Psychophysics of Numerical Representation

A Unified Approach to Single- and Multi-Digit Magnitude Estimation

Christopher J. Young and John E. Opfer

The Ohio State University, Columbus, OH, USA

When estimating the value of Arabic numerals like "3," "33," and "333," is each symbol treated holistically and associated with a different mental magnitude along a common scale (Dehaene, 1989), or is there unique scaling for units, decades, hundreds, and so on (McCloskey, 1992)? In this opinion piece, we argue that the evidence is more consistent with the holistic, single-scaling account than a decomposed, multiple-scaling account.

Is There a Break in Children's Early Mental Number Line?

The contrast between the holistic, single-scaling account and the decomposed, multiple-scaling account is usefully illustrated by two alternative models of children's number-line estimation (Moeller, Pixner, Kaufmann, & Nuerk, 2009; Siegler & Opfer, 2003). On a number-line estimation task, participants are shown a blank line flanked by two numbers (e.g., 0 and 1,000) and asked to estimate the position of a third number (e.g., 150). This estimation task is particularly revealing about representations of numerical magnitude because it transparently reflects the ratio characteristics of the number system. Just as 150 is twice as large as 75, the distance of the estimated position of 150 from 0 should be twice as great as the distance of the estimated position of 75 from 0. More generally, estimated magnitude (y) should increase linearly with actual magnitude (x), with a slope of 1.00, as in the equation $y = x$.

Across a number of studies using this number-line estimation task (Booth & Siegler, 2006; Laski & Siegler, 2007; Opfer & Siegler, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003; Thompson & Opfer, 2008), a systematic difference between younger and older children's estimates has been evident: Younger children's estimates of numerical magnitude typically follow Fechner's law ($y = k \times \ln x$) and increase logarithmically with actual value (Figure 1). In contrast, older children's estimates increase linearly with actual value. This developmental sequence is seen at different ages with different numerical scales. It occurs among preschoolers with the 0–10 scale (Opfer,Thompson, & Furlong, 2010), between kindergarten and second grade with the 0–100 scale (Siegler & Booth, 2004), between second and fourth grade with the 0–1,000 scale (Opfer & Siegler, 2007; Siegler & Opfer, 2003), and between third and sixth grade with the 0–100,000 scale (Thompson & Opfer, 2010). Thus, as shown in Figure 1A, on the 0–100 number-line estimation task, the best-fitting logarithmic function fit kindergartners' estimates better than did the best-fitting linear function (R^2 = .75 vs. R^2 = .49). In contrast, on the same task, the best-fitting linear function fit second-graders' estimates better than did the best-fitting logarithmic function (R^2 = .95 vs. R^2 = .88).

Against this logarithmic-to-linear shift model, an alternative view holds that the logarithmic function that fits young children's estimates is merely a good-fitting approximation of two (or more) linear functions. Ebersbach, Luwel, Frick, Onghena, and Verschaffel (2008), for example, suggested that children's estimates are based on their counting range, with differences between smaller and more familiar numbers seeming larger than differences with less familiar numbers. Similarly, Moeller et al. (2009) proposed that the logarithmic estimates of young children actually indicate different representations of units and decades. In this account, children initially estimate differences among single-digit numbers to be greater than differences among two-digit numbers. If so, the appropriate model for early numerical estimation is not a logarithmic function but a single linear function each to single-digit (1–9) and two-digit (10–99) numbers.

