January 2013

Developmental change in numerical estimation

Emily Slusser
San Jose State University, emily.slusser@sjsu.edu

R Santiago
Wesleyan University

H Barth
Wesleyan University

Developmental Change in Numerical Estimation

Emily Slusser, Rachel Santiago, and Hilary Barth

Wesleyan University

Keywords: cognitive development; number; numerical cognition; proportion; estimation

Word Count: 9848

Abstract

Mental representations of numerical magnitude are commonly thought to undergo discontinuous change over development in the form of a "representational shift". This idea stems from an apparent categorical shift from logarithmic to linear patterns of numerical estimation on tasks that involve translating between numerical magnitudes and spatial positions (such as number-line estimation). However, the observed patterns of performance are broadly consistent with a fundamentally different view, based on psychophysical modeling of proportion estimation, that explains the data without appealing to discontinuous change in mental representations of numerical magnitude. The present study assessed these two theories' abilities to account for the development of numerical estimation in 5- through 10-year-olds. The proportional account explained estimation patterns better than the logarithmic-to-linear-shift account for all age groups, at both group and individual levels. These findings contribute to our understanding of the nature and development of the mental representation of number and have more general implications for theories of cognitive developmental change.

Developmental Change in Numerical Estimation

Contemporary views of cognitive development and learning have been heavily influenced by a large body of work aimed at assessing the development of number representation. This work reveals a developmental sequence that occurs in a similar fashion across multiple age groups, tasks, and timescales (for a review, see Siegler, Thompson, & Opfer, 2009). The developmental sequence comprises an apparent discontinuous change in mental representations of numerical magnitude, often described as a representational shift.

This view of the nature and development of numerical representations has shaped theoretical approaches to cognition, learning, and development (e.g. Opfer & Siegler, 2007; Siegler, Opfer, & Thompson, 2009). For example, some have suggested that these repeating patterns of change across multiple timescales are consistent with views of cognitive development held by contemporary information-processing and dynamic systems theorists (Siegler, Thompson, & Opfer, 2009). Microgenetic studies are thought to demonstrate the rapid replacement of one representation with another (Opfer & Siegler, 2007), thereby providing direct support for an overlapping-waves theory of cognitive change (Siegler, 1996). Furthermore, recent studies probing the origins of mathematics and the nature of cultural influences on mathematical cognition have also drawn on the concept of a representational shift (Dehaene, Izard, Spelke, & Pica, 2008; but see Cantlon, Cordes, Libertus, & Brannon, 2009).

This body of work also has clear links to formal education and mathematics learning. (e.g. Krasa & Shunkwiler, 2009). Performance patterns that are thought to indicate the successful achievement of representational change are correlated with performance on standardized math tests and other measures of mathematical ability (Booth & Siegler, 2008; Laski & Siegler, 2007; Siegler & Booth, 2004; Siegler & Ramani, 2008, 2009; Siegler et al., 2009), and children with mathematical learning disability show a delay in exhibiting these patterns (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Geary, Hoard, Nugent, & Byrd-Craven, 2008). The idea of a representational shift has also led to the development of brief, low-cost interventions that can improve math performance in lower-income children, purportedly by supporting changes in their number representations (Siegler & Ramani, 2008; 2009).

Evidence for the representational shift view comes from a family of tasks that share a common structure: they are typically variations on number-line estimation, involving translations between numerical values and spatial positions on a line. To evaluate children's performance, researchers examine the relationship between the magnitudes represented by given symbols (most commonly Arabic numerals, but sometimes visually presented sets of dots or other representations of magnitude) and participants' estimates of those magnitudes (most commonly marked spatial positions on the number line that researchers convert into corresponding numerical values). When estimates are plotted as y-values against given magnitudes on the x-axis, perfectly accurate performance falls on the line $y = x$. Deviation from this line provides a measure of estimation error. Individuals producing high rates of estimation error tend to overestimate smaller values on the number line, such that their estimates appear better characterized as a logarithmic curve than as a straight line (e.g. Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Booth & Siegler, 2006, 2008; Laski & Siegler, 2007; Opfer & Siegler, 2007; Siegler & Opfer, 2003).

Possibly the first study to note this pattern examined 2nd, 4th, and 6th graders' performance with 0-100 and 0-1000 number lines (Siegler & Opfer, 2003). When asked to estimate the positions of numbers on 0-1000 number lines, 6th graders produced estimates that were linearly related to the presented values, while 2nd graders' estimates were better fit with a logarithmic curve. Fourth graders' estimates, on the other hand, showed both patterns: roughly half were classified as logarithmic and half as linear. Moreover, individual 2nd graders whose estimates were categorized as logarithmic on the 0-1000 number line produced more linear estimates on a smaller, more familiar number range (0-100), showing that the same child could produce both estimation patterns depending on the range of numbers presented. A similar developmental pattern arose for kindergarteners and 1st and 2nd graders tested on a 0-100 number line (Siegler & Booth, 2004). Kindergarteners' estimates were better fit by a logarithmic function, 2nd graders' estimates were better fit by a linear function, and 1st graders were split between the two (for similar findings and interpretation see also Booth & Siegler, 2006, 2008; Laski & Siegler, 2007; Opfer & Siegler, 2007; Opfer & Thompson, 2008). These logarithmic and linear estimation patterns have also been noted in preschool-aged children (Berteletti et al., 2010; Siegler & Ramani, 2008) as well as across language and culture (see Dehaene et al., 2008; Siegler & Mu, 2008).

The representational-shift explanation accounts for these data by supposing that logarithmic or linear estimation patterns provide direct readouts of the participants' mental representations of number. Thus, children have access to multiple types of coexisting mental number representations, drawing upon linear representations for tasks spanning more familiar numerical ranges and logarithmic representations for unfamiliar ranges. Over the course of development and experience, children learn to rely primarily on the mature, linear representation of number that supports accurate, linear estimation.

Proponents of this view have conducted several studies designed to promote the representational shift by encouraging children's use of linear, rather than logarithmic, representations. For example, Siegler and Booth (2004) asked children to estimate the positions of multiple numbers on a single number line. This led to a marked improvement on subsequent number-line tasks, with improvement defined as the degree to which children produced linear estimation patterns. Opfer and Siegler (2007) found that feedback for numbers that should correspond to the greatest discrepancy between logarithmic and linear representations (e.g., 150 on a 0-1000 number line) is particularly effective in improving performance. Further studies have shown similar effects (characterized as a shift from logarithmic to linear estimation performance) using progressive alignment cues (Thompson & Opfer, 2010) and experience with linear numerical board games (Siegler & Ramani, 2008, 2009). Improvement can be dramatic and sudden, appearing even after only a single feedback trial and occurring broadly across the entire range of numbers tested. This is thought to provide particularly strong evidence for the idea of a representational shift, as children shift from profoundly biased, more logarithmic estimation patterns to much more linear patterns within a single testing session (Opfer & Siegler, 2007; but see Barth, Slusser, Cohen, & Paladino, 2011).

Some critics of the representational-shift hypothesis have argued that the choice of a logarithmic model to explain children's estimation performance is problematic, and that relatively younger children's estimates may be better explained by a two-segment linear model, with each segment having a different slope. Rationale for one version of this model stems from the notion that estimation errors may differ when the given numbers fall within vs. outside a child's count range (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008). In support of their hypothesis, these researchers argued that kindergarteners' and 1st graders' number-line estimates were better characterized by a segmented linear model than by a logarithmic one, and that the change point for a particular child (the point at which the two linear models were segmented) was significantly correlated with his or her count range. Other researchers also report evidence for a segmented linear model based on 1st graders' estimates with 0-100 number lines (Moeller, Pixner, Kaufmann, & Nuerk, 2009). These authors postulated that the change point in the segmented linear model represents a change from processing single-digit to two-digit numbers, rather than a move from a familiar to an unfamiliar number range. Both versions of the segmented linear hypothesis describe development in terms of the eventual integration of the two-part representation into one holistic linear representation (see also Moeller & Nuerk, 2011; but see Dehaene et al., 2008; Young & Opfer, 2011).

The use of logarithmic and linear models is nevertheless clearly theoretically motivated in numerical estimation research. In fact, there is a long-standing debate as to whether internal representations of numerical magnitude are organized in a linear or a logarithmic fashion. Some maintain that underlying representations of number are logarithmically organized (an idea rooted in the Weber-Fechner Law stating that the magnitude of sensation is logarithmically related to objective stimulus intensity; see for example Dehaene, 1997). Others posit that underlying representations of number are linearly spaced and that logarithmic patterns may emerge as a result of scalar variability (variability that increases in proportion to numerical magnitude; see for example Brannon, Wusthoff, Gallistel, & Gibbon, 2001; Gallistel & Gelman, 1992; cf. Gallistel, 2011).

Number-line estimation has been identified as a possible means to reconcile this debate. For example, Siegler and Opfer (2003) argued against the idea of a linear representation with scalar variability after failing to observe scalar variability in number-line estimation tasks. However, typical bounded number-line tasks elicit estimates relative to marked endpoints, prompting participants to make judgments about relative numerical magnitude within a restricted range. Because the task puts an upper bound on participants' responses, thereby affecting the variability of their estimates, the absence of scalar variability in number-line estimates does not imply a lack of scalar variability in mental representations of numerical magnitude (see also Cohen & Blanc-Goldhammer, 2011). This suggests that typical number-line estimation tasks are poorly suited to resolving these longstanding debates over logarithmic vs. linear mental representations of number.

