Developmental Psychology
2017, Vol. 53, No. 6, 1088–1097
© 2017 American Psychological Association
0012-1649/17/$12.00 http://dx.doi.org/10.1037/dev0000305

A Unified Framework for Bounded and Unbounded Numerical Estimation

Dan Kim and John E. Opfer
The Ohio State University

Representations of numerical value have been assessed by using bounded (e.g., 0–1,000) and unbounded (e.g., 0-?) number-line tasks, with considerable debate regarding whether 1 or both tasks elicit unique cognitive strategies (e.g., addition or subtraction) and require unique cognitive models. To test this, we examined how well a mixed log-linear model accounted for 86 5- to 9-year-olds' estimates on bounded and unbounded number-line tasks and how well it predicted mathematical performance. Compared with mixtures of 4 alternative models, the mixed log-linear model better predicted 76% of individual children's estimates on bounded number lines and 100% of children's estimates on unbounded number lines. Furthermore, the distribution of estimates was fit better by a Bayesian log-linear model than by a Bayesian distributional model that depicted estimates as being anchored to varying number of reference points. Finally, estimates were generally more logarithmic on unbounded than bounded number lines, but logarithmicity scores on both tasks predicted addition and subtraction skills, whereas model parameters of alternative models failed to do so. Results suggest that the logarithmic-to-linear shift theory provides a simple, unified framework for numerical estimation with high descriptive adequacy and yields uniquely accurate predictions for children's early math proficiency.

Keywords: cognitive development, numerical cognition, number-line estimation, psychophysical function

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In this article, we sought to resolve a debate on what gives rise to developmental changes in numerical estimation and to provide a unified framework for seemingly irreconcilable data regarding the psychophysical functions that link numbers to their magnitude estimates (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Opfer, Siegler, & Young, 2011; Opfer, Thompson, & Kim, 2016; Rouder & Geary, 2014; Siegler & Opfer, 2003; Slusser, Santiago, & Barth, 2013).

Conventionally, developmental changes in numerical estimation have been viewed as following a logarithmic-to-linear shift in representations of numeric magnitude (Dehaene, Izard, Spelke, & Pica, 2008; Opfer & Siegler, 2007; Siegler & Opfer, 2003; Siegler, Thompson, & Opfer, 2009). This shift was first observed on a "bounded" number-line task, in which a target number was estimated on a line flanked by a number at each end (Figure 1A). On this task, young children's placement of numbers typically follows an approximately logarithmic function, but this logarithmic pattern changes to a linear one with age and experience, and the timing of the shift depends on the number ranges tested (Booth & Siegler, 2006; Laski & Siegler, 2007; Opfer & Martens, 2012; Opfer & Siegler, 2007; Opfer, Siegler, & Young, 2011; Opfer & Thompson, 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003; Siegler et al., 2009; Thompson & Opfer, 2008).

Because these logarithmic-to-linear shifts occur at different times in development for different number ranges, logarithmic and linear representations apparently coexist and compete in the same child. For example, on a 0–100 number line, most kindergarteners produce logarithmic estimates, whereas most second graders produce linear ones; on a 0–1,000 number line, most second graders' estimates are logarithmic and most sixth graders' estimates are linear; on 0–10,000 and higher scales, the proportion of oldest children and adults producing linear estimates declines with the magnitude of the scale (Booth & Siegler, 2006; Landy, Silbert, & Goldin, 2013; Siegler & Opfer, 2003; Thompson & Opfer, 2010). Thus, a commonly used model of numerical estimation is a mixed log-linear model (MLLM), in which estimates are predicted as a weighted sum of logarithmic and linear transforms of the number to be estimated (Anobile, Cicchini, & Burr, 2012; Cicchini, Anobile, & Burr, 2014; Dehaene et al., 2008; Karolis, Iuculano, & Butterworth, 2011; Opfer et al., 2016).

Proportional Reasoning Account of "Bounded" Numerical Estimation

Recently, two related challenges to the logarithmic-to-linear shift theory have been raised. The first challenge argued that children's estimates on bounded number lines (e.g., 0–1,000) reflect one of three proportional reasoning strategies (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Slusser et al., 2013), which had been modeled in adults using three different cyclic power models (CPMs; Hollands & Dyre, 2000). By comparing multiple extensions of the CPM with a log or linear model, Slusser et al. (2013) showed that a two-cycle power model provided better fits to 7- to 10-year-olds'

This article was published Online First April 27, 2017.
Dan Kim and John E. Opfer, Department of Psychology, The Ohio State University.
Partial support was provided by IES grant R305A160295. We are grateful to the schools, parents, and children for their support and contribution, and to Sungkyun Cho, Ariel Lindner, Marina Peeva, and Yiwan Wang for their assistance in collecting data.
Correspondence concerning this article should be addressed to Dan Kim, Department of Psychology, Ohio State University, 225 Psychology Building, 1835 Neil Avenue, Columbus, OH 43210. E-mail: kim.3839@osu.edu

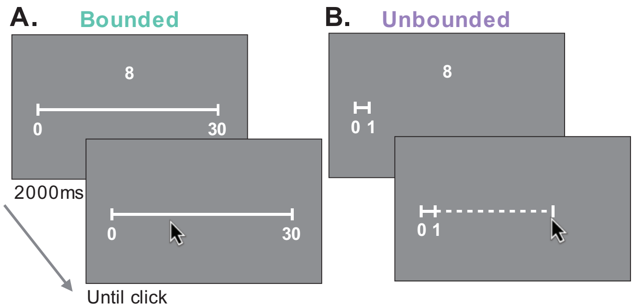

Figure 1. Illustration of the bounded and unbounded number-line task. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

estimates of familiar and unfamiliar numbers, suggesting children use the ends and middle of the number line as reference points. On the other hand, 5- to 6-year-olds who made linear estimates on familiar number scales (0–20) produced estimates of unfamiliar numbers (0–100) that were better fit by a zero-cycle power model, implying the use of the lower end as the only reference point. Based on these findings, Slusser, Santiago, and Barth (2013) argued that logarithmic patterns in young children's estimates did not reflect logarithmic representations of numbers, but arose from their lack of skills to use reference points as anchors and reason about numbers proportionally.