In the next two sections, we argue against this segmental linear interpretation on theoretical and empirical grounds. First, we show that both representations of symbolic and nonsymbolic magnitudes are consistent with Fechner's law and have a similar neurocognitive profile. From this perspective, positing a unique representation just for two-digit Arabic numbers is simply unnecessary, and it violates the principle of Occam's razor by multiplying mental processes

Zeitschrift für Psychologie / Journal of Psychology 2011; Vol. 219(1):58–63
DOI: 10.1027/2151-2604/a000047

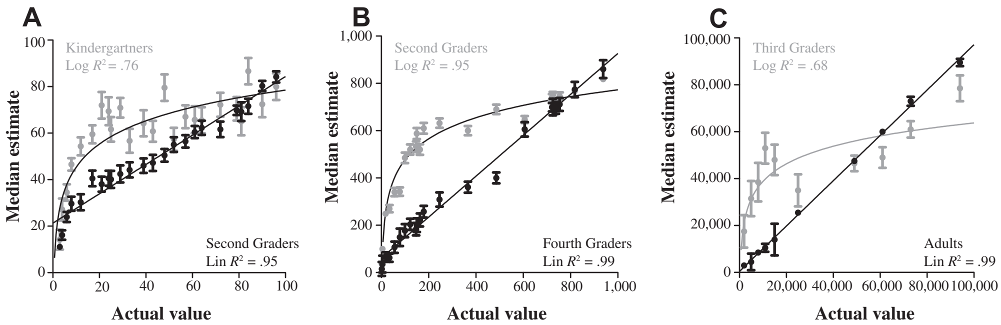

Figure 1. Evidence for a logarithmic-to-linear shift. Data presented in (A) Siegler and Booth (2004), (B) Opfer and Thompson (2008), and (C) Thompson and Opfer (2010).

beyond necessity. Second, we describe appropriate numerical estimation methods and outline a model selection procedure that offers a way to discriminate between the very close predictions of the logarithmic and segmental linear model. Here, we find little support for the idea that a two-linear model is more likely than a logarithmic one.

Representing Numerical Magnitude: The Number Line and Beyond

Representations of numerical magnitude are central to a wide range of activities beyond the number-line estimation task. From understanding the meaning of number symbols (e.g., knowing that "6" or "six" denotes six objects), to comparing the magnitudes of numerals (e.g., knowing that "six" is more than "four"), to estimating quantities (e.g., knowing whether there are closer to 6, 60, or 600 candies in a jar), children must map between alternative quantitative representations, at least one of which is inexact and at least one of which is numerical (Siegler & Booth, 2005). From this perspective, number-line estimation is revealing because it involves one of many possible forms of numerical estimation. Number-line estimation, for example, requires translating a number into a spatial position on a number line or translating a spatial position on a number line into a number. Numerosity estimation also requires translating a nonnumerical quantitative representation (e.g., a picture of candies in a jar) into a number, and computational estimation involves translating from an exact numerical representation (e.g., 75 × 29) to an inexact one (about 2,200). Thus, the mathematical functions translating between external symbolic representations of numbers (e.g., numerals) and internal, analog magnitude representations are highly revealing about how a cognitive system encodes number.

Beyond number-line estimation tasks, many tasks indicate that the function translating objective numeric quantities into a subjective number is logarithmic, thereby following Fechner's law. For example, when selecting the larger of two sets of less than 10 dots, the time required to select the larger set is proportionate to the logarithm of the distance, with the time required to select the larger of 3_5 being less than 5_7 (Buckley & Gilman, 1974). Moreover, the same function describes comparisons of two sets of 10 or more objects (Birnbaum, 1980). Although it is true that comparisons between two sets of 10 or more are always more difficult than comparisons between two sets of < 10, there is clearly no need to postulate a separate high-slope linear representation for sets < 10 and another low-slope linear representation for sets ≧ 10: Fechner's law already generates this prediction – and, unlike a linear scaling of numeric magnitude, it correctly predicts size effects *within* each of the < 10 and ≧ 10 batches.

Just as comparisons of sets of objects are subject to Fechner's law, so too are comparisons of Arabic numerals (Buckley & Gilman, 1974; Dehaene, 1989; Moyer & Landauer, 1967), with the time required to select the larger of two numbers being proportionate to the logarithm of the distance. Thus, the time required to select the larger of 3_5 is less than 5_7 and the time required to select the larger of 30_50 is less than 50_70. Again, although it is true that comparisons of two-digit numbers require more time than comparisons of single-digit numbers, there is no need to postulate separate scaling for single-digit and two-digit numbers: Fechner's law already generates this prediction, and, unlike a linear scaling of numeric magnitude, it correctly predicts size effects within single-digit and two-digit comparisons.