The question of how (and whether) to draw conclusions about internal scales of magnitude from various types of estimation tasks is complex (e.g. Cantlon et al., 2009; Dehaene et al., 2009; Izard & Dehaene, 2008; Krueger, 1989; Laming, 1997; Poulton, 1989; Teghtsoonian, 1973; see also Vlaev, Chater, Stewart, & Brown, 2011). Numerical estimation is no exception, and of course we do not solve this problem here. However, there are clear difficulties with the conclusions about mental representations that are commonly drawn from performance on number-line estimation tasks. This paper will focus on trying to resolve some of these. One difficulty is that the categorization of estimation data as either logarithmic or linear is questionable, even though most studies only consider these two possibilities (cf. Ebersbach et al., 2008; Moeller et al., 2009; Siegler & Opfer, 2003). A second concern is that by applying logarithmic and linear models to number-line estimation data, researchers effectively treat them as simple tasks that require the estimation of single numerical magnitudes in isolation, failing to acknowledge task constraints with important implications for interpretation. A third potential problem is that typical analyses of these tasks attribute variations in number-line estimation solely to *numerical* processing and *numerical* representations, assuming that the *spatial* components of the task do not introduce their own variations. This assumption is deeply problematic given a wealth of research on the estimation of spatial position in children and adults (e.g. Huttenlocher, Hedges, & Vevea, 2000; Huttenlocher, Newcombe, & Sandberg, 1994). We return to the first two points in more detail below, and to the third in the General Discussion.

Theoretical Models Based on Proportional Reasoning

The bounded nature of the typical number-line estimation task strongly suggests a need to look beyond the models of unbounded numerical magnitude most often applied to these data. Of course, related ideas about the importance of endpoints and other reference points in this task have been noted before. Ebersbach and colleagues (2008, p. 13) recognized the problem of applying inappropriate models to explicitly bounded tasks, suggesting that the assumptions of models tested in many previous studies "might need to be adapted to magnitude estimation tasks that involve anchored response scales." These authors also noted that anchors could be provided explicitly (like the marked endpoints) or even generated by the participants themselves. Siegler and Opfer (2003) also considered an informal "landmark-based proportionality model," allowing for the possibility that participants mentally divide the number line in half or in quarters, creating reference points to guide their estimates. Thus, while the existence of reference points in the number-line task has been widely recognized, far less attention has been paid to the important implications of the task's bounded nature for the interpretation of estimation data (but see Cantlon et al., 2009).

Recently, several research groups have approached this problem by applying a psychophysical model of proportion judgment, appropriate for bounded estimation tasks, to number-line estimation data. This model was originally developed for tasks involving judgments of perceptual magnitude (see Hollands & Dyre, 2000; Hollands, Tanaka, & Dyre, 2002; Spence, 1990). It does an excellent job of explaining estimation bias in number-line tasks as well, both for 7-year-olds (Barth & Paladino, 2011) and for adults (Cohen & Blanc-Goldhammer, 2011;

Sullivan, Juhasz, Slattery, & Barth, 2011). Logically, the justification for the use of this model stems from the fact that number-line tasks require the estimation of a smaller magnitude (the value presented) relative to a larger one (the value given at the upper endpoint of the line), thereby eliciting a judgment of a numerical proportion rather than an unbounded numerical magnitude. Thus a bounded number-line task with Arabic numerals requires participants to retrieve the magnitudes associated with the given values from memory (for example, by retrieving the magnitudes associated with "43" and "100" on a 0-100 number line), to estimate the proportion of the two magnitudes, and to produce a corresponding spatial proportion by marking the number line in the appropriate position.

The psychophysical model of proportion judgment discussed here is derived from Stevens' Power Law, which expresses the relationship between the estimated magnitude of a stimulus and its actual magnitude as $y = \alpha x^{\beta}$. The exponent β is a free parameter that may be thought of as a quantification of bias associated with estimating a particular type of stimulus magnitude (such as brightness, area, or length), and α is a scaling parameter. Here, "estimated magnitude" would be the participant's estimated position on a number line (which the researcher typically converts to a corresponding numerical value) and "actual magnitude" would be the value of the presented numeral (see Figure 1A). Importantly, like the logarithmic and linear functions used to model estimates in many number-line studies, Stevens' Power Law has been considered as a model of numerical estimation in various tasks (e.g. Cordes, Gelman, Gallistel, & Whalen, 2001; Izard & Dehaene, 2008; Shepard, Kilpatric, & Cunningham, 1975; Siegler & Opfer, 2003; Stevens & Galanter, 1957; see also Krueger, 1989, for a review).

Spence (1990) suggested that bias in proportion judgments arises from the biases associated with estimating the component part and whole magnitudes. He adapted the power law for proportional magnitude judgments, modifying the basic power function to define perceived magnitude (y) in terms of the presented range (such that the scaling parameter included in most formulations of Stevens' Law cancels out, leaving the exponent β as the single free parameter). When this model is applied to a 0-100 number line task, for example, estimates are predicted by the function $y = x^{\beta} / (x^{\beta} + (100 - x)^{\beta})$. Spence's model predicts that estimates of proportions will take the form of S-shaped or reverse S-shaped curves, depending on the particular value of β in question (see Figure 1B). Spence's original formulation cannot account for the patterns of performance that would arise from an observer using reference points in addition to the two endpoints. However, a generalized form (the Cyclical Power Model, or CPM; Hollands & Dyre, 2000) predicts a pattern of over- and underestimation, akin to that predicted by Spence's model, which repeats between every pair of reference points used. Thus the generalized model is equivalent to the basic Spence model when the number of reference points is fixed at two (the endpoints; see Figure 1B, the "one-cycle" version of the model). But if three reference points are used, as when estimates are made relative to the two endpoints plus a given or inferred midpoint, a "two-cycle" version of the model results (see Figure 1C).

The patterns of estimation bias predicted by the psychophysical model described above have been clearly and uncontroversially demonstrated in a variety of perceptual tasks that were either explicitly or implicitly proportional (Hollands & Dyre, 2000; Hollands, Tanaka, & Dyre, 2002). For example, this pattern was found in a study in which participants estimated the relative quantities of black dots and white dots in a mixed collection (Varey, Mellers, & Birnbaum, 1990; see also Erlick, 1964; Spence, 1990; Spence & Krizel, 1994; Stevens & Galanter, 1957). Yet the idea that these patterns can also be seen in numerical estimation data is less readily accepted (see Opfer, Siegler, & Young, 2011; for a reply, see Barth et al., 2011). To our knowledge, Barth and

Paladino (2011) were the first to call attention to this pattern in children's number-line estimation data, as were Cohen and Blanc-Goldhammer (2011) for adults' data, even though the same pattern can be seen in estimates gathered in many previous studies (see, for example, Booth & Siegler, 2006, *Figure 2*; Ebersbach et al., 2008, *Figure 2;* Laski & Siegler, 2007, *Figures 1 & 2*; Moeller et al, 2009, *Figure 3*; Siegler & Booth, 2004, *Figure 3*; Siegler & Mu, 2008, *Figure 1*).

Why have these systematic patterns of estimation bias apparently gone unnoticed in number-line studies? We can think of two possibilities. First, there is a resounding tendency for researchers to sample heavily from the lower end of the number line and scarcely from the upper end. This is because most studies aim specifically to distinguish between logarithmic and linear fits in the context of the representational-shift hypothesis, rather than to entertain alternative hypotheses (cf. Ebersbach et al., 2008; Moeller et al, 2009). This practice focuses on participants' propensity to overestimate small numbers, but yields little data to reveal the details of underestimation patterns for larger numbers. Second, contingent on the value of the exponent (β) and on the participant's use of reference points, unbounded and cyclical power models may closely resemble logarithmic or linear models (see Figure 1), particularly if numbers near the upper end of the range are sparsely sampled.

Developmental Change in the Proportion-Judgment Account

If the observed developmental patterns in these estimation data do not indicate a log-to-linear shift, then what *is* the nature of the observed change over development? The proportion-judgment account of number-line estimation can accommodate a notion of gradual change quite different from the categorical change required by the representational-shift account. Developmental change in numerical estimation, here, may be described in terms of (at least) two distinct kinds of changes. The first of these is change in the value of the β parameter, which reflects the degree of bias associated with the estimation of individual magnitudes (such as the estimation of magnitudes of different Arabic numerals in typical number-line tasks). The observed β parameter may change gradually with age or experience, such that estimates are more accurate for older and more experienced observers (presumably with values of β eventually converging on 1, at which point proportion-judgment models are equivalent to $y = x$; see Figure 1). Some evidence in support of this idea has been found in number-line tasks (Barth & Paladino, 2011) and in perceptual tasks (Hollands & Dyre, 2000; Spence & Krizel, 1994).

Second, learning and development may lead to changes in the use of reference points, including both marked endpoints and, potentially, an inferred midpoint. Our theoretical account predicts that increased accuracy is linked to the number of reference points utilized by a participant and offers a quantitative explanation of this link. For example, a participant with a poor understanding of the task structure or an incomplete knowledge of the numerical range in question may use only the lower endpoint value, treating the task as an open-ended magnitude judgment rather than a proportion judgment. This participant should therefore produce estimates well described by an unbounded power function [1] (Stevens' Power Law; see Figure 1A).

Alternatively, a participant appropriately referencing both the given lower and upper endpoint values on the number line would produce the pattern of over- and under-estimation predicted by the one-cycle version of the proportional model (Figure 1B). A participant who infers a third reference point at the line's midpoint would produce the cyclical pattern of over- and under-estimation corresponding to the two-cycle version of the model (Figure 1C). While each progression – from an unbounded power to a one-cycle proportional to a two-cycle proportional version of the model – results in an increase in overall accuracy, this view requires no representational shift to explain developmental change from more logarithmic-looking to more linear-looking estimation patterns.

The Present Study

Prior work demonstrated that the proportion estimation model described earlier can account for estimation patterns arising from a single number-line task presented to 5- and 7-year-olds (Barth & Paladino, 2011). That study, however, provided little information about change over development. To address this question, the present study investigates how predictions of the proportion-judgment account map onto performance across age and experience through a series of cross-sectional experiments, spanning 5 years. We also present children with multiple numerical ranges to evaluate the claim that different estimation patterns for different ranges within children indicate the presence of multiple types of mental number representations (e.g. Siegler & Opfer, 2003). For all experiments, we compare the predictions of the proportional approach to those of the log-to-linear representational-shift account [2].

Broadly, we predict that children's estimation data will be better explained by the proportion-judgment account than by the representational-shift account. We also predict that estimates will follow the developmental progression described earlier, from the unbounded power function (reflecting effective use of the lower endpoint, but not the upper) to a one-cycle version of the proportional model (reflecting the use of both endpoints) to a two-cycle version (reflecting the use of both endpoints plus a midpoint). Finally, we predict that values of the β parameter will tend to change with age and experience, such that they will be closer to 1 (corresponding to perfect accuracy) for older children and for children tested on a more familiar number range.

