Citation: Kim, D., & Opfer, J. E. (2021). Dynamics Versus Development in Numerosity Estimation: A Computational Model Accurately Predicts a Developmental Reversal. *Cognitive Science*, *45*(10), e13049.

Dynamics vs Development in Numerosity Estimation: A Computational Model Accurately Predicts a Developmental Reversal

Dan Kim* and John E. Opfer

Department of Psychology, The Ohio State University

1835 Neil Avenue, Columbus, OH 43210, USA

* Corresponding author

E-mail: kim.3839@osu.edu

Abstract

Perceptual judgments result from a dynamic process, but little is known about the dynamics of number-line estimation. A recent study proposed a computational model that combined a model of trial-to-trial changes with a model for the internal scaling of discrete numbers. Here, we tested a surprising prediction of the model—a situation in which children's estimates of numerosity would be better than those of adults. Consistent with the model simulations, task contexts led to a clear developmental reversal: children made more adult-like, linear estimates when to-be-estimated numbers were descending over trials (i.e., backward condition), whereas adults became more like children with logarithmic estimates when numbers were ascending (i.e., forward condition). In addition, adults' estimates were subject to inter-trial differences regardless of stimulus order. In contrast, children were not able to use the trial-to-trial dynamics unless stimuli varied systematically, indicating the limited cognitive capacity for dynamic updates. Together, the model adequately predicts both developmental and trial-to-trial changes in number line tasks.

Keywords: numerical cognition, numerosity perception, cognitive development, dynamic models.

Dynamics vs Development in Numerosity Estimation: A Computational Model Accurately Predicts a Developmental Reversal

1. Introduction

Magnitude estimation is not a static process but changes dynamically from trial to trial. In the past half-century, numerous studies have shown that quantities encountered on previous trials can have substantial effects on the estimate of a current quantity, such as time, size, length, weight, and numerosity of a set (Gilden et al., 1995; Helson et al., 1954; Parducci, 1963, 1965; Petrov & Anderson, 2005; Petzschner et al., 2015). Recently, researchers began noticing dynamic effects in number-line estimation tasks (Cicchini et al., 2014; D. Kim & Opfer, 2018), which are used in such diverse fields as numerical cognition and education.

The underlying mechanisms of dynamic effects in number-line estimation are not fully understood. In one model, the internal scaling of number is linear, and compression results from dynamic updates (Cicchini et al., 2014). In another, the internal scaling of number is logarithmic, and decompression results from memory for previous trials (D. Kim & Opfer, 2018). Here we tested a novel and unexpected prediction of the latter model. Specifically, if trial-to-trial changes reflect participants' memory for previous numbers (and not the internal scaling of number), highly memorable sequences of numbers (e.g., testing numbers in an ascending or descending order) could result in children providing more linear numerical estimates than adults, thereby reversing the typical "log-to-linear shift" (Siegler et al., 2009).

1.1. Development in number-line estimation

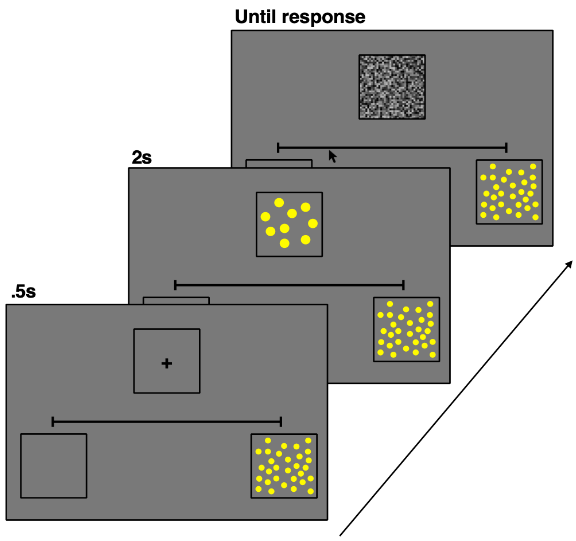

Figure 1. Illustration of a number-line estimation task.

In a number-line estimation task, participants are asked to place a numerical magnitude (e.g., 9 dots) on a line flanked by two numbers (e.g., 0 and 30 dots; Fig. 1). Previous literature has shown that numerical estimates increase logarithmically with actual magnitude in young children and more linearly in older children and adults (the “logarithmic-to-linear shift”; Berteletti et al., 2010; Booth & Siegler, 2006; Opfer et al., 2019; Sasanguie et al., 2012; Sella et al., 2015; Siegler et al., 2009; Siegler & Opfer, 2003). The log-to-linear shifts are still evident in various versions of number-line tasks that differ in anchoring instruction, boundedness, or number-line size (D. Kim & Opfer, 2017; D. Kim & Opfer, 2020; Lee et al., 2019; Opfer et al., 2016; Thompson & Opfer, 2010).