Just as there is no break in the mental number line for comparisons of nonsymbolic and symbolic numbers, there seems to be a very high overlap between the neural regions that encode ones and tens. That is, when comparing either single-digit or two-digit numbers, a common area in the intraparietal sulcus (IPS) moderates the distance effect (Ansari, Dhital, & Siong, 2006; Dehaene, Dupoux, & Mehler, 1990). At present, there appears to be no clear evidence for topographical organization of number-sensitive neurons (Nieder & Merten, 2007), although there is some evidence that the left IPS may be more responsive to large single-digit numbers (5–9) than small single-digit numbers (1–5) (Naccache & Dehaene, 2001; Stanescu-Cosson et al., 2000). There is also a tendency for smaller number tasks to be impaired by more anterior transcranial magnetic stimulation (TMS) and larger number tasks to be impaired by posterior parietal TMS (Göbel, Walsh, & Rushworth, 2001a, 2001b). However, when Göbel, Johansen-Berg, Behrens, and Rushworth (2004) directly tested for differences between single- versus two-digit number comparison, they found no brain areas significantly more active for single-digit number comparison than double-digit number comparison, and the opposite contrast only yielded activation within areas associated with general visual processing.

Evaluating the Logarithmic and Decade-Break Models

In addition to good theoretical reasons to expect one scale for single- and multi-digit numbers, we believe that the observed results from the number-line estimation task also favor one scaling solution to two or more. In this section, we present a new analysis of previously collected data (Laski & Siegler, 2007; Opfer & Siegler, 2007; Opfer & Thompson, 2008; Siegler & Booth, 2004; Thompson & Opfer, 2008) in an effort to determine (1) whether the previously observed superiority of the logarithmic function on 0–100 and 0–1,000 tasks required fitting a function over numbers that varied in having a single, two, or three digits and (2) whether the previously observed superiority of the linear function for data within an order of magnitude (e.g., 1–9 and 10–99) implies separate representations for single-digit, two-digit, and three-digit numbers.

Our choice of studies and subjects for reanalysis was necessarily selective, and we were guided by two principles that may have affected results reported by Moeller et al. (2009). The first principle was to exclude data from subjects who had had feedback on the task (e.g., Experiment 1 of Booth & Siegler, 2006, or Experiment 2 of Opfer & Siegler, 2007) or given the opportunity to compare estimates over different ranges (e.g., Siegler & Opfer, 2003; Thompson & Opfer, 2010). Given only a single trial of feedback, subjects typically recalibrate their estimates quite broadly compared to subjects given no feedback (Opfer & Siegler, 2007; Opfer & Thompson, 2008; Thompson & Opfer, 2008). Also, given the opportunity to compare one number-line task (e.g., 0–100) to a second task (e.g., 0–1,000), children often perform better on the second task than when given no such opportunity (Thompson & Opfer, 2010). One concern we have is that a similar type of effect might have influenced the responses of children in Moeller et al.'s experiment, where children were presented, invariably, first with a 0–10 version of the number-line task followed by a 0–100 version.

A second principle that we followed was to aim for population homogeneity. In any developmental study, one often finds a mixture of immature and mature response patterns within a given age group, even when the developmental sequence is quite stable. For example, in Siegler and Opfer's (2003) study of number-line estimation, the logarithmic model was the best fitting one for 91% of second graders, 44% of fourth graders, 28% of sixth graders, and 3% of adults. As illustrated in Figure 2, the particular mixture of logarithmic and linear response patterns can have a large effect on the fit of the logarithmic function to mean responses. Thus, in a hypothetical mixed population made entirely out of linear and logarithmic subjects, averaged together, the logarithmic model does not provide a better account until the population is over 70% logarithmic, whereas the segmental linear model retains a strong fit across all mixtures *even though no responses were generated by a segmental linear function*. For this reason, we analyzed the results of children best fit by the logarithmic and linear functions separately.