Experiment 1

Method

Participants. A total of 33 five- and six-year-old children (16 males and 17 females, mean age 5;11) completed the task. Most children were recruited through a database of families residing in the Central Connecticut area and were tested in a quiet laboratory room. Some children were recruited and tested at local venues such as a nearby children's museum. No questions were asked about socio-economic status, race, or ethnicity, but children were presumably representative of the community from which they were drawn. In this community, 84% of adults have a high-school diploma and 30% have a bachelor's degree. Most residents identify as white (80%), black (12%) or Asian (3%) (U.S. Census Bureau, 2000).

Stimuli. Children were presented with booklets of pages (27.9 x 10.8 cm) with a 23 cm line printed in the center of each page. The left end of the line was marked with 0 and the right end of the line was marked with 20 or 100, depending on the condition. The target number for each trial was printed 2 cm above the center of the number line, as in many previous studies of number-line estimation (e.g. Booth & Siegler, 2006; Opfer & Siegler, 2007; Siegler & Opfer, 2003; Thompson & Opfer, 2010).

Design. Each child completed two conditions. The "familiar" condition used a smaller numerical range, likely to be more familiar to the child, with a number line bounded by 0 and 20. The "unfamiliar" condition used a larger numerical range, likely to be less familiar to this age group, with a number line bounded by 0 and 100. Children always completed the familiar number range first [3]. Test trials used a selection of numbers sampled roughly evenly across the given number range (0-20 or 0-100 depending on the condition), with no numbers repeated (see Appendix A for a complete list of numbers tested). The order of the trials was randomized for each child.

Procedure. The procedure was similar to that reported in Barth and Paladino (2011), which was modeled directly on the procedure of Booth and Siegler (2006). Experimenters first introduced the task for the familiar number range: “We’re going to play a game with number lines. I’ll ask you to show me where you think some number should go on the number line. When you decide where the number goes, make a line through the number line.” There was one practice trial at the beginning of each condition; for the initial 0-20 condition, the experimenter prompted children to mark a new (blank) number line to show where 10 should go. After responding, children were shown a number line marked in the middle. The experimenter asked if they knew why 10 went there and then explained, “Because 10 is half of 20, it goes right in the middle between 0 and 20. So 10 goes right there, but it’s the only number that goes there.” On the first test trial, if children marked the halfway point, the experimenter reminded them that only 10 goes in the middle. Test trials immediately followed the single practice trial. The experimenter read off the Arabic numeral printed above each number line saying, “Where would you put [X]?” for each trial. After the children responded by marking the line, experimenters concluded the trial by saying, “Thank you.” When switching to the unfamiliar condition, experimenters would say, “Now we’re going to play the game with different numbers. Zero still goes here at this end, but now 100 is here at the other end.” This was followed by the same single practice trial described above, using 50 as the halfway point.

Results

Data from children producing responses that were uncorrelated with the presented number (Spearman rank correlation, r_s, $p>.05$) or were negatively correlated (r_s, $p<.05$) on either or both conditions were excluded from the following analyses (n=8). Data from children who marked over 90% of their responses within a single region comprising 10% of the number line on either or both conditions were also excluded (n=5). This resulted in a total of 20 children: 5 five-year-olds (mean age 5;7) and 15 six-year-olds (mean age 6;6). We deliberately chose these exclusion criteria because data from children with extreme response biases or uncorrelated responses are difficult or impossible to interpret.

As a measure of general accuracy, each child's mean Percent Absolute Error (see Booth & Siegler, 2006; 2008) was computed for each task. This was calculated by dividing the absolute difference between the number corresponding to the child's estimate and the presented number by the numerical range, then multiplying the quotient by 100 to express a percentage:

$$Percent\ Absolute\ Error\ (PAE) = \frac{|Estimated\ Position - Presented\ Number|}{Numerical\ Range} \times 100$$

Patterns of estimation biases were evaluated by comparing the models described earlier to those comprising the representational-shift account (logarithmic and linear models). Models comprising the proportion-judgment account included an unbounded power function (a single-parameter version of Stevens' Power Law, with a fixed scaling factor; see Figure 1A) and the one- and two-cycle versions of the proportional power model (see Figures 1B and 1C; Hollands & Dyre, 2000). Constraints were set such that no model was allowed to project values less than zero.

Formal model comparisons determined which model best predicted children's performance patterns [4]. Comparisons were made using AICc scores (Akaike's Information Criterion, corrected for small sample sizes) as in Barth and Paladino (2011). Differences in AICc scores provide a measure of how well different models can explain data, taking into account both "goodness of fit" and model complexity, where complexity is defined in terms of number of parameters (e.g. Burnham & Anderson, 2002). Because the number of parameters in a model is not the only measure of complexity, we also used a second model selection technique for all of our group analyses (Leave-One-Out Cross-Validation or LOOCV, which assesses how a model will generalize to an independent data set [5], Browne, 2000 ; see Opfer et al., 2011, and Barth et al., 2011 for recent discussion). The findings reported here are based on AICc and LOOCV analyses, but we also report R^2 values for each fit because of this measure's greater familiarity to readers [6].

Familiar range (0-20). Performance on this task was highly accurate, with a mean PAE of just under 10% (see Figure 2A). There was little evidence of systematic estimation bias in the group median estimates, resulting in rather linear performance ($R^2 = .980$) and a relatively poor fit of the logarithmic model ($R^2 = .776$). Further, the linear model was ranked first over the two-cycle ($R^2 = .935$), one-cycle ($R^2 = .932$), and unbounded power functions ($R^2 = .897$). With a slope of .891 (intercept of ~ 0), group median performance on this numerical range is very near true accuracy ($y = x$) (see Appendix B). This result is consistent with 5- and 6-year-olds' medians considered separately (see Figures 2B and 2C).

An examination of individual children's estimates reveals, however, that only 30% of the children (n=6) produced estimates that were best explained by a linear model. All other children produced estimates that were best predicted by the proportional reasoning account (unbounded power, n=6; one-cycle, n=5; two-cycle, n=3).

Unfamiliar range (0-100). Performance was less accurate on the 0-100 number range (see Figure 2D), with a mean PAE of 15.7%. When only linear and logarithmic models are considered, our 5- to 6-year-olds' group median performance on the 0-100 number-line is better characterized by the linear model (R^2=.944) than by the logarithmic model (R^2=.838). However, the unbounded power function (R^2=.952) is preferred over all others considered here for the group median data, ranking first over the linear model as well as the one-cycle (R^2=.927), or two-cycle (R^2=.766) proportional model (see Appendix B).

The majority of individual children (65%, n=13) produced estimation patterns for the 0-100 task that were best explained by the proportion-judgment account. Most of the remaining children (n=6) produced estimates best characterized by a logarithmic model.

Finally, we examined performance patterns of 5-year-olds (the youngest portion of our sample) and 6-year-olds (the majority of the sample) on the 0-100 task separately (see Figures 2E and 2F). Although the logarithmic model fits 5-year-olds' median data reasonably well (R^2=.864), the unbounded power function (R^2=.866) offers the best explanation (with comparatively little support for linear, one-cycle, or two-cycle models as explanations of the 5-year-olds' median data). In contrast, 6-year-olds' median performance is more linear (R^2=.931) than logarithmic (R^2=.864). However, the AICc comparison and LOOCV analysis shows that 6-year-olds' systematic estimation bias (see Figure 2F) is best characterized by the one-cycle proportion-judgment model (R^2=.926). This characterization is also borne out in the individual data (with nine children exhibiting one-cycle patterns, eight children exhibiting two-cycle patterns, and seven simply showing linear estimation patterns).

Comparison across tasks. Unsurprisingly, overall accuracy decreased from the 0-20 number range to the 0-100 range. The linear model was preferred for group median estimates on the 0-20 task, while the unbounded power function was preferred for group medians on the 0-100 task. When 5- and 6-year-olds' performance patterns are considered separately, 5-year-olds' group median estimates were linear for the more familiar range, but highly biased for the less familiar range such that the unbounded power function was preferred for the 0-100 task. Likewise, 6-year-olds' group median estimates were best fit by a linear model for the familiar range, but by a one-cycle proportional model for the unfamiliar range.

Discussion

Two important findings emerge from these data. First, for the more familiar 0-20 numerical range, group median estimates were both quite linear and highly accurate. But at the individual level, even for the familiar range, estimates were better explained by the proportion-judgment account, suggesting that the linear group median estimates do not provide evidence of the use of linear mental representations by individual children. Second, for the less familiar 0-100 numerical range, estimation biases were apparent at the group level as well as at the individual level, with group medians best described by the unbounded power function. The estimates of the 5-year-olds in the sample (the low end of the age range) were also best described by the unbounded power function. The 6-year-olds' median estimates, on the other hand, were best explained by the one-cycle proportional model, consistent with the idea that these children were better able to apply appropriate proportional reasoning.

This change from an unbounded power function to a one-cycle version of the proportional model may reflect children's developing understanding of the upper end of the number line. That is, 5-year-old children's understanding of the numerical magnitudes represented by the upper end of the less-familiar number line may be limited (e.g. Barth, Starr, & Sullivan, 2009; Lipton & Spelke, 2005), rendering the reference point marked on the right side of the number line (100) unviable (cf. Barth & Paladino, 2011). These children's estimates may therefore be made relative to only one reference point (the left endpoint) such that they are effectively open-ended. Only when children become more familiar with the entire number range do they begin to understand how to use both available reference points (0 and 100), yielding a pattern of estimation predicted by the one-cycle proportional model.

Overall, these findings provide evidence of the developmental progression predicted by the proportion-judgment account. While our observation of improved accuracy for the more familiar numerical range is consistent with the findings of previous studies, the present findings did not support the idea of a logarithmic-to-linear representational shift. Rather, the observed changes were better explained by the idea that the youngest children were more likely to make open-ended judgments, particularly on the less familiar range, producing estimates consistent with an unbounded power function. The older children were better able to make appropriate proportional estimates, due probably to a greater understanding of the numerical magnitudes involved and perhaps of the task's proportional structure as well.