Conventionally, the developmental log-to-linear shifts are viewed as evidence for representational changes that take place with age and education. According to this view, human and non-human species possess logarithmic representations for number that are responsible for log estimates in young children. With schooling and experience with numbers, children acquire accurate, linear representations, which lead to more linear estimates in older children and adults (Dehaene et al., 2008; Dehaene, 2011; Siegler et al., 2009; Siegler & Opfer, 2003).

Logarithmic and linear representations of number are not mutually exclusive but coexist and compete with one another. For example, individual second graders who generate linear estimates on 0-100 number lines also produce logarithmic estimates on 0- 1,000 number lines (Siegler & Opfer, 2003). In adults, too, numerical estimates can be logarithmic when stimuli are extremely large (Landy et al., 2013) or attention is shared by multiple tasks (Anobile et al., 2012). Thus, the logarithmic-to-linear shift does not appear in a stage-like manner, but occurs gradually, starting with numbers with which children have the greatest familiarity (Siegler et al., 2009). The co-existence of linear and logarithmic representations also helps to explain why aligning 0-1,000 to 0-100 number-line problems results in much more linear number-line estimates for the 0-1,000 problems (Opfer & Siegler, 2007; Opfer & Thompson, 2008).

The co-existence of logarithmic and linear representations of number can be well captured by a mixed log-linear model (MLLM; Anobile et al., 2012), which is formulated as follows:

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

where y denotes the estimate of a given number (x) on a number line with an upper-bound (U). a is a scaling parameter. λ is the index of logarithmic compression in estimates, measuring the relative contribution of log representations in comparison with that of linear representations. When estimates are perfectly logarithmic, meaning that there is no contribution of linear representations, λ equals 1. When estimates are perfectly linear, which indicates no influence of log representations in estimates, λ equals 0.

The logarithmicity component (λ) adequately captures the log-to-linear changes over the course of development (D. Kim & Opfer, 2018; D. Kim & Opfer, 2020; Opfer et al., 2016, 2019). Estimates of numbers in young children exhibit high values of logarithmicity components, which suggests that young children rely heavily on log representations. The values of logarithmicity components decrease with age, indicating that the dependence on logarithmic representations decreases with age. For example, Opfer et al. (2016) found that the λ value in 3-year-olds' estimates was .73 on 0-20 number lines, and the λ value in 5-year-olds' estimates dropped to .07 on the same number lines, presenting a log-to-linear developmental shift. Besides the representational-shift account, it's worth noting that there are alternative accounts for the source of log-linear patterns in number-line estimates. The alternatives include the views that log estimates result from children's lack of familiarity with numbers (Ebersbach et al., 2008; but see Thompson & Opfer, 2010) or of proportional reasoning skills (Zax, Slusser, & Barth, 2019; but see Opfer, Young, & Siegler, 2011; Opfer, Thompson, & D. Kim, 2016; D. Kim & Opfer, 2017). Table A1 summarizes previous number-line studies testing the representational-shift and alternatives accounts using psychophysical models.

1.2. Dynamics in number-line estimation

In addition to number-line estimates changing with age and experience, estimates also change from trial to trial in ways that mirror the dynamics of similarity ratings. Like similarity ratings, the estimate of a number on the number line involves judging the similarity of a given number (e.g., 9 dots) to an upper-bound number (e.g., 30 dots) on a continuum (D. Kim & Opfer, 2018). In a number-line task, an estimate moves towards an upper-bound as a given number is more similar in magnitude to the upper-bound number. On the other hand, if a given number is dissimilar to the upper-bound number, the estimate moves away from the upper-bound.

Similarity judgment has been shown to vary dynamically depending on previous trials. Typically, a pair of objects is judged to be more similar if preceded by highly dissimilar pairs and to be judged more dissimilar if preceded by highly similar pairs (Sjöberg & Thorslund, 1979; Tversky, 1977). For example, two string instruments (e.g., violin and harp) were rated to be more similar when previous trials included pairs of different types of instruments, (e.g., clarinet and harp as a non-string and string instrument pair) than when only the pairs of the same type of instruments (e.g., banjo and harp as a string instrument pair) appeared on previous trials (Sjöberg & Thorslund, 1979).