A somewhat similar account was proposed by Rouder and Geary (2014). Within this account, variance in children's estimates is thought to vary with the number-to-be-estimated because children estimate the position of a number by anchoring their responses against reference points (e.g., 0, 50, and 100 on a 0–100 number line). Although this idea had been tested informally in adult subjects (Cohen & Blanc-Goldhammer, 2011), the tests were purely descriptive and failed to simulate the compressive or cyclic patterns in estimates by falsely assuming an equal variance across given numerals. To address this issue, Rouder and Geary used hierarchical, Bayesian modeling to assess the descriptive adequacy of three distributional models that implemented both the unequal-variance assumption and use of reference points. Fitting the distributional anchor-based models to 5-year longitudinal data, they showed a developmental increase in the number of anchors that were used in estimation: Estimates in most first graders were best predicted by a two-anchor model, whereas a three-anchor model explained data of older children better than the other two models. Also, a one-anchor model that is computationally the most similar to a log model provided the worst fitting to estimates for all grades. Taken together, the authors reached the conclusion that the developmental shifts in number-line estimation resulted from the improvement in the use of reference points.

In defense of the log-to-linear account, however, a recent study by Opfer and colleagues (2016) showed that strong fits of the CPMs to number-line data was purely an artifact of an unusual anchoring procedure used by both Slusser et al. (2013) and Rouder and Geary (2014). Specifically, unlike the standard "free" numerical estimation procedure, these investigators had used an "anchoring" procedure in which they pointed to the middle of the number line and told children the number that would go there. Unsurprisingly, children's estimates near this point mirrored the information they were given. As a result, the superior fit of the cyclic models over a logarithmic one obtained only for anchored estimates rather than free estimates (Table 1). Moreover, Opfer et al. (2016) found that *both* free and anchored estimates were better fit by MLLM than a mixed cyclic power model (MCPM) that included all variants of the CPMs proposed. Within this MLLM, the effect of anchoring could be traced to simply decreasing the logarithmicity of estimates. A major question left outstanding by Opfer et al. (2016), however, is how well the MLLM model compares with Rouder and Geary's (2014) anchor-based distributional models.

Measurement Skills Approach to Numerical Estimation

A second challenge to the logarithmic-to-linear shift account comes from the measurement-skills account proposed by Cohen and colleagues (2011, 2014). Within this account, log-like patterns in estimates are proposed to arise because conventional number-line tasks are "bounded" by two numbers, thereby requiring subtraction skills that young children may not have acquired (Figure 1A). For example, when a target number such as 70 is given on a 0–100 number line, a participant may need to consider both the number 70 and the difference 30 after subtracting 70 from 100. In contrast, estimation of numbers on a number line "unbounded" by the larger endpoint is thought to require only simple addition (Figure 1B is thought to more accurately reveal numeric magnitude judgments and is thought to be best modeled using three more variants of power models called "scallop power models" (SPMs; Cohen & Blanc-Goldhammer, 2011).

To test this second challenge, Cohen and Sarnecka (2014) examined the effects of boundedness in number-line tasks by giving both bounded and unbounded tasks to 3- to 8-year-olds. Model fits of a log or linear function were compared with those of multiple variations of the CPM for the bounded estimates and to those of SPMs for the unbounded estimates. In the bounded condition, one of the extensions of cyclic models—the subtraction bias cyclic model—provided better fits to estimates, and the response bias parameters (βs) changed with age. For the unbounded number-line estimates, in contrast, one of three SPMs predicted the estimates better, but the estimation bias parameters (βs) stayed constant across age groups. These findings led the researchers to conclude that log-to-linear shifts in numerical representation do not exist, but these changes instead stemmed from the poor use of advanced arithmetic strategies, such as subtraction, in young children. Also, with the finding that response bias parameters (βs) in the unbounded condition did not change with age, they argued that young children could scale numbers as accurately as older children and adults in unbounded number lines, requiring simple addition skills only. They further suggested that numerical estimation be made on an unbounded number line to rule out the influence of lacking arithmetic and measurement skills and for a better assessment of number representations.

The Current Study

In this article, we tested three hypotheses regarding bounded and unbounded numerical estimation. The primary hypothesis is that the cognitive process of estimation in the two tasks is essentially similar and does not require positing seven estimation strategies associated with 10 unique psychophysical functions. From this

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
Differences in Methods and Findings Among Studies Examining Number-Line Estimation

Study	Participants	Number Scale	Models compared	Best-fitting model for majority
Bounded free				
Siegler and Opfer (2003)	Age 7 to 11, and adults	0–100, 0–1,000	Log, linear	Log for younger children in 0–1,000, linear for young children in 0–100 and older children and adults on both
Siegler and Booth (2004)	Age 5–8	0–100	Log, linear, exponential	Log for younger and linear for older children
Laski and Siegler (2007)	Age 5–8	0–100	Log, linear	Log for younger and linear for older children on the pretest, more linear on the posttest after training with discrepant feedback
Opfer and Siegler (2007)	Age 8	0–1,000	Log, linear	Log before training, linear after training with discrepant feedback
Opfer and Thompson (2008)	Age 7	0–1,000	Log, linear	Log before training, linear after training with discrepant feedback
Thompson and Opfer (2008)	Age 7–9	0–1,000	Log, linear	Log for younger and linear for older children on the pretest, more linear on the posttest after training
Opfer et al. (2011)	Kindergarten through fourth graders	0–100 for kindergarten through second grade, 0–1,000 for first through fourth	Log, linear vs. 1CPM, 2CPM	Log for younger children on both scales, linear for older children on both scales
Cohen and Blanc-Goldhammer (2011; P)	Adults	0–26	Linear, 1CPM, 2CPM	One of CPMs
Opfer and Martens (2012)	Age 6–17, and adults for WS, and age 6–11, and adults for TD	0–1,000	Log, linear	Log for WS children and adults (even after training) and TD younger children, linear for TD older children and adults
Cohen and Sarnecka (2014; P)	Age 3–8	0–20	Log, linear, SBCM, 1CPM, 2CPM	One of CPMs
Opfer et al. (2016)	Age 6–8	0–1,000	Log vs. 2CPM, MLLM vs. MCPM	Log, MLLM
Anchored				
Siegler and Booth (2004)	Age 5–8	0–100	Log, linear, exponential	Log for younger and linear for older children
Booth and Siegler (2006)	Age 5–10	0–100 for age 5–9, 0–1,000 for age 8 and 10	Log, linear, exponential	Log for younger and linear for older children in both 0–100 and 0–1000
Barth and Paladino (2011)	Age 5 and 7	0–100	Log, linear, 1CPM, 2CPM	Log and one of CPMs for age 5, one of CPMs for age 7
Slusser et al. (2013)	Age 5–10	0–20 and 0–100 for age 5–6, 0–100 and 0–1,000 for age 7–8, 0–1,000 and 0–100,000 for age 8–10	Log, linear, 0CPM, 1CPM, 2CPM	One of CPMs
Rouder and Geary (2014)	Age 6–11 (longitudinal)	0–100	1- to 3-anchor models	2-anchor model for younger children (1st grade), 3-anchor model for older children (2nd-5th grade)
Opfer et al. (2016)	Age 6–8	0–1,000	Log vs. 2CPM, MLLM vs. MCPM	2CPM, MLLM
Unbounded free				
Cohen and Blanc-Goldhammer (2011; P)	Adults	0–1 (numbers <25)	Linear, 1SPM, 2SPM, multiple SPM	One of SPMs
Cohen and Sarnecka (2014; P)	Age 3–8	0–1 (numbers <20)	Log, linear, 1SPM, 2SPM, multiple SPM	One of SPMs
Anchored				
None				