Do the predictions of the proportion-judgment account outperform those of the log-to-linear shift account for other age groups and numerical ranges as well? In Experiment 2, we asked older children to perform similar numerical estimation tasks.

Experiment 2

This experiment evaluated the predictions of the proportion-judgment account, comparing these to the predictions of the representational-shift account, with a slightly older age group (7- to 8-year-old children). Children again performed typical number-line estimation tasks on more and less familiar numerical ranges. The specific numerical ranges used in this experiment were 0-100 and 0-1000.

Method

Participants. Twenty-four 7- and 8-year-old children (16 males and 8 females, mean age 8;0) completed the tasks of Experiment 2. The sample included 11 seven-year-olds (mean age

7;5) and 13 eight-year-olds (mean age 8;6). They were drawn from the same population as Experiment 1 and tested under the same circumstances.

Stimuli. Stimuli included booklets identical to those used in Experiment 1, with the exception that each number line was labeled with a 0 at the left end and a 100 or 1000 at the right end (depending on the condition).

Design. As with Experiment 1, each child completed two conditions. For this age group, number lines for the "familiar" range were bounded by 0 and 100 and number lines for the "unfamiliar" range were bounded by 0 and 1000. Each condition presented a series of numbers sampled roughly evenly across the entire range (see Appendix A). The order of trials was randomized for each child.

Procedure. The procedure for this experiment was the same as the procedure for Experiment 1, except that Experiment 2 used larger numerical ranges as described above.

Results

All of the children in this age group produced responses correlated with the presented number (r_s, $p<.05$) and no child responded only within a small portion (<10%) of the number line; therefore data from all of the children were included. Analyses were conducted as in Experiment 1: an overall measure of accuracy (PAE) was computed, and explanations of children's estimates were evaluated by comparing AICc scores and LOOCV indices for logarithmic, linear, and proportion-judgment models.

Familiar range (0-100). The two-cycle version of the proportional model (consistent with the use of the two explicit endpoint values plus a midpoint; Hollands & Dyre, 2000) provided the best fit for the 7-to-8-year-old group's median data on the 0-100 task (R^2=.990), ranking it first over linear (R^2=.986), one-cycle (R^2=.984), and logarithmic (R^2=.856) models for group performance (see Figure 3A; Appendix C). Analyses of individuals' estimates reveal a similar trend, with the majority of the children (71%) producing one- or two-cycle estimation patterns (n=9 and n=8 respectively). The remaining children's estimates were best fit by a linear model (n=7).

Two-cycle proportion-judgment models are also preferred for both the 7-year-olds' and 8-year-olds' estimates considered separately (Figures 3B; 3C; and Appendix B). Interestingly, the value of the exponent (β) for the 2-cycle fit of 7-year-olds' estimates (β=.589) is notably lower than the value for 8-year-olds' (β=.717).

Unfamiliar range (0-1000). The two-cycle version of the proportional model (R^2=.994) ranked first over linear (R^2=.980), one-cycle (R^2=.967), or logarithmic (R^2=.560) models for group performance on the less familiar range (0-1000) (see Figure 3D; Appendix C). Again, analyses of individual children's estimates show a similar distribution, with the majority of children (75%) producing estimates best accounted for by one-cycle (n=3) or two-cycle (n=14) versions of the proportional model.

The two-cycle model provides the best account of estimation performance in both the 7-year-old and 8-year-old groups (see Figures 3E and 3F). As in the familiar number-line task, β-values increased across age groups (β=.348 for 7-year-olds' medians and β=.560 for 8-year-olds' medians), consistent with more accurate estimation by the older children.

Comparison across tasks. Even though performance on both the familiar and unfamiliar tasks shows the same 2-cycle pattern of over- and underestimation, mean PAE scores reveal that performance not only improves across age, but also across tasks (with mean PAE of 6.8% on the 0-100 range and 10.4% on the 0-1000 range). Consistent with this finding of increased accuracy, observed β-values also increase across tasks. This is true for median estimates of the entire 7-

and 8-year-old group (β=.492 for the 0-1000 condition and β=.669 for the 0-100 condition) as well as for 7-year-olds (β=.348 for the 0-1000 condition and β=.589 for the 0-100 condition) and 8-year-olds (β=.560 for the 0-1000 condition and β=.717 for the 0-100 condition) considered separately. In fact, β-values for all individuals producing consistent proportion-judgment patterns across tasks were higher (closer to 1) on the familiar range than on the unfamiliar range.

Discussion

Number-line estimation performance in the 7- and 8-year-old age range was best characterized by the two-cycle version of the proportional model. This pattern of performance is predicted when observers use both the endpoints of the number line and a midpoint as reference points for their estimates. It appears that the 7- and 8-year-olds we tested were able to make use of a central reference point, and that estimation accuracy benefited as a result. Children's estimation patterns provided no support for the log-to-linear-shift account.

These findings, especially when considered in concert with the findings of Experiment 1 on the 0-100 task, show quantitatively that one source of increased accuracy in number-line estimation comes from the use of additional reference points. The youngest children in Experiment 1 were apparently unable to make consistent use of upper endpoints (reflected by estimates best described by the unbounded power function). The older children in Experiment 1 produced estimates that were best explained by the one-cycle proportional model, consistent with the reliable use of both lower and upper explicit endpoint values. The slightly older children of Experiment 2, however, produce the estimation patterns that are characteristic of the use of 3 reference points (the two explicit endpoints plus a midpoint). Therefore the findings of Experiments 1 and 2 show the predicted developmental progression of one source of increased accuracy in this task: the appropriate use of additional reference points (see Figure 1).

Experiment 2 further shows that improved estimation accuracy does not arise solely from the use of additional reference points. Children's estimates also revealed an overall decrease in estimation bias with age (reflected by β-values that were closer to 1 for older children) and for more familiar ranges. The idea that improved accuracy is associated with this kind of change, as predicted by the proportion-judgment account, is distinct from the idea that additional reference points confer improved accuracy.

In Experiment 3, 8- to 10-year-old children performed similar tasks for even larger numerical ranges. Again, the predictions of the proportion-judgment account were compared to those of the log-to-linear shift account.

Experiment 3

This experiment evaluated number-line estimation by 8- to 10-year-old children. Children again performed typical number-line estimation tasks on more and less familiar numerical ranges. The specific numerical ranges used in this experiment were 0-1000 and 0-100000.

Method

Participants. Participants included 30 eight-, nine-, and ten-year-old children (19 males and 11 females, mean age 9;3) drawn from the same population and tested in the same locations as Experiments 1 and 2.

Stimuli. Test booklets were identical to those of Experiments 1 and 2 except that number lines were labeled with a 0 at the left end and a 1000 or 100000 at the right end (depending on condition).

Design. Children completed each of two conditions: the familiar range was bounded by 0 and 1000 and the unfamiliar range was bounded by 0 and 100000. Test trials were randomly selected from the entire number range (see Appendix A). The order of the trials was randomized for each child.

Procedure. Procedures were identical to those of Experiments 1 and 2 except for the numbers used.

Results

One child's data was excluded from further analyses because responses were not correlated with the presented numbers on one of the two conditions (r_s, $p>.05$). Another child was excluded for developmental delays (by parental report). This resulted in a total of 28 children: 12 eight-year-olds (mean age 8;4), 8 nine-year-olds (mean age 9;7), and 8 ten-year-olds (mean age 10;5).

Familiar range (0-1000). As with the previous two experiments, performance on the familiar number range was relatively accurate, with mean PAE of 8.1%. As in Experiment 2, the two-cycle proportional model (R^2=.990) was ranked first over linear (R^2=.986), one-cycle (R^2=.982) and logarithmic (R^2=.482) models for group performance on the familiar 0-1000 task (see Figure 4A; Appendix D). The β-value for the two-cycle fit was .682 (slightly higher than the .492 β-value for the 7- to 8-year-old group on the same task).

An evaluation of individual performance revealed that many of the children who produced estimates best characterized by a proportional model showed a one-cycle pattern (n=9). Moreover, estimates from most of the children producing a one-cycle pattern follow a distinct *under*-then-*over* pattern, with corresponding β-values greater than 1 (n=5, mean age 9;11). This unexpected finding, which contrasts with the over-then-under patterns found in the previous experiments (see Figures 2F and 3), will be addressed further in later sections.

Analyses of the 8- and 9-year-olds' estimates considered separately (Figures 4B and 4C) revealed a pattern similar to that of the 7- and 8-year-olds of Experiment 2, with median estimates in both cases best characterized by the two-cycle proportional model (with a β-value of .477 for 8-year-olds and .717 for 9-year olds). Ten-year-olds' performance, on the other hand, was best characterized by a one-cycle version (R^2=.979). And, consistent with the aforementioned analysis of individual children in which the older children produced under-then-over estimation patterns, 10-year-olds' median estimates show the same pattern of bias, resulting in β-values that exceed 1 (β=1.243; Figure 4D).

Unfamiliar range (0-100000). As a group, 8- to 10-year-olds' estimates (Figure 4E) are best explained by the two-cycle version of the proportional model (R^2=.989, β=.540), outperforming linear (R^2=.970), one-cycle (R^2=.965), or logarithmic (R^2=.309) models (see Appendix D). Analyses of individual performance on this range yield results similar to those from the familiar range. A majority of children (71%) produce estimates best predicted by the proportional model; some children produce estimates best explained by a one-cycle version (n=6), rather than a two-cycle version (n=13). Estimates of all children producing a one-cycle pattern showed an *under*-then-*over* pattern, with β-values greater than 1 (mean age 9;11).

Separating children according to year-of-age shows that the two-cycle version of the proportional model provides the best account of estimation biases for both 8- and 9-year-olds (R^2=.944 and R^2=.980 respectively), with the β-values obtained for group medians increasing across age groups (β=.349 for 8-year-olds and β=.547 for 9-year-olds; see Figures 4F and 4G). Ten-year-olds' performance, on the other hand, was best characterized by a one-cycle version (R^2=.978) with a β-value over 1 (β=1.398), corresponding to an under-then-over pattern of bias (Figure 4H).