Kim and Opfer (2018) integrated this context-dependence into the MLLM and developed the dynamic mixed log-linear model (D-MLLM) that includes trial-by-trial calibration in a response scale ($\lambda_i, i > 1$), which is similar to the adaptation level (Helson et al., 1954), the average response level (Petrov & Anderson, 2005), or an internal function that links a psychological value to a response (Birnbaum, 1999):

$$\lambda_i = \lambda_{i-1}(1 - w(x_{i-1} - x_i))\,.$$
$$\text{If} \quad \lambda_i < 0, \text{ then } \lambda_i = 0; \text{ and if } \lambda_i > 1, \text{ then } \lambda_i = 1.$$

While the logarithmicity on the first trial (λ_1) shows the psychological representation prior to any dynamic effects, the logarithmicity after the first trial (λ_i) changes based on previous logarithmicity (λ_{i-1}) and similarity of current and previous numbers to the upper-bound ($(U - x_i) - (U - x_{i-1}) = x_{i-1} - x_i$). There is also a weight parameter (w) that represents a memory component for previous trials. The parameter w is equal to the mean slope of regression lines of current estimates (y_i) on previous numbers (x_{i-1};

Fig. 4 in Kim & Opfer, 2018). This weight parameter determines the adjustment in response scaling (Eq. 2) as well as the relative influence of previous trials (Eq. 4). If estimators are unable to remember or otherwise make use of numbers given on previous trials (i.e., $w = 0$), there are no dynamic effects on estimates.

The dynamic logarithmicity (λ_i) is next combined with the psychological scaling (p_i), which is similar to the conventional MLLM (Eq. 1), except that it reflects trial-by-trial dynamics:

$$p_i = a(\lambda_i \frac{U}{ln(U)} ln(x_i) + (1 - \lambda_i)x_i)\,.$$

The contribution of this internal scaling (p_i) is relatively weighted in comparison to numbers on previous trials (Cicchini et al., 2014), and they together produce an estimate of a current number (y_i):

$$y_i = (1 - w)p_i + w \cdot x_{i-1}, 0 \leq w \leq 1.$$

The D-MLLM predicts how numerical estimates will change as a function of the preceding trials. As in overestimation of similarity between similar objects (e.g., violin and harp) after dissimilar trials (e.g., clarinet and harp), the D-MLLM predicts that dissimilar stimuli (small numbers) to the upper-bound on previous trials increase the logarithmicity component on the current trial of a more similar number (large number; Eq. 2). The increased logarithmicity in turn leads to a more logarithmic estimate of the current number (Eq. 3-4). In contrast, previous trials of similar stimuli (large numbers) to the upper-bound prior to the current trial of a less similar number (smaller number) decrease logarithmicity. The decreased logarithmicity then causes a more linear estimate of the current number. Together, the D-MLLM provides unique predictions for order effects in numerical estimates: estimates are *more logarithmic* if to-be-estimated numbers are given in *forward* order (small → large numbers) and *more linear* if to-be-estimated numbers are given in *backward* order (large → small numbers; Fig. 2A).

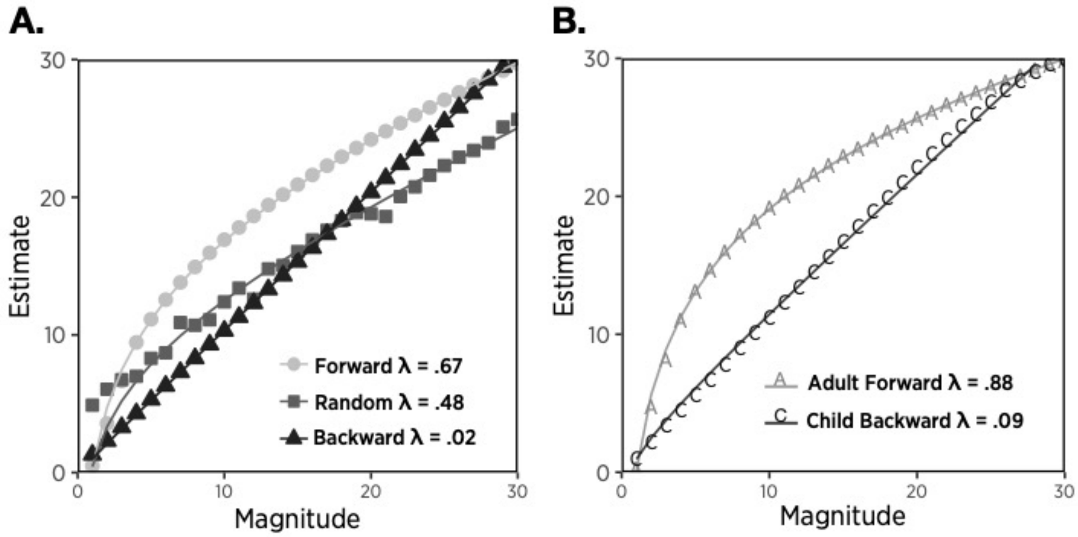

Figure 2. Median estimates of 100-participant simulations using the D-MLLM (A). a was set to 1, the λ_1 was .5, and w was .3 for all orders. A developmental reversal by stimulus order was predicted in D-MLLM simulations (B): λ_1 was set to .72 for adults and 1 for children as found in actual data (D. Kim & Opfer, 2018). a was set to 1, and w was .1 for both age groups.