Note. CPM = cyclic power model; WS = Williams syndrome group; TD = typically developing group; SBCM = subtraction bias cyclic model; MLLM = mixed log-linear model; MCPM = mixed cyclic power model. (P) indicates that the study included practice trials, in which numbers used in practice are not stated. Thus, it is unclear whether the midpoint was used for practice in the studies although they are classified as *free* estimation.

perspective, bounded and unbounded numerical estimates—like free and anchored numerical estimates—are best viewed as reflecting children's representations of numeric magnitude, best modeled using the same MLLM, and that individuals with the most logarithmic performance on one task will have the most logarithmic performance on the other.

The second hypothesis is that both tasks are equally well suited for characterizing numerical magnitude representations, and both tasks will be associated with use of numbers in other contexts, such as addition and subtraction. If this view is correct and the log-linear model is the best model, then we would expect that the logarithmicity parameter of the MLLM will be more strongly associated with addition and subtraction than *even the model parameter of the subtraction bias cyclic model (SBCM) that allegedly tracks subtraction skills*.

Finally, the third hypothesis is that because number-line estimation tasks do assess children's representations of numeric magnitude, providing children with more numbers against which to anchor their estimates will result in improved performance, leading to bounded numerical estimates being more accurate and more linear than unbounded estimates. This prediction is directly at odds with the measurement-skills hypothesis and is the simplest to address empirically.

To test these three hypotheses, we nearly replicated the procedure used by Cohen and Sarnecka (2014) with only a couple of exceptions, such as the use of a fully balanced design that could detect order effects and the administration of a battery of math tests (addition and subtraction). Then, we pitted the MLLM tested by Opfer et al. (2016) against the MCPMs, the MSPM, and the three distributional anchor-based models. The MCPMs and MSPM were compared with the MLLM using ordinary least squares (OLS) analyses. To take into account variance in estimates across experimental conditions and participants, the MLLM and the distributional models were compared in Bayesian contexts as done in Rouder and Geary (2014).

The MLLM consists of logarithmic and linear components and is defined as:

$$y = a\left(\lambda \frac{U}{\ln(U)} \ln(x) + (1 - \lambda)x\right), \quad (1)$$

in which y indicates an estimate of number x on a 0-U number line. Also, a denotes a scaling parameter, and λ is a logarithmicity index that measures the degrees of logarithmic compression in estimates. If estimation is perfectly linear, a λ value converges to 0, whereas the value of the logarithmicity index gets close to 1 as estimation shows more logarithmic compression.

The MCPMs were formulated as proposed by Hollands and Dyre (2000; also see Opfer, Thompson, & Kim, 2016, for details). The first MCPM (MCPM1) is formalized based on Slusser et al.'s (2013) study that hypothesized that the number of reference points used for estimation changes in development: Children with poor proportion skills would only use a single reference point, that is, the lower bound (zero-cyclic power model), and then learn to use the lower and upper bounds (one-cycle power model), and the middle point with the two endpoints (two-cycle power model) as they become more familiar with the number range. The MCPM1 is defined as:

$$y = w_1 \cdot 0CPM + w_2 \cdot 1CPM + w_3 \cdot 2CPM, \quad (2)$$

where each of w_1, w_2, and w_3 denotes a weight for each variant of the CPM, respectively. Each weight and the sum of weights are constrained to be between 0 and 1, so that contribution of three models in a response can be assessed individually.

The MCPM2 is identical to the MCPM1 except that 0CPM is replaced with SBCM in the MCPM2 as proposed by Cohen and Sarnecka (2014). The SBCM was similar to 1CPM, but includes an additional parameter (s) that is associated with the subtraction bias. The following is the equation of the MCPM2:

$$y = w_1 \cdot SBCM + w_2 \cdot 1CPM + w_3 \cdot 2CPM. \quad (3)$$

The MSPM for unbounded number-line tasks is also formed in the same manner based on Cohen and Blanc-Goldhammer's (2011) assumption that the unbounded tasks are solved using an addition strategy. The following is the formalization of the mixed model:

$$y = w_1 \cdot 1SPM + w_2 \cdot 2SPM + w_3 \cdot m\mu ltiple\ SPM. \quad (4)$$

In the model, 1SPM indicates the single scallop model, 2SPM the dual scallop model, and multiple SPM the multiple scallop model. The same constraints set on weights in the MCPMs are set in this model.