Comparison across tasks. Overall group performance was more accurate for the familiar range (mean PAE=8.1%) than the unfamiliar range (mean PAE=10.7%). The two-cycle version of the proportional model was preferred as an explanation of group performance for both the familiar and unfamiliar ranges, with the degree of bias smaller for the familiar range (β=.682) than for the unfamiliar range (β=.540). This finding parallels the results of comparisons of 7- and 8-year-olds' performance on the 0-100 and 0-1000 number lines from Experiment 2.

Comparisons across the familiar and unfamiliar number-line tasks also showed that the *under*-then-*over* pattern emerges with age and does not necessarily change according to familiarity with a given number range. Rather, 10-year-olds' group median estimates follow this pattern on the more familiar 0-1000 range as well as the less familiar 0-100000 range. On the individual level, three of the five children who produced this pattern on the 0-1000 range did so on the 0-100000 range as well, while the other two did not; three more 10-year-olds showed this pattern on the 0-1000000 but not the 0-1000 range. These findings provide some evidence that this pattern may be characteristic of children's number-line estimation strategies at later stages of development but that it is not contingent on their overall familiarity with the number range.

Discussion

Number-line estimation performance in the 8- to 10-year-old age group was best characterized by the two-cycle proportional model, suggesting that these children were able to make use of central reference points when making their estimates, much like the 7- and 8-year-olds tested on smaller numerical ranges in Experiment 2. Children's estimation patterns again provided no support for the log-to-linear-shift account. We discuss the overall findings from this experiment in combination with those of the other two experiments below.

General Discussion

The idea that mental representations of numerical magnitude undergo a categorical shift has had a major influence on theoretical approaches to mathematical cognition as well as to cognitive development more broadly. Performance patterns on various numerical estimation tasks led to the development of this idea. However, the present work provides evidence for a different interpretation, building upon previous research showing that typical patterns of number-line estimation performance are predicted by psychophysical models of proportion estimation (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Sullivan et al., 2011). In the present paper, we evaluated the relative abilities of the representational-shift and the proportion-judgment views to account for children's estimates and further explored the sources of developmental change underlying these patterns of performance.

Children in three age groups (5- to 6-year-olds in Experiment 1, 7- to 8-year-olds in Experiment 2, and 8- to 10-year-olds in Experiment 3) completed typical number-line estimation tasks for both a familiar and a less familiar numerical range. We assessed the explanatory power of the quantitative models comprising each theoretical view for median estimates provided by each of the targeted broad age groups (5- to 6-year-olds, 7- to 8-year-olds, and 8- to 10-year-olds). We also evaluated explanations of performance at more fine-grained subgroups (5-, 6-, 7-, 8-, 9-, and 10-year-olds) and at the individual level. Overall, these data provide overwhelming evidence in favor of the proportion-judgment account: for both group and individual analyses [7], this account provided the best explanation of estimation patterns. We also emphasize that our findings do not rest on the choice of a particular model selection technique: in all cases, AICc and LOOCV analyses yielded consistent results.

Improvements in estimation performance are well characterized in terms of the sources of change described earlier. First, children's patterns of performance suggest that accuracy can improve through changes in the use of reference points (see Figure 1). The data suggest that very young children (such as 5-year-olds; see also Barth & Paladino, 2011) do not evaluate the upper endpoint of the number line appropriately. Some may be unable to use this reference point at all, treating the task as an open-ended magnitude judgment and producing highly biased estimates well characterized by an unbounded power function (Figure 1A). A child entirely lacking the ability to reason about proportions might also produce this pattern of performance; however, due to substantial evidence for various forms of proportional reasoning in young children (Barth, Baron, Carey, & Spelke, 2009; Boyer, Levine, & Huttenlocher, 2008; Boyer & Levine, in press; Duffy, Huttenlocher, & Levine, 2005; Jeong, Levine, & Huttenlocher, 2007; McCrink & Spelke, 2010; Sophian & Wood, 1997; Spinillo & Bryant, 1991, 1999) and even infants (McCrink & Wynn, 2007), it seems more likely that this pattern arises from the failure to appropriately apply a proportional strategy to the task, or from a lack of accurate knowledge of the meaning of the upper endpoint numeral (and probably other numerals at the high end of the range), rather than from a total lack of proportional competence.

Slightly older children produce estimation patterns that suggest they are able to appropriately consider both endpoints when deciding on the location of the given number. This ability, appearing around age 6 in our task, allows children to make more accurate judgments and results in a pattern of bias characterized best by the one-cycle version of the proportional power model (Figure 1B). This might arise from a more robust ability to apply a proportional strategy to the task, or from a better understanding of the meanings of the larger numerals involved (consistent with previous findings; e.g. Barth, Starr, & Sullivan, 2009; Lipton & Spelke, 2005; see also Matthews & Chesney, 2011, for related findings in adults), or both. Somewhat older children (7- to 10-year-olds in this sample) appear able to use a midpoint value as a reference in addition to the two explicitly labeled endpoints. These children show a biased estimation pattern that repeats itself around the midpoint, producing estimates best characterized by the two-cycle version of the proportional power model (Figure 1C; Hollands & Dyre, 2000).

A second source of improvement inherent to the proportion-judgment account is reflected in the value of the β parameter, which gradually increases with age. For example, consider children's performance on the 0-1000 task: across Experiments 2 and 3, a total of 52 children between 7 and 10 years of age completed this task. Of the forty children whose individual estimation patterns were best predicted by models of proportion judgment, corresponding β-values strongly correlate with age in months, $r_s(40)=.652$, $p<.001$ (see Figure 5).

What We Can Conclude About Mental Representations of Number

It is tempting to conclude from these and other recent findings that, once the proportional nature of the number-line estimation task has been taken into account, we may make simple and useful inferences about the nature and development of mental number representations from the data it yields. For example, the underlying power function in the proportion-judgment model could describe a representation of numerical magnitude that is highly compressed in younger children and becomes gradually less compressed with age, as evidenced by smooth change in the β parameter from values far below 1 to values near 1 (see also Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Sullivan et al., 2011).

Although these ideas about compressed numerical magnitude representations may be correct (and broadly consistent with findings from other paradigms, e.g. Merten & Nieder, 2009), we urge caution in making such inferences directly from paradigms structured like typical number-line tasks for the following reason. These kinds of tasks involve the production of a spatial proportion – involving lengths or distances – as well as the estimation of a numerical proportion; or, in the case of inverse "position-to-number" tasks (Ashcraft & Moore, 2011; Siegler & Opfer, 2003), the estimation of a spatial proportion and the production of a numerical proportion. Attributing performance patterns to numerical processing alone requires assuming that the spatial component of the task does not in itself contribute to variations in performance – that our estimation and production of spatial proportion is veridical. But a substantial set of findings shows that the estimation of spatial position is biased. For example, experimental paradigms structurally similar to number-line estimation have shown that both children and adults exhibit systematic bias when placing a mark on a remembered position within a linear space, or finding a hidden object in a long rectangular sandbox (e.g. Huttenlocher et al., 1994). Thus, because similar patterns of estimation bias are seen in a spatial task with no numbers, we cannot assume that bias in a spatial-numerical mapping task arises from numerical (not spatial) processing [8].

Clearly, however, the β values observed in the present data do reflect bias in numerical processing, and not just in spatial processing. This is demonstrated by the finding that children produced more biased estimates for larger, less familiar numerical ranges than for smaller, more familiar ranges even though each of these tasks has identical spatial components. For instance, 7- and 8-year-olds' performance on both the more and less familiar number ranges was best characterized by the two-cycle version of the proportion-judgment model, but the value of β was closer to 1 on the familiar task than the unfamiliar task (β=.669 and β=.492, respectively; see Figures 3A and 3D). In fact, 21 of the 23 children who produced the same proportion-judgment estimation pattern across tasks yielded β-values that were closer to 1 on the familiar range (see Teghtsoonian, 1973, for related findings in different tasks). This finding suggests that numerical processing does contribute, but that β is not a simple index of some stable characteristic of the child's mental representation of numerical magnitude (this may not be surprising given the considerable debate over the psychological meaning of the parameters of Stevens' Law; e.g. Laming, 1997; Teghtsoonian, 1973). Moreover, given the well-known spatial biases that arise in similarly structured tasks, spatial processing likely also contributes to the values of β we observe here, and to the estimates of children and adults in number-line tasks in general.

Questions for Future Research

An unexpected finding arose from the oldest children's data. We found that the direction of bias seen in younger children (the typical over-then-under pattern seen for example in our 7- and 8-year-olds' data, Figure 3) reversed itself with age and experience, with the oldest children (our 10-year-olds) producing instead an under-then-over pattern such that their observed β-values were greater than 1 (see Figures 4D and 4H). This suggests that as children's estimates gradually become less biased, their estimates do not simply become more and more accurate, constantly approaching perfect performance (with β-values growing until they approximate 1). Rather, these under-then-over estimation patterns in our oldest children suggest that the values of β may eventually "overshoot" 1. To our knowledge, the emergence of this reversal in the direction of number-line estimation bias over development has not yet been reported, so the reason for it is not yet well understood. However, this finding appears to be robust, as a slight under-then-over pattern has been observed in adult's estimation performance (see Cohen & Blanc-Goldhammer, 2011) and a reversal in the direction of bias in older vs. younger children has also been observed in a related task (Ashcraft & Moore, 2011; Slusser & Barth, manuscript in preparation).

Two findings in particular may be relevant to future investigations of this reversal. First, β-values over 1 were associated only with children whose estimates were best explained by the one-cycle version of the proportion-judgment model, suggesting that these children did not make their estimates in relation to unmarked central reference points. Second, the tendency to produce the under-then-over pattern seems to be consistent within children: those who generated this estimation pattern for a more familiar numerical range were also likely do so for a less familiar range. One possible speculative explanation of the reversal is that older children (and perhaps adults) are, at least implicitly, aware of an erroneous tendency to overestimate values on the lower end of the number line and therefore overcorrect for this bias, resulting in the opposite under-then-over trend. Further studies, however, are needed to explore the reasons for this pattern of performance.