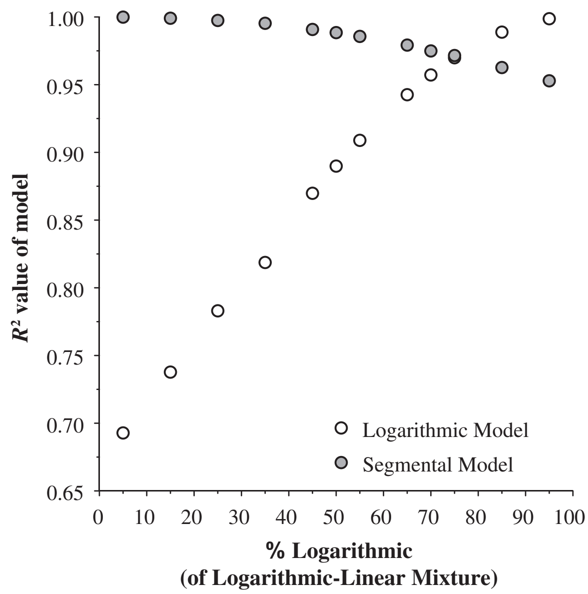

Figure 2. Segmental linear overfitting of simulated mixed population. Goodness-of-fit measures for the segmental and logarithmic functions to averaged data that is generated by a mixture of logarithmic and linear functions.

0–100 Analysis

To examine the fit of the logarithmic and linear functions, 88 children (age range = 5.8–8.2 years) were selected from Experiment 1 of Laski and Siegler (2007) and Experiment 1 of Siegler and Booth (2004); these children had previously been categorized as generating logarithmic estimates, and thus provided the best test of the decade-break alternative.

We first examined the fit of the logarithmic and linear functions to the average estimate given for all the numbers (2, 3, 4, 5, 6, 8, 12, 17, 21, 24, 25, 26, 29, 33, 34, 39, 42, 43, 46, 48, 52, 54, 57, 58, 61, 64, 67, 72, 73, 78, 79, 81, 82, 84, 89, 90, 92, 96). The best-fitting function for all the estimates was logarithmic (log R^2 = .93, lin R^2 = .71) and is shown in gray in Figure 3. We next examined the fit of the logarithmic and linear functions for selected ranges of data (Figure 3: left panel, 1–25, 1–50, 1–75, 1–99; right panel, 1–9, 10–25, 10–50, 10–75, 10–99). From the perspective of the logarithmic-to-linear shift hypotheses, all of these selected ranges are essentially arbitrary subdivisions over one function, but from the perspective of the segmental linear model, linear functions are expected to have poorer fits for ranges that cross over the single-/two-digit divide