In Bayesian modeling, a hierarchical MLLM and three distributional models were compared. For better model convergence, the logarithmicity component (λ) of the MLLM is logit transformed. The modified Bayesian MLLM is defined as:

$$y_{ijk} = a_j\left(\frac{U_r \cdot \ln(x_{ijk}) \cdot \exp(\mathrm{log}it(\lambda_j)) + x_{ijk} \cdot \ln(U_r)}{\ln(U_r) \cdot (\exp(\mathrm{log}it(\lambda_j)) + 1)}\right) + e_{ijk}, \quad (4)$$

where the transformed λ is defined as follows:

$$\mathrm{log}it(\lambda_j) = \ln\left(\frac{\lambda_j}{1 - \lambda_j}\right). \quad (5)$$

In the model, e denotes error that is normally distributed with zero mean ($e_{ijk} \sim$ Normal(0, σ_j)), while i indicates the number of trials in bounded ($k = 1$) and unbounded condition ($k = 2$) for j participants ($j = 1, 2, \ldots 86$) and r the number ranges ($r = 1, 2, 3$). The model is hierarchically extended using the priors as follows:

$$a_j \sim Normal(\mu_a, \sigma_a), \quad (6)$$
$$\mathrm{log}it(\lambda_j) \sim Normal(\mu_{\mathrm{log}it(\lambda)}, \sigma_{\mathrm{log}it(\lambda)}).$$
$$\sigma_j \sim Cauchy(0, 10)$$

The Gaussian distribution was used for the priors of μ_a and $\mu_{logit(\lambda)}$, whereas σ_a and $\sigma_{logit(\lambda)}$ had the Cauchy priors. Although the parameter λ is transformed to constrain λ to be between 0 and 1 more efficiently in Bayesian analyses, the Bayesian MLLM (Eq. 4) is mathematically identical to the conventional MLLM (Eq. 1). In the current study, fitting of the MLLM was pitted against that of each of the hierarchical models by Rouder and Geary (2014). The same hierarchical priors used in the MLLM were applied to the anchor-based models: the Gaussian priors for μ_α and μ_β and the Cauchy priors for σ_α and σ_β.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Experiment

Method

Subject. Thirty 5- to 6-year-old kindergarten, 30 first-grade, and 26 second-grade students were recruited in Columbus, OH (kindergarteners: 17 female children, M = 5.90 years, SD = .32 years; first graders: 19 female children, M = 6.74 years, SD = .39 years; second graders: 21 female children, M = 7.91 years, SD = .46 years).

Materials and procedure. Participants completed both bounded and unbounded number-line tasks given in a counterbalanced order. In the bounded condition, a number was shown for 2,000 ms above a number line flanked by 0 and 30/100/1,000 (Figure 1A). On every trial, the mouse cursor was reset to be located at the 0 point and moved only horizontally on a number line. Participants were instructed to estimate a given number on a number line with the following instruction:

> Now we're going to play a game with numbers. This is a number line. In this game, each number line will have a 0 at one end and 30/100/1,000 at the other end. There will be a number up here. Your job is to show me where that number goes on a number line like this one. When you decide where the number goes, you have to drag this little mark to where the number should go. When you're ready to go again, press the green bar (spacebar with a green sticker on) on the keyboard.

The unbounded number-line task was identical to the bounded one except that a single-unit line (0–1) was presented instead of the full range number line (Figure 1B). An experimenter introduced the task with the following instruction that was created based on Cohen and Sarnecka (2014):

> Now we're going to play a game with numbers. This is a number line. In this game, each number line will have a 0 here and 1 over here. All the other numbers go after 1. There will be a number up here. Your job is to show me where that number goes on a number line like this one. When you decide where the number goes, you have to drag this little mark to where the number should go. When you're ready to go again, press the green bar on the keyboard.

The length between 0 and 1 was adjusted based on the ranges of number lines. In other words, a displayed length for 0–1 intervals was the shortest for a 0–1,000 number line, whereas it was the longest in the 0–30 number line task.

Based on Slusser et al. (2013), different number ranges that would elicit compressive estimates were used depending on children's grades. Thus, 5- and 6-year-old kindergarteners were given 0–30 number lines, with to-be-estimated numbers sampled evenly from a 0–30 range: 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 19, 20, 21, 22, 23, 27, 28, 29, and 30. First graders were given 0–100 number lines, with to-be-estimated numbers sampled evenly from 0 and 100: 3, 4, 6, 8, 12, 17, 21, 23, 25, 29, 33, 39, 43, 48, 52, 57, 61, 64, 72, 79, 81, 84, 90, 96, and 100. Second graders were given 0–1,000 number lines, with to-be-estimated numbers sampled from 0–1,000: 2, 5, 18, 34, 56, 78, 100, 122, 147, 150, 163, 179, 246, 366, 486, 606, 722, 725, 738, 754, 818, 938, and 1,000. In both bounded and unbounded conditions, the same target numbers were randomly presented. To keep children's attention on the tasks but to avoid visual distraction over the course of estimation, a neutral sound was produced once a stimulus was displayed and once a response was made. This procedure differed trivially from the use of visual reinforcements in Cohen and Sarnecka (2014). The tasks started after an instruction without any practice, and there was no feedback provided over trials. Unlike Cohen and Sarnecka (2014), number lines stayed the same and did not vary in locations or lengths over trials. This difference in procedure was to give children the best possible chance at performing better on the unbounded than bounded task.

Upon completion of the two number-line tasks, arithmetic performance was assessed with paper-and-pencil addition and subtraction tests based on Geary, Bow-Thomas, Liu, and Siegler (1996). The addition test consisted of 50 one-digit addend problems. The subtraction items were generated by rearrangement: first addends in addition problems were used as subtrahends, while the second addends became the answers. Participants were given tests in random order and asked to solve as many problems as possible within 1 min for each test.

Results

Our results are divided into two sections, "Conventional Model Selection Techniques" and "Bayesian Extensions." In the first section, we examine the descriptive adequacy and predictive power of the MLLM against the mixed cyclic power and mixed SPMs. Here we rely solely on conventional OLS methods. In the next section, we adapt a Bayesian version of the MLLM to compare with a Bayesian version of the CPM approach employed by Rouder and Geary (2014).

Conventional Model Selection Techniques

To test whether the cognitive process of estimation differed for the bounded and unbounded tasks, children's estimates were examined using combinations of the psychophysical functions proposed by Cohen and Sarnecka (2014) and by Opfer, Thompson, and Kim (2016). For the bounded condition, individual data were fit by the MLLM, MCPM1, and MCPM2, whereas estimates on unbounded number lines were fit with the MLLM and MSPM with the OLS methods. The model fits were then assessed by comparing corrected Akaike information criterion (AICc) values (Table 2).

Against the proportional reasoning and measurement-skills accounts, none of the MCPMs was the best fitting model for bounded number-line estimation. Overall, 76% of individual children's estimates were better fit by the MLLM than alternative models. Also against the measurement-skills account of unbounded number-line estimation, 100% of children produced estimates that were better described by the MLLM than the MSPM. These results suggest that logarithmically distorted estimates in a number-line task do not stem from a lack of proportion, subtraction, or addition skills, but come from representations of numerical magnitude that are well predicted with the MLLM.