Conclusions

We believe these data provide strong evidence against two prominent theoretical ideas: that children's number-line estimates transparently indicate the forms of their mental representations of number, and that developmental changes in estimation patterns implicate a discontinuous shift from logarithmic to linear mental representations. These studies show that understanding number-line estimation and structurally similar tasks in terms of proportion estimation can explicate patterns of bias in children's performance. The systematic patterns emerging from estimation performance across this 5-year span also provides evidence that increased task proficiency across development is attributable to a process of developmental change with at least one gradual component.

This work further shows that models of proportion estimation, developed originally for perceptual tasks, can be usefully applied to tasks that involve more abstract assessments of numerical magnitude. That is to say, children's (and adults') estimates of numerical magnitude in bounded tasks apparently share many characteristics with estimates of non-numerical perceptual magnitude in tasks that are both explicitly and implicitly bounded (see Cohen & Blanc-Goldhammer, 2011; Hollands & Dyre, 2000). We also speculate that these findings, despite the rather specific type of task investigated here, may relate usefully to other work showing that proportional reasoning is perhaps more ubiquitous than has commonly been thought (see Balci & Gallistel, 2006; Hollands & Dyre, 2000).

References

Ashcraft, M., Moore, A. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology, 111*, 246-267.

Balci, F., & Gallistel, C. R. (2006). Cross-domain transfer of quantitative discriminations: Is it all a matter of proportion? *Psychonomic Bulletin & Review, 13*, 636-642.

Barth, H., Baron, A., Spelke, E., & Carey, S. (2009). Children's multiplicative transformations of discrete and continuous quantities. *Journal of Experimental Child Psychology, 103*, 441-454.

Barth, H.C., & Paladino, A.M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14,* 125-135.

Barth, H., Slusser, E., Cohen, D., & Paladino, A.M. (2011). A sense of proportion: Commentary on Opfer, Siegler, & Young. *Developmental Science, 14*, 1205-1206.

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

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology, 46*, 545-551.

Booth, J.L., & Siegler, R.S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42,* 189-201.

Booth, J.L., & Siegler, R.S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*, 1016-1031.

Boyer, T.W., & Levine, S.C. (in press). Child proportional scaling: Is 1/3 = 2/6 = 3/9 = 4/12? *Journal of Experimental Child Psychology*.

Boyer, T.W., Levine, S.C., & Huttenlocher, J. (2008). The development of proportional reasoning: Where young children go wrong. *Developmental Psychology, 44*, 1478-1490.

Brannon, E.M., Wusthoff, C.J., Gallistel, C.R., & Gibbon, J. (2001). Numerical subtraction in the pigeon: evidence for a linear subjective number scale. *Psychological Science, 12*, 238-243.

Browne, M. (2000). Cross-validation methods. *Journal of Mathematical Psychology, 44*, 108-132.

Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference. A practical information-theoretic approach (2nd ed). New York: Springer.

Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on "Log or linear? Distinct intuitions of the number scale in Western and Amazonian Indigene cultures." *Science*, *323*, 38b-38b.

Cohen, D.J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review*, 331-338.

Cordes, S., Gelman, R., Gallistel, C.R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review, 8*, 698-707.

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

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Response to Comment on "Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures." *Science, 320*, 1217-1220.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2009). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science, 320*, 1217-1220.

Duffy, S., Huttenlocher, J., & Crawford, L. E. (2006). Children use categories to maximize estimation. *Developmental Science, 9*, 597-603.

Duffy, S., Huttenlocher, J., & Levine, S.C. (2005). It is all relative: How young children encode extent. *Journal of Cognition and Development*, 6, 51-63.

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.

Erlick, D.E. (1964). Absolute judgments of discrete quantities randomly distributed over time. *Journal of Experimental Psychology, 67*, 475-482.

Gallistel, C. R. (2011). Mental magnitudes. In *Space, time, and number in the brain: Searching for the foundations of mathematical thought*. S. Dehaene & E. Brannon (Eds). London: Academic Press.

Gallistel, C.R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44,* 43-74.

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33*, 277-299.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*, 1343-1359.

Hollands, J.G., & Dyre, B.P. (2000). Bias in proportion judgments: the cyclical power model. *Psychological Review, 107*, 500-524.

Hollands, J.G., Tanaka, T., & Dyre, B.P. (2002). Understanding bias in proportion production. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 563-574.

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

Jeong, Y., Levine, S., & Huttenlocher, J. (2007). The development of proportional reasoning: Effect of continuous versus discrete quantities. *Journal of Cognition and Development, 8*, 237-256.

Krasa, N., & Shunkwiler, S. (2009). Number sense and number nonsense: Understanding the challenges of learning math. Baltimore: Brookes Publishing.

Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral and Brain Sciences, 12*, 251–320.

Laming, D. (1997). *The measurement of sensation*. New York: Oxford University Press.

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 78*, 1723-1743.

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

Matthews, P. G., & Chesney, D. (2011). Straightening up: Number line estimates shift from log to linear with additional information. *Proceedings of the Cognitive Science Society*, 1936-1941.

McCrink, K., & Spelke, E.S. (2010). Core multiplication in childhood. *Cognition, 116*, 204-216.

McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science, 18*, 740-745.

Merten, K. & Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. *Journal of Cognitive Neuroscience, 21*, 333-346.

Moeller, K. & Nuerk, H.-C. (2011). Psychophysics of numerical representation: why seemingly logarithmic representations may be multi-linear. *Zeitschrift für Psychologie/Journal of Psychology*, *219*, 65-71.

Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2009). Children's early mental number line: logarithmic or decomposed linear? *Journal of Experimental Child Psychology, 103*, 503-515.

Opfer, J.E., & Siegler, R.S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55,* 169-195.

Opfer, J.E., Siegler, R.S., & Young, C.J. (2011). The powers of noise-fitting: Reply to Barth and Paladino. *Developmental Science, 14*, 1194-1204.

Opfer, J.E., & Thompson, C.A. (2008). The trouble with transfer: insights from microgenetic changes in the representation of numerical magnitude. *Child Development, 79,* 788-804.

Poulton, E.C. (1989). *Bias in quantifying judgments*. Sussex: Lawrence Erlbaum.

Shepard, R.N., Kilpatric, D.W., & Cunningham, J.P. (1975). The internal representation of numbers. *Cognitive Psychology, 7,* 82-138.

Siegler, R.S., & Booth, J.L. (2004). Development of numerical estimation in young children. *Child Development 75*, 428-444.

Siegler, R.S., & Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychological Science, 19,* 759-763.

Siegler, R.S., & Opfer, J.E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychological Science 14,* 237-250.

Siegler, R.S., & Ramani, G.B. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79,* 375-394.

Siegler, R.S., & Ramani, G.B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology, 10*, 545-560.

Siegler, R.S., Thompson, C. A., & Opfer, J.E. (2009). The logarithmic-to-linear shift: one learning sequence, many tasks, many time scales. *Mind, Brain, and Education, 3*, 143-150.

Siegler, R.S., Thompson, C.A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*, 273-296.

Slusser, E., & Barth, H. (n.d.). An Excel worksheet for proportion-judgment analyses of number-line data. Retrieved from http://eslusser.faculty.wesleyan.edu on 7/7/11.

Sophian, C., & Wood, A. (1997). Proportional reasoning in young children: The parts and the whole of it. *Journal of Educational Psychology, 89*, 309-317.

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 683-692.

Spence, I., & Krizel, P. (1994). Children's perception of proportion in graphs. *Child Development, 65,* 1193-1213.

Spinillo, A., & Bryant, P. (1991). Children's proportional judgments: The importance of "half". *Child Development, 62*, 427-440.

Spinillo, A., & Bryant, P. (1999). Proportional reasoning in young children: Part-part comparisons about continuous and discontinuous quantity. *Mathematical Cognition, 5*, 181-197.

Stevens, S.S., & Galanter, E.H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology, 54,* 377-411.

Sullivan, J., Juhasz, B., Slattery, T., & Barth, H. (2011). Adults' number-line estimation strategies: evidence from eye movements. *Psychonomic Bulletin & Review, 18,* 557-563.

Teghtsoonian, R. (1973). Range effects in psychophysical scaling and a revision of Stevens' Law. *American Journal of Psychology, 86*, 3-27.

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.

U. S. Census Bureau (2000). *State and Country QuickFacts. Data derived from Population Estimates, 2000*. Retrieved July 13, 2011 from http://www.census.gov/

Varey, C.A., Mellers, B.A., & Birnbaum, M.H. (1990). Judgments of proportions. *Journal of Experimental Psychology, 16,* 613-625.

Vlaev, I., Chater, N., Stewart, N., & Brown, G. (2011). Does the brain calculate value? *Trends in Cognitive Science, 15*, 546-554.

Young, C.J., & Opfer, J.E. (2011). Psychophysics of numerical representation: A unified approach to single- and multi-digit magnitude estimation. *Zeitschrift für Psychologie/Journal of Psychology, 219*, 58-63.

Acknowledgments

This work was supported in part by NSF-DRL0950252 to HB, a Wesleyan University Psychology Department Postdoctoral Fellowship to ES, and the Wesleyan McNair Fellows Program. We thank the participating families, schools, teachers, and administrators who made this work possible. We also thank Elizabeth Chase for her helpful feedback and Anima Acheampong, Shipra Kanjlia, Mattie Liskow, and Kyle MacDonald for their assistance with stimulus preparation and data collection.

Footnotes

[1] It is important to note that this pattern of performance (estimates that are well described by simple power or log functions) could arise in this task from many possible strategies likely to be used by young children with a poor understanding of the task structure and/or little knowledge of the numbers presented. The use of essentially open-ended magnitude judgments is just one example; counting up from the left side of the number line by arbitrarily chosen units is another. Some children, moreover, may try to reference the upper endpoint but lack the accurate knowledge of its value that is necessary for a reasonably accurate proportion estimate (that is, they try to compare the magnitude of "30" to the magnitude of "100," but they don't yet know what "100" means). These children will also produce similar power- or logarithmic-looking estimates (Barth & Paladino, 2011). For these and other reasons, the applicability of a particular type of function to number-line estimation patterns should not be taken as evidence for a corresponding mental representation of number.