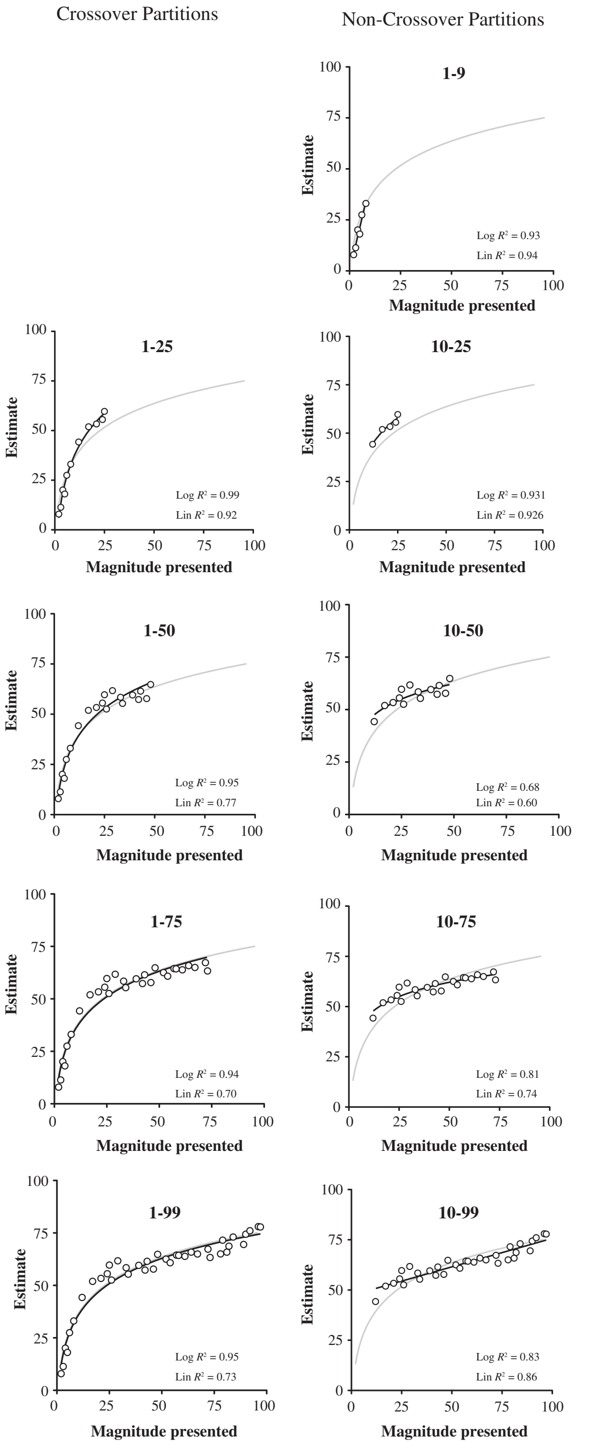

Figure 3. Reanalysis of data from Laski and Siegler (2007) and Siegler and Booth (2004). The best-fitting function for all the estimates was logarithmic (log R^2 = .93, lin R^2 = .71) and is shown in gray. The left panel (crossover partitions) depicts the fit of the linear and logarithmic functions for numeric ranges that crossed over the single-/two-digit divide. The right panel (non-crossover partitions) depicts the fit of the linear and logarithmic functions for numeric ranges that did not cross over the single-/two-digit divide.

and to have better fits for ranges that do not cross over the divide. As illustrated in Figure 3, the logarithmic function provided the better fit for 7/9 partitions (4/4 crossover, 3/5 noncrossover). Indeed, among these multiple comparisons, the *only* partitions where the linear function provided the best fit were for 1–9 (log R^2 = .93, lin R^2 = .94) and 10–99 (log R^2 = .83, lin R^2 = .86), where Moeller et al. also found a better fit of the linear to logarithmic function. Does this suggest that we have separate scales for representing single-digit and two-digit numbers? Clearly not: if two-digit numbers were represented on the same linear scale, then the linear function would have also provided the best fit for 10–25, 10–50, and 10–75, as well as 10–99. Instead, we see that the superior fit of the logarithmic function holds for most two-digit ranges as well as most ranges that bridge the supposed single- to two-digit divide.

0–1,000 Analysis

The method for the 0–1,000 analysis was the same as that of the 0–100 analysis. Here, 138 subjects (age range = 6.4–11.2 years) were drawn from the pretest phases of Opfer and Siegler (2007), Opfer and Thompson (2008), and Thompson and Opfer (2008).