Would variants of the power models based on arithmetic strategies predict children's actual arithmetic performance? To address this question, we next correlated each child's addition and subtraction performance with the best-fitting parameter values from the models. If response patterns in the bounded condition came from poor subtraction skills, parameters, such as βs of the MCPM1 & MCPM2 and s of MCPM2, would be expected to correlate with arithmetic scores. As presented in Table 3, however, none of the βs

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Mean and SD (in Parentheses) of Akaike information criterion (AICc) Values From Individuals' Estimates and Percentage of Individuals Best Fit by Each Model

	AICc			% Children best fit by each model		
Variable	0–30	0–100	0–1,000	0–30	0–100	0–1,000
Bounded						
MLLM	69.81 (14.87)	132.30 (21.03)	218.77 (19.44)	96.67	80	50
MCPM1	81.30 (14.99)	138.46 (24.62)	239.46 (23.27)	0	10	3.85
MCPM2	79.88 (17.48)	140.24 (24.07)	220.43 (19.33)	3.33	10	46.15
Unbounded						
MLLM	70.35 (19.89)	139.60 (20.30)	236.11 (20.19)	100	100	100
MSPM	94.01 (18.67)	167.83 (23.25)	280.92 (25.71)	0	0	0

Note. MLLM = mixed log-linear model; MCPM = mixed cyclic power model.

from the MCPMs was a reliable predictor for either addition or subtraction scores: only β of MCPM1 correlated with subtraction scores, but such correlation was not found in the subtraction bias parameter *s* of MCPM2. This correlation also remained insignificant if adjusted parameter values (absolute values of βs − 1 or *s* − 1) were used for analyses. In contrast, the logarithmicity parameter λ of the MLLM reliably predicted *both* addition and subtraction performance, $r(84) = -.42$, $p < .001$ for addition; $r(84) = -.36$, $p < .01$ for subtraction. These findings support the idea that logarithmic estimates on a number line should be viewed as a reflection of logarithmic representation of numbers, which interferes with use of numbers in other contexts.

We also examined relations between the parameter βs of the MSPM and arithmetic achievement in the unbounded tasks. According to the measurement-skills account, these parameters in the scallop models should reflect representation of numeric magnitude most accurately of all because the task is supposedly pure of the sins of boundedness and the model is correct. Against this idea, however, none of the parameter βs were significantly correlated with arithmetic performance, whereas the MLLM parameter λ again showed a significant correlation with addition and subtraction, $r(84) = -.33$, $p < .01$ for addition; $r(84) = -.34$, $p < .01$ for subtraction (Table 3). These findings again support the idea that logarithmic estimates on a number line should be viewed as a reflection of logarithmic representation of numbers, which interferes with use of numbers in other contexts.

The final issue we examined was to quantify the effect of boundedness on accuracy of numerical estimates. If the bounded number-line tasks required more challenging measurement skills, the bounded tasks would be expected to elicit more erroneous and compressive estimates. On the other hand, if the two endpoints provide better anchors against which to judge large numbers, estimates in the bounded condition would be expected to yield less logarithmic and more accurate estimates. Consistent with the latter account, on all the 0–30/100/1,000 number-line tasks, averaged percent absolute errors (PAEs) in median estimates were always greater in the unbounded than bounded conditions (bounded M = 12%, SD = 5%, unbounded M = 17%, SD = 7% for 0–30 range; bounded M = 12%, SD = 4%, unbounded M = 14%, SD = 7%, for 0–100 range; bounded M = 20%, SD = 12%, unbounded M = 24%, SD = 14%, for 0–1,000 range).

Also, in all three number ranges, the values of the logarithmicity parameter (λ) were greater for median estimates in the unbounded condition than those in the bounded condition (Figure 2; bounded λ = .30, unbounded λ = .76 for 0–30 range; bounded λ = .42, unbounded λ = .55 for 0–100 range; bounded λ = .73, unbounded λ = .75 for 0–1,000 range). These results do not support the measurement-skills prediction that unbounded number-line tasks are easier and require less advanced mensuration skills than conventional bounded tasks.

Finally, the same analyses were repeated on individual children's data. For the 0–30 number lines, greater PAEs were again observed in the unbounded than bounded condition (M = 17%, SD = 7% for bounded; M = 21%, SD = 7% for unbounded), $t(29) = 3.82$, $p < .001$, $d = .69$. For the 0–100 number lines, greater PAEs were also observed in the unbounded than bounded condition (M = 15%, SD = 7% for bounded; M = 19%, SD = 9% for unbounded), $t(29) = 3.51$, $p < .01$, $d = .64$. For 0–1,000 number lines, greater PAEs were again observed in the unbounded

Table 3
Correlation Between Parameter Values and Arithmetic Performance After Number-Line Ranges Were Controlled For

Variable	Addition	Subtraction
Bounded		
MLLM		
λ	$-.42^{***}$	$-.36^{**}$
MCPM1		
β_{0CPM}	−.01	−.10
β_{1CPM}	.17	$.27^{*}$
β_{2CPM}	−.08	.10
MCPM2		
s	.13	.06
β_{SBCM}	.02	−.09
β_{1CPM}	.17	.14
β_{2CPM}	−.06	−.13
Unbounded		
MLLM		
λ	$-.33^{**}$	$-.34^{**}$
MSPM		
β_{1SPM}	−.05	.00
β_{2SPM}	.19	.01
β_{MSPM}	.16	.15

Note. MLLM = mixed log-linear model; MCPM = mixed cyclic power model.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