Using the D-MLLM, Kim and Opfer (2018) demonstrated the interaction between logarithmic representations and dynamic processes over trials in 0-30 numerosity estimation. More specifically, adults were logarithmic on the first trial (λ_1 = .47 to .85) but became more linear over trials using inter-trial differences ($\lambda_{\text{last trial}}$ = .04 to .18). However, this was only true when numerosities were given in *random* order. When they estimated the same numerosity for every trial, such that there were zero inter-trial differences ($x_{i-1} - x_i = 0$), adult participants remained logarithmic from the first to the last trial (λ_1 = .87, $\lambda_{\text{last trial}}$ = .40).

In contrast, in the identical number-line task, in which numerosity was given in random order, children failed to incorporate previous trials into current estimates ($w = 0$) and yielded numerosity estimates that were consistently logarithmic from the first to the last trials (λ_1 = 1.00, $\lambda_{\text{last trial}}$ = 1.00). This constantly high logarithmicity was adequately simulated by the D-MLLM with a zero w. Kim and Opfer (2018) postulated that the lack of between-trial dependence in children might be due to limited memory resources required to track different numbers trial to trial (Ciesielski et al., 2006; Kwon et al., 2002). They further suggested that if to-be estimated numerosities were given in a certain way that eases memory loads, such as consistently forward and backward order, children might be able to dynamically update previous trials onto current estimates. This speculation provides a very interesting prediction for a developmental reversal in numerosity estimation: when numbers are given in backward order, so children can use previous trials to decrease logarithmicity, they will become as linear as adults. On the other hand, if numbers are given in forward order, such that logarithmicity increases over trials, even adults will produce more logarithmic estimates that look like children's estimates (Fig. 2B).

1.3. Overview of Studies

In the current Studies, we test this novel prediction of the D-MLLM by manipulating order of to-be-estimated stimuli. Specifically, we asked adults (Study 1, Study 3) and children (Study 2) to estimate numerosities given in random, forward, and/or backward order on number lines. If number-line estimation is free from trial-to-trial dynamics, there will be no difference in logarithmicity across forward, backward, and random conditions. Specifically, adults in the forward condition will be as linear as in the random condition, whereas children in the backward condition will be as logarithmic as in the random condition, without a predicted developmental reversal. In contrast, if log compression results from the interaction between log representations and dynamic updates of previous trials, estimates will be the most logarithmic in the forward condition and the most linear in the backward condition. Indeed, given previously observed values for the parameters of the D-MLLM, children would be expected to produce more linear estimates than adults in the backward condition, whereas adults would be expected to produce more logarithmic estimates in the forward condition. As should be evident, this is not a trivial prediction, and it appears to be a unique implication of the computational model.

2. Study 1

In Study 1, we revisit dynamics in adults' numerosity estimates and test if salient contexts (forward and backward order) drive increased or decreased logarithmicity as predicted by the D-MLLM (D. Kim & Opfer, 2018). Extending Kim and Opfer (2018), we use larger to-be-estimated numerosities (5 to 495 dots) than the previous study (5 to 29 dots). Large numerosities are perceptually challenging and lead to more logarithmic estimates in adults (Lee et al., 2019). Therefore, this manipulation might impede dynamic calibrations for linear estimation in the random condition (via large λs or a small w of the D-MLLM), putting adults on an equal footing with children.

2.1. Method

2.1.1. Participants

We collected data from 49 undergraduate students (21 females, M = 20.01 years, SD = 1.71 years; 45% Caucasian, 29% Asian, 10% African American, 6% Hispanic, 4% Native American, 2% Multiracial, and 4% no answer) at The Ohio State University. Participants received course credit in return for their participation.