We first examined the fit of the logarithmic and linear functions to all the numbers tested (18, 34, 56, 78, 100, 122, 147, 150, 163, 179, 246, 366, 486, 606, 722, 725, 738, 754, 818, 938). The best-fitting function for all the estimates was logarithmic (log R^2 = .97, lin R^2 = .65) and is shown in gray in Figure 4. We next examined the fit of the logarithmic and linear functions for selected ranges of data (Figure 4: left panel, 0–99, 10–250, 10–500, 10–750, 10–999; right panel, 10–99, 100–250, 100–500, 100–750, 100–999). Again, from the perspective of the logarithmic-to-linear shift hypotheses, all of these selected ranges are essentially arbitrary subdivisions over one function, but from the perspective of the segmental linear model, linear functions are expected to have poorer fits for ranges that cross over the two-/three-digit divide and to have better fits for ranges that do not cross over the divide. As illustrated in Figure 4, the logarithmic function provided the better fit for 9/10 partitions (5/5 crossover, 4/5 noncrossover). Among these multiple comparisons, the only partition where the linear function provided the best fit was for 100–999 (log R^2 = .91, lin R^2 = .94). Thus, on the 0–1,000 task, it seems much more clear that two- and three-digit numbers were not represented on two different linear scales. Instead, we see that the superior fit of the logarithmic function holds for the two-digit range, most three-digit ranges, and all ranges that bridged the supposed two- to three-digit divide.

In our analysis of the 0–100 and 0–1,000 tasks, we followed Moeller et al.'s (2009) lead in closely examining the fit of the logarithmic and linear functions for selected ranges of numbers. Given no prior hypotheses regarding the best range to select for either the logarithmic or linear function, we did not find any particular superiority of the linear function for single- versus two-digit ranges nor for two- versus three-digit ranges. Indeed, in the few cases where the linear