[2] We do not test the predictions of segmented linear models for two reasons. First, they were explicitly tested by Barth & Paladino (2011), and model selection procedures (Burnham & Anderson, 2002) found them to be unsupported. Second, visual inspection strongly suggests that the data we report here provide little if any support for such models.

[3] As part of another study, children first completed a Position to Number (PN) task in which they were given a marked number line and asked which number went with that mark. Data from the PN task will not be discussed here.

[4] A Microsoft Excel worksheet for performing simple versions of these analyses (Slusser & Barth, n.d.) is freely available at http://eslusser.faculty.wesleyan.edu/.

[5] This analysis first divides each data set into a calibration sample of N-1 and a validation sample of 1. The model of interest is then fit to the calibration sample. This is repeated such that each observation in the sample is used once as the validation sample. An error index (the mean standard error or MSE) for each iteration is calculated, with the average MSE across all iterations summarizing the model's fit.

[6] Comparisons of R^2 values do not take model complexity into account so they should not be used for comparing the models tested here.

[7] Though the 5- and 6-year-old group's median performance on the 0-20 number-line was ostensibly linear, individual analyses showed that the majority of these children (14 out of 20) produced estimates that are more consistent with the proportion-judgment account.

[8] Very similar patterns of bias arise in a variety of other tasks involving the estimation of magnitudes within a bounded response range (e.g. the width of schematic fish, Duffy, Huttenlocher, & Crawford, 2006; Huttenlocher, Hedges, & Vevea, 2000; and the lightness of a grey square, Huttenlocher et al., 2000). Although these studies are conceptually situated in a larger literature examining category effects on stimulus judgment, it is worth noting that many of these findings appear consistent with the proportion estimation models used here. The relation between these models – the less constrained but perhaps more broadly generalizable Category Adjustment Model (e.g. Huttenlocher et al., 2000) and the more parsimonious but possibly less general Cyclical Power Model of Proportion Judgment (e.g. Hollands & Dyre, 2000; Hollands et al., 2002) – remains to be determined.

Figure Captions

Figure 1. Predictions of the proportion judgment account. Figure 1A depicts a one-parameter unbounded power function (with the scaling factor, α, fixed at 10). Figures 1B and C depict one- and two-cycle versions of the proportional power model, respectively. The two-cycle version of the model depicts the predicted estimation pattern for observers who use the midpoint of the number line as a reference point.

Figure 2. Median estimates of 5- and 6-year-olds on each task. Estimated number corresponds to the marked position on the number line. The solid line represents the preferred model. The dashed line shows $y=x$.

Figure 3. Median estimates of the 7- and 8-year-olds on each task. Estimated number corresponds to the marked position on the number line. The solid line represents the preferred model. The dashed line shows $y=x$.

Figure 4. Median estimates of 8- to 10-year-olds on each task. The solid line represents the preferred model. The dashed line shows $y=x$.

Figure 5. Values of the β parameter corresponding to the estimates of each child whose performance is best predicted by the proportion judgment account on the 0-1000 number-line task (n=40).

Figure 1.

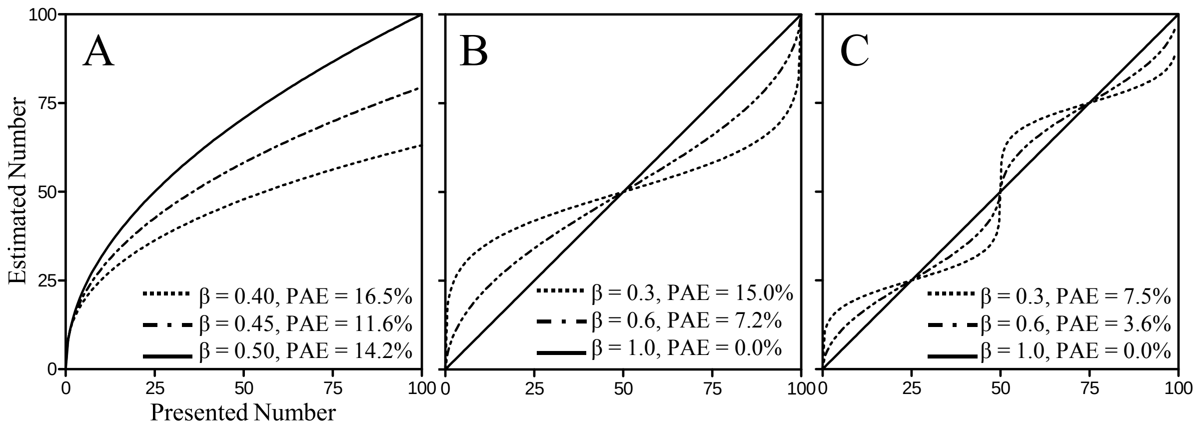

A
B
C
Estimated Number
Presented Number
β = 0.40, PAE = 16.5%
β = 0.45, PAE = 11.6%
β = 0.50, PAE = 14.2%
β = 0.3, PAE = 15.0%
β = 0.6, PAE = 7.2%
β = 1.0, PAE = 0.0%
β = 0.3, PAE = 7.5%
β = 0.6, PAE = 3.6%
β = 1.0, PAE = 0.0%

Figure 2.

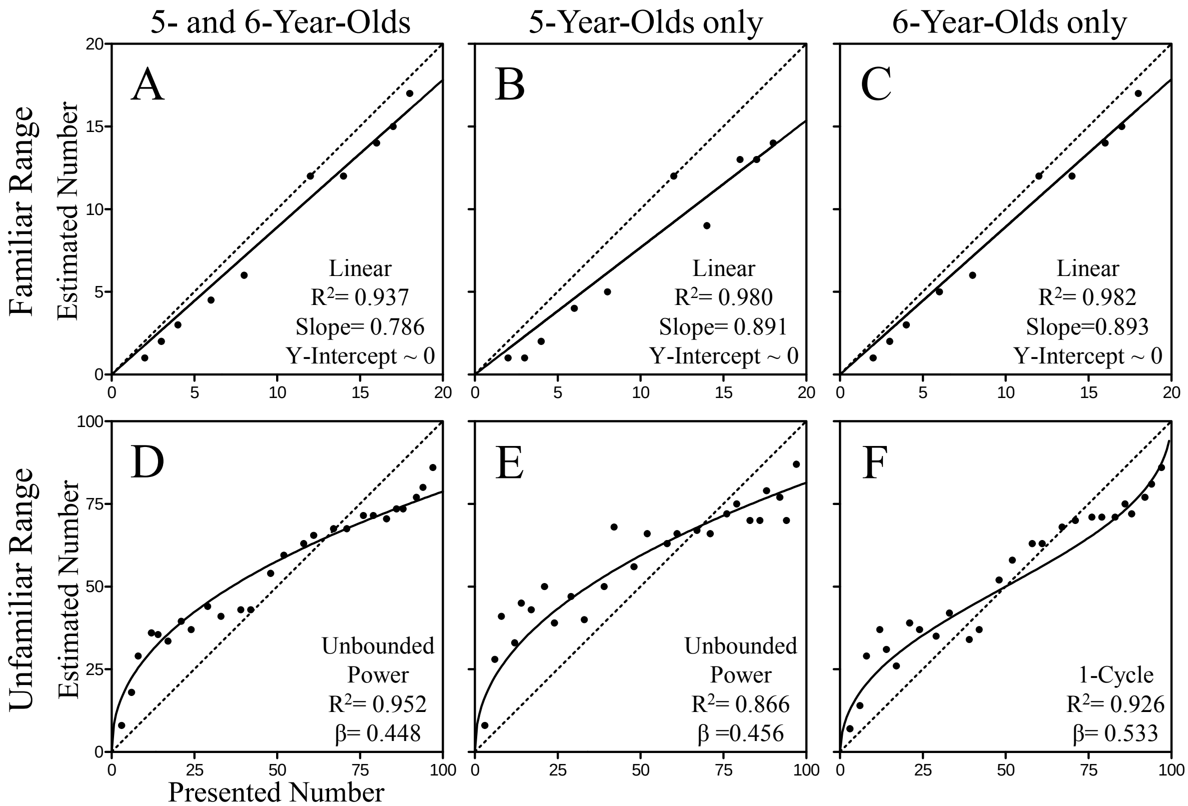
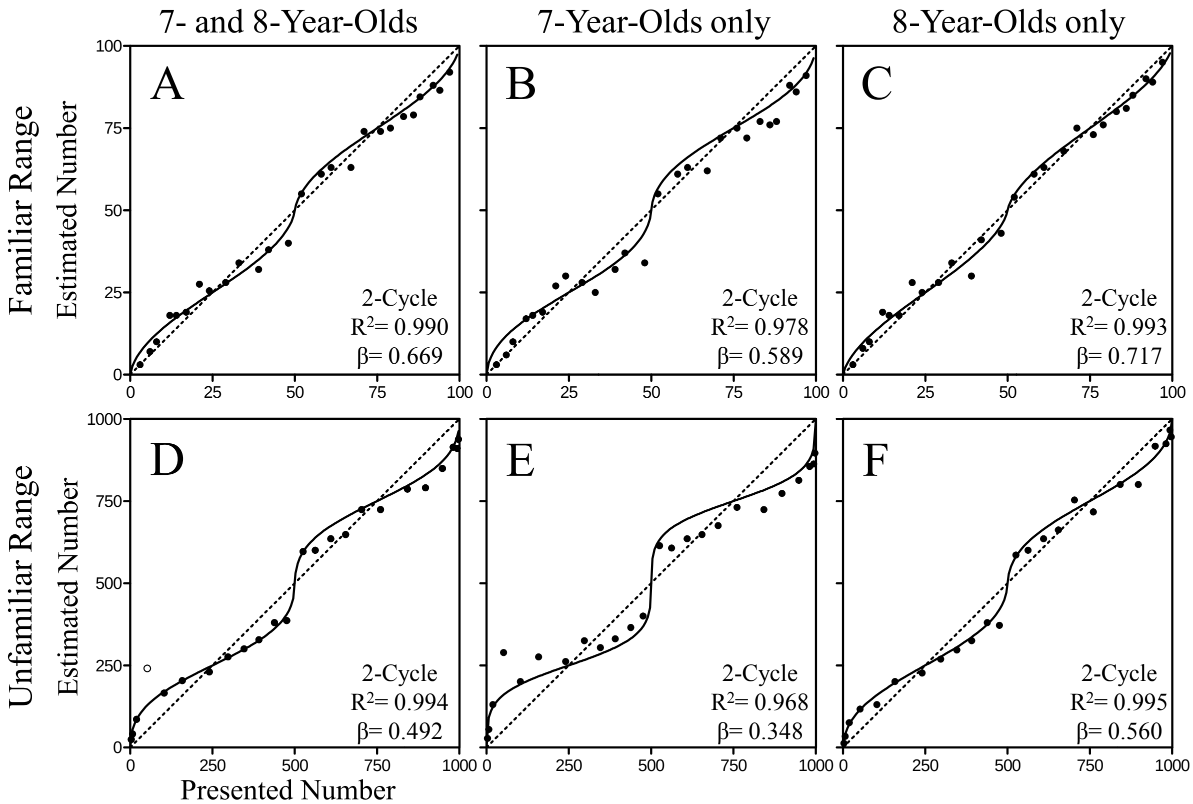
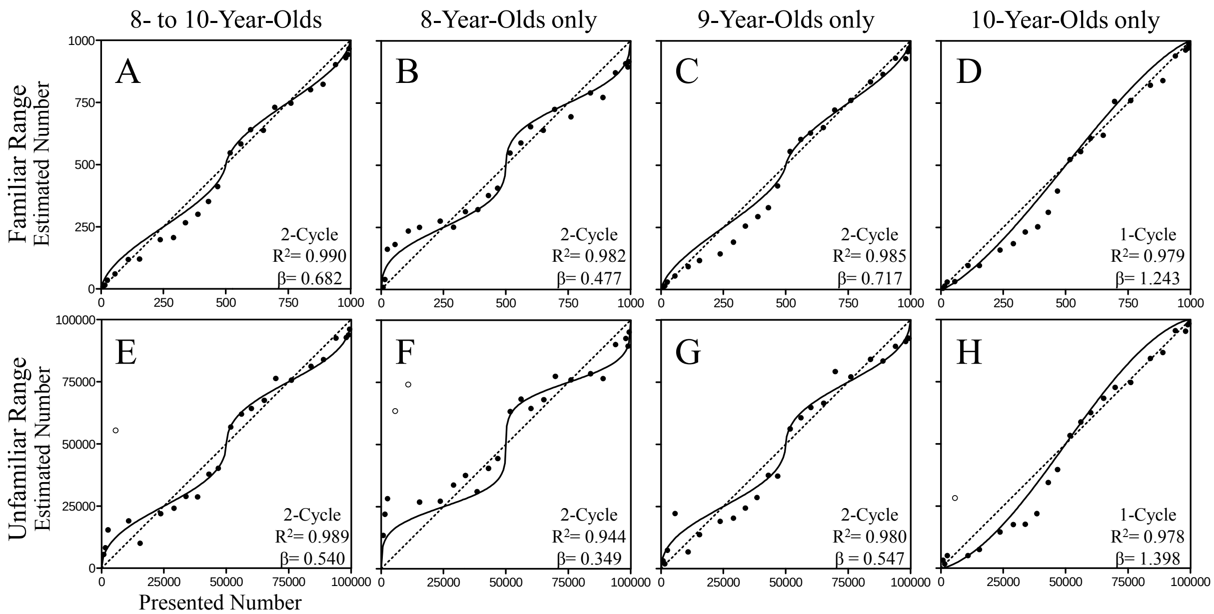

