

We posit a structural mapping hypothesis to account for the mappings adults make between large number words and magnitudes. By this hypothesis, adults recruit associatively mapped information about small numbers in order to map larger number words to magnitudes. We know that even large number words bear some relation to ANS representations of magnitudes, because adults' processing of verbal number words exhibit signatures typical of those found in perceptual judgments of numerosity using the ANS (e.g., Duncan & McFarland, 1980). Because of this, we suggest that structural mappings are constructed and supported by knowledge of associatively mapped magnitudes. This process may recruit more domain-general analogical or comparative mechanisms previously linked to the acquisition and extrapolation of spatial, numerical and categorical information during development (e.g., Gentner & Namy, 2006). As a result, structural links between number words and magnitudes may be based on analogy, proportional reasoning, or an understanding of the ordinality of both the verbal and nonverbal number systems. While each of these possible mechanisms for structure mapping is theoretically plausible and consistent with the current data, the present study cannot disambiguate between them. However, future research manipulating the availability and content of 'anchor' sets will allow us to construct a precise model of how small-number information is incorporated into structural mappings, by exploring how manipulating the availability of information about small quantities influences judgments about large quantities.

The present study also raises several developmental questions about the acquisition of number language. While much has been learned in recent years about the procedure of number word learning, little is known of the mechanisms that drive this learning. For example, we know that immediately after learning how counting represents sets, many children fail to map larger number words to larger sets in estimation tasks—however, by age 5, most children successfully provide larger estimates for larger magnitudes (Le Corre & Carey, 2008, Barth, Starr, & Sullivan, 2009). Clearly, 5-year olds have learned something about the count system that the 4-year olds have not—but what? The present study demonstrates that it is unlikely that these older children have improved at estimation solely because they have expanded their system of associative mappings between number words and magnitudes: even adults showed little to no evidence of any direct associative link between words like "one hundred" and sets of 100 things. Instead, children who are successful estimators must have learned something about the structural mapping of the count sequence onto magnitudes. What kind of structural relationships have these children learned?

Additionally, if children's structural mappings are constructed early in life, how much of the count sequence must be associatively mapped in order to support adult-like structural mappings? While the present study suggests that adults may have associative mappings for magnitudes as large as 20, we do not know which of these associative mappings are necessary to support a structure mapping. It may be the case that children need only to associatively map the smallest numbers (e.g., 1-10) in their count system in order to have enough information about the number system to develop a rich structural mapping between number words and magnitudes— by this theory, any additional associative mappings gained en route to adulthood (e.g., mappings between 10-20) are simply the result of additional experience with number words and magnitudes, and are not necessary to support structure mappings. However, it is also possible that children must have an adult-like system of associative mappings in order to support a structure mapping system. By this view, children would construct associative mappings between words and magnitudes for the numbers 1-20 *prior* to creating a structure mapping for larger number words. A continuation of the present line of research with children will help distinguish between these two possibilities, and in doing so, shed light on the nature of the inferences that children make about number words as they construct mappings between number language and magnitudes.

A better understanding of the roles of structural and associative mapping in the development of number knowledge may also help to illuminate other poorly understood developmental phenomena in numerical cognition. For example, one signature of immature estimation ability is a tendency of young children to provide 'logarithmic' looking estimates: overestimating small numbers and underestimating large numbers. This tendency largely disappears in the 0-100 range by age 7, and in the 0-1000 range by age 9 (Booth & Siegler, 2006). What new information about the number system do these older children have? Several researchers have posited that the shift from immature-and-logarithmic to mature-and-linear patterns of estimation stems from a change in these children's underlying numerical representation (Siegler & Opfer, 2003; Booth & Siegler, 2006), or from an increased ability to reason about proportions (Barth & Paladino, 2010). However, the present study leads to an additional (though perhaps not mutually exclusive) hypothesis. Perhaps the shift towards more adult-like estimation can be best explained either by a realignment of structural mappings or to a refinement in the accuracy of the associative mappings that support these structure mappings. These two possible sources of the log-to-linear shift in estimation lead to distinct predictions of how both the younger and older children will reason about small and large numbers.

In conclusion, the present study failed to find evidence of associative links between most number words and magnitudes. Instead, the present study demonstrates that number word meaning is constructed through multiple mechanisms, and not necessarily through associations to real-world exemplars of their referents. By emphasizing the relative roles of associative and structure mappings, we hope to provide a new lens through which to view many of the developmental questions about number language acquisition, and in doing so, to open up new avenues for investigating the kinds of inferences we make about number words.

Acknowledgments

Thanks to Jennifer Audet, Amanda Chamberlain, David Huang, Michael Sim, and Tony Wang for help with data collection.

References

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representation in adults. *Cognition, 86*, 201-221.

Barth, H., Starr, A., & Sullivan, J. (2009). Children's mappings of large number words to numerosities. *Cognitive Development, 24*, 248-264.

Barth, H., & Paladino, A. (in press). The development of numerical estimation in children: evidence against a representational shift. *Developmental Science*.

Booth, J., & Siegler, R. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41*, 189-201.

Duncan, E. & McFarland, C. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: an additive-factors analysis. *Memory and Cognition, 8,* 612-622.

Gentner, D., & Namy, L. (2006). Analogical Processes in Language Learning. *Current Directions in Psychological Science, 15,* 297-301.

Le Corre, M., & Carey, S. (2008). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*, 395-438.

Lipton, J., & Spelke, E. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development, 76*, 978-988.

Holloway, I., & Ansari, D. (2008). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*, 17-29.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*, 1221-1247.

Mundy, E., & Gilmore, C. (2009). Children's mapping between symbolic and nonsymbolic representation of number. *Journal of Experimental Child Psychology, 103*, 490-502.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron, 53*, 293-305.

Shuman, M. (unpublished thesis). Computational characterization of numerosity perception and encoding.

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

Thompson, C., & Opfer, J. (in press). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development*.

Verguts, T., & Fias, W. (2006). Priming reveals differential coding of symbolic and non-symbolic quantities. *Cognition, 105*, 380-394.

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

Wynn, K. (1992). Children's acquisition of number words and the counting system. *Cognitive Psychology, 42*, 220-251.