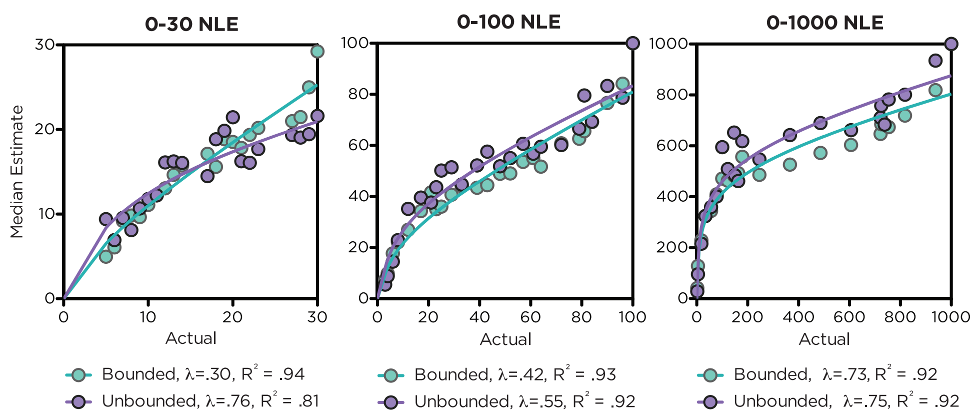

Figure 2. Median estimates in 0–30, 0–100, and 0–1,000 number lines with and without bounds. See the online article for the color version of this figure.

than bounded condition (M = 20%, SD = 8% for bounded; M = 27%, SD = 11% for unbounded), $t(25) = 5.21$, $p < .001$, d = 1.02. Therefore, for all the three number ranges, conventional number lines with two bounds elicited more accurate responses than unbounded number lines. These results are not at all consistent with the measurement-skills account, which depicts the bounded version of the task as more difficult, but the findings are in line with the previous findings about increased accuracy in anchored estimation (Opfer et al., 2016; Slusser et al., 2013).

Although number-line tasks without an upper bound produced more erroneous estimates, it is not clear whether the unboundedness increased random noise or logarithmic compression in estimates. To address this issue, individual children's degrees of logarithmicity were obtained with the MLLM fitting.

As shown in Figure 3, in all the three number ranges, the values of the logarithmicity measure were greater for a number line without bounds. In 0–30 and 0–100 number-line tasks, estimates in the unbounded condition showed significantly greater logarithmic compression in the 0–30 range (M = .51, SD = .38 for bounded; M = .76, SD = .31 for unbounded), $t(29) = 4.33$, $p <$ .001, d = .79, and in the 0–100 range (M = .51, SD = .31 for bounded; M = .65, SD = .30 for unbounded), $t(29) = 3.37$, $p <$ .01, d = .62. Although logarithmicity was slightly greater for the unbounded 0–1,000 number-line task than for the bounded one, the difference was not statistically significant (M = .70, SD = .32 for bounded; M = .74, SD = .32 for unbounded), $p > .05$. Against the arithmetic strategy hypothesis, negatively accelerating patterns in estimates were observed in both bounded and unbounded number-line tasks, and these patterns were stronger when the number line did not have a right-end bound. Even if the number-line task without the right endpoint revealed more accurate representation of numbers as the measurement-skills account claimed, what it revealed in the unbounded task was more logarithmic compression.

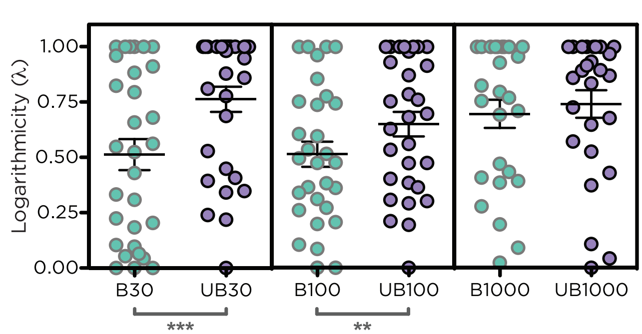

Figure 3. Individual logarithmicity (λ) across the three number-line tasks. ** $p < .01$. *** $p < .001$. See the online article for the color version of this figure.

Bayesian Extensions

A Bayesian framework for the proportionality account was suggested in previous work by Rouder and Geary (2014). We compared that anchor-based distributional framework here against a Bayesian extension of the MLLM.

To make analyses equivalent to previous work, we implemented hierarchical Bayesian modeling of the MLLM and the three psychophysical functions from Rouder and Geary (2014) using the package *rstan* (Stan Development Team, 2016b), which is one of the interfaces to Stan (Stan Development Team, 2016a). Given that a three anchor model captures scallop patterns that may be observed in unbounded estimates, and more importantly, estimates in the unbounded task did not show qualitatively distinct patterns from bounded estimates, we used the combined data of bounded and unbounded estimates to obtain more reliable modeling outcomes.

To examine model fits of the MLLM and three anchor-based models, model predictive accuracy was compared using approximate leave-one-out cross-validation (LOO; Vehtari, Gelman, & Gabry, 2016a) conducted via the R package *loo* (Vehtari, Gelman, & Gabry, 2016b). Next, like Rouder and Geary (2014), we scaled all numbers and estimates to 0–1. Then, to compare model fits on a common scale, we transformed predicted values to correspond to actual magnitudes.

We found that, contrary to the proportional reasoning account, the estimated predictive accuracy for the MLLM was greater than that of the 1- and 3-anchor models. Specifically, the LOO expected log pointwise prediction density ($elpd_{loo}$) was −18,061.23 for the MLLM, whereas −18,109.23 for the 1-anchor and −18,964.98 for the 3-anchor model. A significance test using standard errors in the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

differences yielded the result that the MLLM was significantly more predictive than the others. The remaining model comparison, the 2-anchor model versus MLLM, yielded LOO values that were nominally in favor of the 2-anchor model ($elpd_{loo} = -18011.16$), but the difference in the prediction accuracy between the 2-anchor model and the MLLM was not significant (95% CIs [−104.29, 4.16]). Even though LOO allows us to compare model predictability directly derived from the Bayesian models, caution should be taken in interpretations of LOO values and the use of the standard errors in model prediction differences (Vehtari et al., 2016a).

With a tie in the ability of the MLLM and 2-anchor model to predict numerical estimates, we wondered if both models accurately predict data for other tasks and if both models accurately predict the distribution of responses. Our reasoning was that if the 2-anchor model accounts for children's numerical estimation as well as the MLLM, the model parameter β should be associated with numerical performance in other contexts. To address this, we tested the relationship between parameters of the two competitive models (β vs. λ) and mathematics proficiency by conducting correlation analyses after controlling for the number-line ranges. βs of the 2-anchor model showed significant associations with both scores, $r(84) = .31$, $p < .01$ for addition and $r(84) = .24$, $p < .05$ for subtraction, but λs from the Bayesian MLLM were more strongly correlated with mathematics proficiency, $r(84) = -.38$, $p < .001$ for addition; $r(84) = -.36$, $p < .001$ for subtraction. Therefore, even though the Bayesian 2-anchor model somewhat correlated with performance in other numerical tasks, the relations were weaker and less reliable than with the Bayesian MLLM.