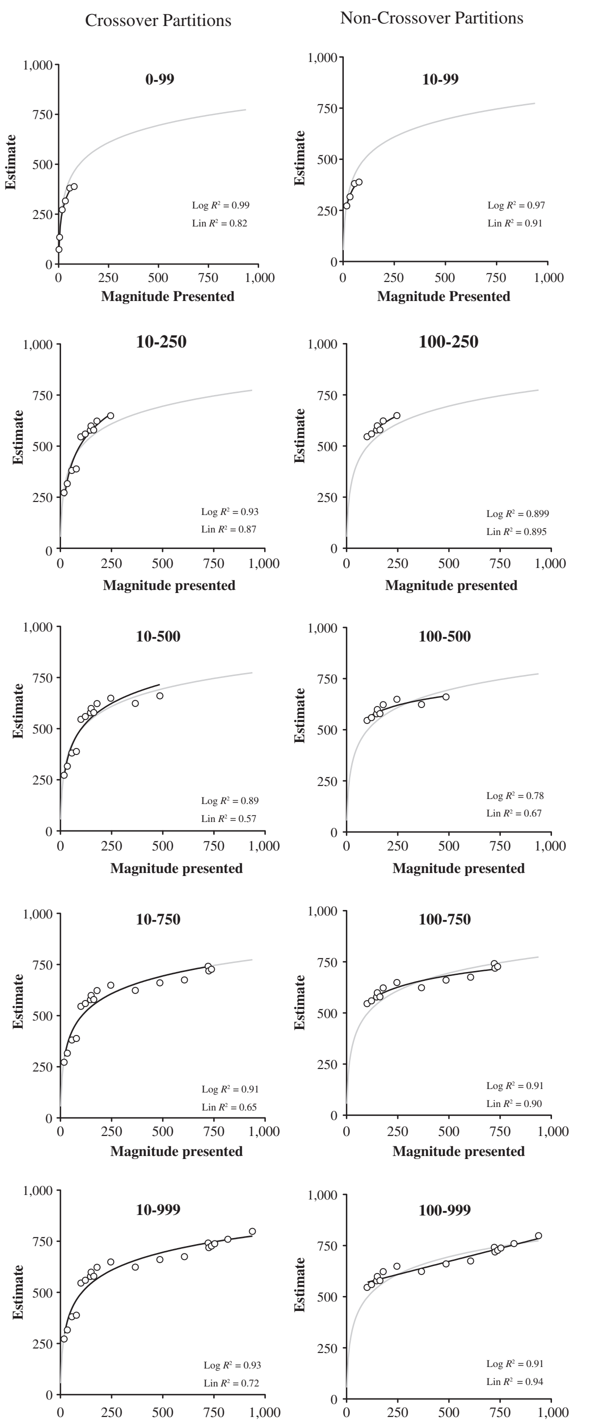

Figure 4. Reanalysis of data from Opfer and Siegler (2007), Opfer and Thompson (2008), and Thompson and Opfer (2008). The best-fitting function for all the estimates was logarithmic (log R^2 = .97, lin R^2 = .65) and is shown in gray. The left panel (crossover partitions) depicts the fit of the linear and logarithmic functions for numeric ranges that crossed over the two-/three-digit divide. The right panel (non-crossover partitions) depicts the fit of the linear and logarithmic functions for numeric ranges that did not cross over the two-/three-digit divide.

function provided the best fit (e.g., 10–99), it only did so on one task (0–100) but not the other (0–1,000). In our view, contradictory findings such as this point to a larger error in attempting to subdivide one continuous function into component discrete functions. In other domains (such as computer science), discretization of a continuous function leads to unreliable errors despite its efficacy in a specific situation. In our view, the unreliability of the segmental linear model provides another example of the dangers of discretization.

Acknowledgments

The authors would like to thank Julie Booth, Elida Laski, Robert Siegler, and Clarissa Thompson for kindly sharing their data with us.

References

Ansari, D., Dhital, B., & Siong, S. C. (2006). Parametric effects of numerical distance on the intraparietal sulcus during passive viewing of rapid numerosity changes. *Brain Research, 1067*, 181–188.

Birnbaum, M. H. (1980). Comparison of two theories of "ratio" and "difference" judgments. *Journal of Experimental Psychology: General, 109*, 304–319.

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

Buckley, P. B., & Gilman, C. B. (1974). Comparison of digits and dot patterns. *Journal of Experimental Psychology, 103*, 1131–1136.

Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception & Psychophysics, 45*, 557–566.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 626–641.

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.

Göbel, S., Johansen-Berg, H., Behrens, T., & Rushworth, M. F. S. (2004). Response-selection-related parietal activation during number comparison. *Journal of Cognitive Neuroscience, 16*, 1536–1551.

Göbel, S., Walsh, V., & Rushworth, M. F. S. (2001a). rTMS disrupts the representation of small numbers in supramarginal gyrus. *Neuroimage, 13*, 409.

Göbel, S., Walsh, V., & Rushworth, M. F. S. (2001b). The mental number line and the human angular gyrus. *Neuroimage, 14*, 1278–1289.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 76*, 1723–1743.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dycalculia. *Cognition, 44*, 107–157.

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology, 103*, 503–515.
Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215*, 1519–1520.
Naccache, L., & Dehaene, S. (2001). The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex, 11*, 966–974.
Nieder, A., & Merten, K. (2007). A labeled-line code for small and large numerosities in the monkey pre-frontal cortex. *The Journal of Neuroscience, 27*, 5986–5993.
Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55*, 169–195.
Opfer, J. E., & Thompson, C. A. (2008). The trouble with transfer: Insights from microgenetic changes in the representation of numerical magnitude. *Child Development, 79*, 790–806.
Opfer, J. E., Thompson, C. A., & Furlong, E. E. (2010). Early development of spatial-numeric associations: Evidence from spatial and quantitative performance of preschoolers. *Developmental Science, 13*, 761–771.
Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428–444.
Siegler, R. S., & Booth, J. L. (2005). Development of numerical estimation: A review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 197–212). Boca Raton, FL: CRC Press.
Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237–243.
Stanescu-Cosson, R., Pinel, P., van De Moortele, P, van De Moortele, P. F., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain, 123*, 2240–2255.
Thompson, C. A., & Opfer, J. E. (2008). Costs and benefits of representational change: Effect of context on age and sex differences in magnitude estimation. *Journal of Experimental Child Psychology, 101*, 20–51.
Thompson, C. A., & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development, 81*, 1768–1786.

Christopher J. Young

Department of Psychology
Ohio State University
1835 Neil Ave.
Columbus
OH 43210
USA
E-mail young.1202@osu.edu