5- and 6-Year-Olds
5-Year-Olds only
6-Year-Olds only
Familiar Range
Estimated Number
A
Linear
R2= 0.937
Slope= 0.786
Y-Intercept ~ 0
B
Linear
R2= 0.980
Slope= 0.891
Y-Intercept ~ 0
C
Linear
R2= 0.982
Slope=0.893
Y-Intercept ~ 0
Unfamiliar Range
Estimated Number
D
Unbounded
Power
R2= 0.952
β= 0.448
E
Unbounded
Power
R2= 0.866
β =0.456
F
1-Cycle
R2= 0.926
β= 0.533
Presented Number

Figure 3.



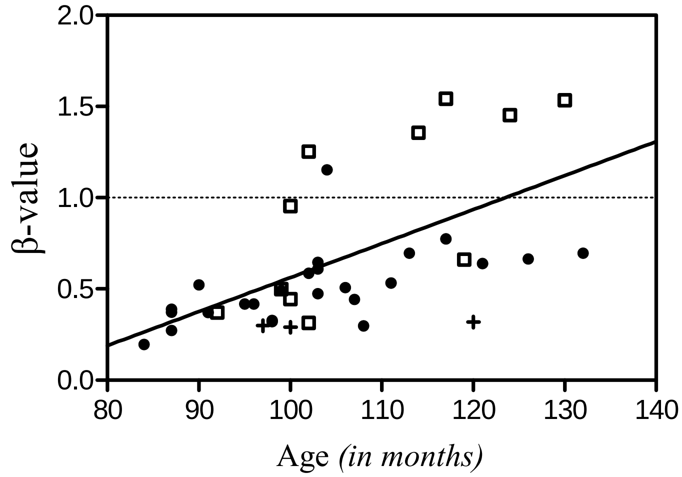
Note. The 7- and 8-year-olds' median estimate for the number 52 on the 0-1000 number line (represented by an open circle) was a statistical outlier and was excluded from the corresponding analyses. Nearly half of the children produced estimates that were much too high, tending to place "52" near the position for "500".

Figure 4.



Note. Median estimates represented as open circles (corresponding to 5652 and 10870 for 8-year-olds' performance, 5652 for 10-year-olds' performance, and 5652 for the 8-10 year-old group performance on the 0-100000 number line) were statistical outliers and excluded from analyses. Again, these outliers result from many children misinterpreting the decimal value – placing, for example, the number 5652 closer to 500000 than 5000.

Figure 5.



+ Unbounded Version *(n=3)*

□ One-Cycle Version *(n=12)*

• Two-Cycle Version *(n=25)*

Appendix A

Complete List of Numbers Presented in Test Trials for Each Experiment

0-20 Number Line (Experiment 1): 2, 3, 4, 6, 8, 12, 14, 16, 17, and 18

0-100 Number Line (Experiments 1 & 2): 3, 6, 8, 12, 14, 17, 21, 24, 29, 33, 39, 42, 48, 52, 58, 61, 67, 71, 76, 79, 83, 86, 88, 92, 94, and 97.

0-1000 Number Line (Experiment 2): 3, 7, 19, 52, 103, 158, 240, 297, 346, 391, 438, 475, 525, 562, 609, 654, 703, 760, 842, 897, 948, 981, 993, and 997

0-1000 Number Line (Experiment 3): 8, 15, 25, 56, 109, 154, 237, 290, 338, 388, 430, 467, 517, 560, 599, 650, 696, 761, 839, 889, 939, 980, 989, and 993

0-100000 Number Line (Experiment 3): 870, 1522, 2609, 5652, 10870, 15435, 23696, 29022, 33805, 38478, 43043, 46739, 51739, 56087, 60000, 65217, 69783, 76087, 83913, 88913, 93913, 98043, 98913, and 99348.

Appendix B

Estimates of Relative Support for Each Model, Experiment 1

		0-20 Number Line			0-100 Number Line		
		R^2	AICc	Δ AICc	R^2	AICc	Δ AICc
Log-to-Linear Shift Models	Logarithmic Model	0.776	23.678	24.269	0.844	111.982	32.795
	Linear Model	**0.980***	**-0.591***	---	0.944	85.451	6.263
Proportion Judgment Models	Unbounded Version	0.897	13.969	14.559	**0.952***	**79.187***	---
	One-Cycle Version	0.931	9.851	10.441	0.926	90.421	11.234
	Two-Cycle Version	0.935	9.295	9.885	0.765	120.583	41.396

* Indicates the preferred model (i.e., the model yielding the lowest AICc value and lowest LOOCV error index).

Notes: Δ AICc refers to the difference in AICc values compared to the preferred model. As a guide to interpreting these results, we include the benchmarks proposed by Burnham and Anderson (2002) below:

> As a rough rule of thumb, models having a Δ within 1-2 of the [preferred] model have substantial support and should receive considerations in making inferences. Models having Δ within about 4-7 of the [preferred] model have considerably less support, while models with Δ > 10 have either essentially no support and might be omitted from further consideration or at least fail to explain some substantial structural variation in the data. (pp. 446).

Appendix C

Estimates of Relative Support for Each Model, Experiment 2

		0-100 Number Line			0-1000 Number Line		
		R^2	AICc	Δ AICc	R^2	AICc	Δ AICc
Log-to-Linear Shift Models	Logarithmic Model	0.644	150.790	95.797	0.566	256.732	89.182
	Linear Model	0.986	66.007	11.013	0.970	192.574	25.024
Proportion Judgment Models	Unbounded Version	0.817	131.563	76.569	0.704	245.563	78.013
	One-Cycle Version	0.984	68.640	13.647	0.957	199.459	31.909
	Two-Cycle Version	**0.990***	**54.994***	---	**0.989***	**167.550***	---

Appendix D

Estimates of Relative Support for Each Model, Experiment 3

		0-1000 Number Line			**0-100000 Number Line**		
		R^2	AICc	Δ AICc	R^2	AICc	Δ AICc
Log-to-Linear Shift Models	Logarithmic Model	0.368	233.711	84.626	0.277	450.321	91.916
	Linear Model	0.981	159.843	10.758	0.966	383.217	24.812
Proportion Judgment Models	Unbounded Version	0.552	224.489	75.404	0.417	443.577	85.173
	One-Cycle Version	0.975	163.903	14.818	0.962	383.368	24.963
	Two-Cycle Version	**0.988***	**149.085***	---	**0.988***	**358.405***	---