More critically, we reexamined the validity of the Bayesian 2-anchor model in terms of distribution of responses. According to Rouder and Geary (2014), estimates show greater variation as estimated magnitudes are more distant from reference points (e.g., 0 and 100 for a 0–100 number line). To test this, we calculated relative distance of estimated magnitudes to either lower or upper end that was closest. For example, relative distance for number 39 and 81 on 0–100 number lines were .39 and .19, respectively (39/100 – 0 = .39; 1 – 81/100 = .19). When regressed as a function of relative distance, the standard deviations of estimates did not increase with relative distance between given numbers and reference points ($p > .05$). Similar results were found across all conditions ($p > .05$) except for unbounded estimates on 0–1,000 number lines. However, the responses in the unbounded 0–1,000 scales became *less* variable with relative distance to an anchor ($b = -.22$, $p < .05$), as opposed to the increase in estimation variability as argued by the distributional model. Also, when we regressed individuals' PAEs against relative distance between a given number and an anchor, less than 14% of children presented significantly increasing patterns in estimation errors (13.95% of children for unbounded and 11.63% of children for bounded PAEs). Taken together, the findings are not supportive of the argument that children's estimates show log-like compression because of the poor use of the reference points.

Discussion

In this article, we attempted to address the nature of developmental changes in numerical estimation and to provide a unified framework for seemingly irreconcilable data regarding the psychophysical functions that link numbers to their magnitude estimates (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Opfer et al., 2016; Rouder & Geary, 2014; Siegler & Opfer, 2003; Slusser et al., 2013). Specifically, we wished to see whether the logarithmic-to-linear shift existed on both bounded and unbounded number lines.

Evidence for Changes in Representations Versus Proportional Reasoning

Like Opfer et al. (2016), we found that a single MLLM provided better fits to children's data than combinations of seven alternative psychophysical functions. In Opfer et al. (2016), the MLLM was tested against MCPM1, and the MLLM predicted the majority of children's estimates regardless of whether their estimates were anchored. In the current study, this issue was revisited (with the addition of the MCPM2 proposed by Cohen & Sarnecka, 2014), and we again found that the simple log-linear model explained children's estimates better than alternative models regardless of whether estimates were bounded or unbounded. These results are consistent with Opfer et al.'s (2016) conclusion that providing children with anchoring instructions increases the linearity of estimates, but are still well described by the log-linear framework.

A similar conclusion is warranted by our findings that a single MLLM model is preferred to an anchor-based distributional model. In Rouder and Geary (2014), children were given a practice trial, in which they were asked to estimate 50 on a 0–100 number line and then provided with the accurate location of the midpoint. As the authors noted, the practice on the midpoint could have elicited the better use of reference points, causing better fits of a 2- or 3-anchor model to a 1-anchor model (which is somewhat similar to the log-linear model). The present results suggest the authors' speculation was correct. When children are given no anchoring instructions, as in the present study, we find no signature of an anchor-based pattern of estimates, but we find the same MLLM performs as well as or better than every Bayesian version of the anchor-based distributional model.

Although the log-to-linear shift account provides better prediction for number-line estimation than models derived from the proportional reasoning account, our findings do not necessarily conflict with the idea that strategies can be involved in older children and adults producing *linear* estimates. This idea was first proposed by Siegler and Opfer (2003), who showed that adults' (but not young children's) estimates had the greatest variation as one moved away from the quartiles on the 0–1,000 number line. Another piece of evidence for the use of reference points can be found in the M-shaped patterns in adults' response errors (Ashcraft & Moore, 2012; Cohen & Blanc-Goldhammer, 2011; Link, Huber, Nuerk, & Moeller, 2014; though see Peeters, Degrande, Ebersbach, Verschaffel, & Luwel, 2016). What is interesting about these signatures of a landmark-based strategy, however, is that they first appear long after children are already estimating the positions of numbers linearly. For example, second graders estimate the position of numbers on a 0–100 number line linearly, yet with no evidence of adults' quartile-based strategy. These results suggest that strategies on the number-line task are not at all evident among logarithmic estimates and not necessary for linear ones. One possibility is that landmark-based strategies are deployed as a slow, safe, and sure route to accuracy—much like children's use of addition strategies on problems they already have the ability to retrieve from memory (Siegler, 1988).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Evidence for Changes in Representations Versus Measurement Skills

As Cohen and Sarnecka (2014) stated, "For those children whose data are best fit by the logarithmic function, we are skeptical that this reflects a logarithmically organized quantity representation. If it did, we should expect to see the same pattern on the unbounded task . . ." (p. 1,650). In the teeth of this skepticism, we found robust evidence for a logarithmically organized quantity representation on the unbounded task, as well as the bounded task.

Evidence came from multiple sources. First, the MLLM, which had previously unified discrepant data from free and anchored numerical estimation (Opfer et al., 2016), predicted data from bounded and unbounded tasks better than rival psychophysical models. In addition, for both tasks, the logarithmicity parameter was significantly greater than zero in each of the three age groups tested. This success of a single simple model over many competing complex models suggests that there may be a simple cognitive process lurking behind many varieties of numerical estimation.

Second, individual differences were stable across the two tasks: children whose estimates were most logarithmic in the bounded task were also most logarithmic patterns in the unbounded task, $r(84) = .73$, $p < .001$. This wouldn't be expected if the two tasks elicited radically different estimation strategies or if only one of the tasks provided an accurate picture of children's numeric magnitude judgments.

Third, although the two tasks differed in overall accuracy, it was not in the direction predicted by the measurement-skills account (Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014). Against the measurement-skills account arguing that unbounded number lines require easier arithmetic strategies, such as simple addition, the magnitude of errors and logarithmicity values of both median and individuals' estimates were always greater for the unbounded condition across all three number ranges. Therefore, even if the more difficult unbounded tasks provide a better picture of the internal mental number line, what they show is *stronger* evidence for logarithmic representation of numbers.

Finally, the parameters expected by the strategy- and skill-based accounts to track arithmetic biases unique to each task were not even weakly associated with actual arithmetic performance, calling into question the psychological meaning of these parameters (Table 3). In contrast, the logarithmicity component of both the OLS and Bayesian MLLM was reliably correlated with children's arithmetic proficiency regardless of whether bounded or unbounded number lines were used. These findings are wholly inconsistent with the idea that bounded number lines require greater arithmetic skills than unbounded ones.

Why might our results have differed so strongly from those of Cohen and colleagues (2011, 2014)? Two reasons—methodological and analytical—seem likely. First, unlike Cohen and Sarnecka (2014), we counterbalanced the order of bounded and unbounded tasks so that any effects of task order would not influence our overall results. Counterbalancing also made it possible to test the presence of order effects. And we found that the order of tasks sometimes influenced estimates. For example, on the 0–1,000 task, more erroneous estimates were produced in both tasks if the unbounded task was first presented, $F(1, 24) = 9.79$, $p < .01$, $\eta_p^2 = .29$, suggesting that the greater difficulty of the unbounded task resulted in fatigue or confusion for the second task. This is critical because it suggests that unbounded tasks are not intrinsically easier for children—quite the opposite.

Another important difference came from the analytic strategies employed. Cohen and Sarnecka (2014) tested the log or linear model against the SBCM *or* 1CPM *or* 2CPM model for bounded estimates, while comparing the linear model to the 1SPM *or* 2SPM *or* multiple SPM for unbounded estimates. Because the number of *or*'s corresponds to greater overall model complexity (i.e., greater degrees of freedom), a poorer fit for one of *two* family A models (or just an A model) than one of *three* family B models is ambiguous. Our own analytic strategy, in contrast, was to explicitly include all of the same family of models in the same equation (e.g., MLLM, MCPM1&2, and MSPM), and thereby include the correct number of degrees of freedom. This strategy allowed us to improve the validity of our model comparison. In addition, it yielded two unique insights: that individual differences in the logarithmicity of estimates were stable across tasks and that individual differences in logarithmicity of estimates correlated with math skills better than alternative models. These findings not only directly contradicted the predictions of the measurement-skills account, they could not even be tested using Cohen and Sarnecka (2014)'s analytic approach.

In conclusion, the present study finds that the log-to-linear shift theory provides a simple, unified framework for numerical estimation that is high in descriptive adequacy for both bounded and unbounded number lines and is strongly related to children's mathematics performance. An interesting question for future studies is whether the same framework provides a good account of magnitude estimation more broadly.

References

Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition, 122,* 454–459. http://dx.doi.org/10.1016/j.cognition.2011.11.006

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology, 111,* 246–267. http://dx.doi.org/10.1016/j.jecp.2011.08.005

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

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42,* 189–201. http://dx.doi.org/10.1037/0012-1649.41.6.189

Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 7867–7872. http://dx.doi.org/10.1073/pnas.1402785111

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review, 18,* 331–338. http://dx.doi.org/10.3758/s13423-011-0059-z

Cohen, D. J., & Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology, 50,* 1640–1652. http://dx.doi.org/10.1037/a0035901

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science, 320,* 1217–1220. http://dx.doi.org/10.1126/science.1156540

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Geary, D. C., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development, 67,* 2022–2044. http://dx.doi.org/10.2307/1131607

Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review, 107,* 500–524. http://dx.doi.org/10.1037/0033-295X.107.3.500

Karolis, V., Iuculano, T., & Butterworth, B. (2011). Mapping numerical magnitudes along the right lines: Differentiating between scale and bias. *Journal of Experimental Psychology: General, 140,* 693–706. http://dx.doi.org/10.1037/a0024255

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive Science, 37,* 775–799. http://dx.doi.org/10.1111/cogs.12028

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. http://dx.doi.org/10.1111/j.1467-8624.2007.01087.x

Link, T., Huber, S., Nuerk, H. C., & Moeller, K. (2014). Unbounding the mental number line-new evidence on children's spatial representation of numbers. *Frontiers in Psychology, 4,* 1021. http://dx.doi.org/10.3389/fpsyg.2013.01021

Opfer, J. E., & Martens, M. A. (2012). Learning without representational change: Development of numerical estimation in individuals with Williams syndrome. *Developmental Science, 15,* 863–875. http://dx.doi.org/10.1111/j.1467-7687.2012.01187.x

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55,* 169–195. http://dx.doi.org/10.1016/j.cogpsych.2006.09.002

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. http://dx.doi.org/10.1111/j.1467-7687.2011.01070.x

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. http://dx.doi.org/10.1111/j.1467-8624.2008.01158.x

Opfer, J. E., Thompson, C. A., & Kim, D. (2016). Free versus anchored numerical estimation: A unified approach. *Cognition, 149,* 11–17. http://dx.doi.org/10.1016/j.cognition.2015.11.015

Peeters, D., Degrande, T., Ebersbach, M., Verschaffel, L., & Luwel, K. (2016). Children's use of number line estimation strategies. *European Journal of Psychology of Education, 31,* 117–134. http://dx.doi.org/10.1007/s10212-015-0251-z

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science, 17,* 525–536. http://dx.doi.org/10.1111/desc.12166

Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development, 59,* 833–851. http://dx.doi.org/10.2307/1130252

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

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

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: the Official Journal of the International Mind, Brain, and Education Society, 3,* 143–150. http://dx.doi.org/10.1111/j.1751-228X.2009.01064.x

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

Stan Development Team. (2016a). *Stan modeling language user's guide and reference manual, Version 2.12.0*. Retrieved from http://mc-stan.org

Stan Development Team. (2016b). RStan: The R interface to Stan (Version 2.10.1) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/rstan/index.html

Thompson, C. A., & Opfer, J. E. (2008). Costs and benefits of representational change: Effects of context on age and sex differences in symbolic magnitude estimation. *Journal of Experimental Child Psychology, 101,* 20–51. http://dx.doi.org/10.1016/j.jecp.2008.02.003

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. http://dx.doi.org/10.1111/j.1467-8624.2010.01509.x

Vehtari, A., Gelman, A., & Gabry, J. (2016a). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. Advance online publication.

Vehtari, A., Gelman, A., & Gabry, J. (2016b). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 0.1.6. Retrieved from https://cran.r-project.org/web/packages/loo/index.html

Received April 22, 2016
Revision received October 22, 2016
Accepted December 22, 2016 ■