2.1.2. Materials and procedure

Participants completed a number-line estimation task. The task consisted of three conditions (forward, backward, and random) given over three blocks in counterbalanced order. Each condition included 36 to-be-estimated numerosities that were non-subitizable and evenly sampled from the 0-500 range: 5, 19, 33, 47, 61, 75, 89, 103, 117, 131, 145, 159, 173, 187, 201, 215, 229, 243, 257, 271, 285, 299, 313, 327, 341, 355, 369, 383, 397, 411, 425, 439, 453, 467, 481, and 495. The conditions were the same except for stimulus order. In the forward condition, the stimuli were presented in ascending order (5 → 495), whereas the stimuli were given in descending order (495 → 5) in the backward condition. In the random condition, to-be-estimated numbers were presented in random order. The stimuli were estimated only once per condition–i.e., 36 trials per condition.

On every trial, a given number of dots above a line flanked with 0 and 500 dots were displayed for 2 seconds and then masked with a random noise image (Fig. 1). Participants were asked to click on the line where a given number of dots belonged. There was no time limit for response. The size of dots was controlled; The dots of a stimulus were the same in size as the dots at the upper bound on half the trials and the same in total area as the upper-bound dots on half the trials. There was neither practice nor feedback provided.

2.2. Results and discussion

First, we computed trial-to-trial logarithmicity of median estimates collapsed over participants in the random condition. The random condition was the only condition, to which the MLLM could be fit trial by trial. In the forward and backward conditions, because every participant received numerosities in the same order, there was only one stimulus per trial number that the model could use, which makes it impossible to compute the logarithmicity component trial by trial. Compared to the previous study using the 0-30 number lines ($w =$ .08 to .13 in D. Kim & Opfer, 2018), the inter-trial dependence in the current Study was slightly smaller ($w = .06$), but still significant ($p =$ .01 in bootstrapping significance test). Despite the moderate inter-trial dependence, the logarithmic components on the first and the last trials did not differ so much (λ_1 = .42 and λ_{36} = .46), and there was no decrease in logarithmicity over trials ($b = -.09$, $p = .16$ in a log-log transformed regression). The absence of the log-to-linear change might be due to large numerosities in the current Study that might have increased perceptual demands and hindered trial-to-trial linearization. When estimates were simulated from the D-MLLM using w and λ_1 from the data, simulated estimates explained the actual data moderately well ($R^2 = .83$ in median estimates), although they exhibited the log-to-linear change over trials ($b = -.23$, $p < .01$).

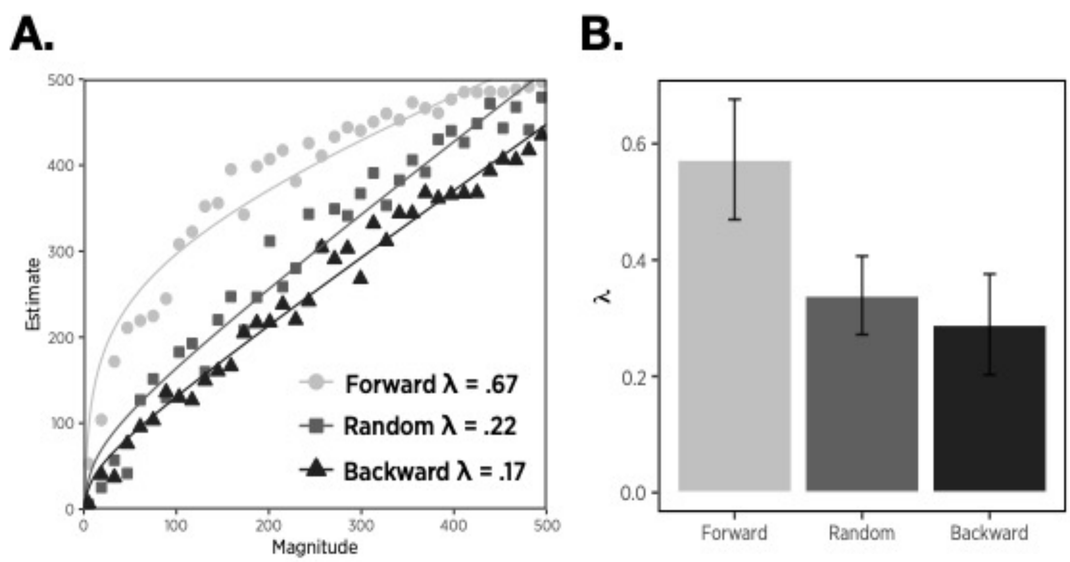

Figure 3. Median estimates (A) and mean logarithmicity values of individual adults (B) in the first block in Study 1. Error bars indicate ± SEM.

Next, we examined whether dynamic effects predicted by the D-MLLM were evident across conditions. To do so, we computed logarithmicity (λ) in median estimates. Median estimates did not seem to differ by condition, presenting similar logarithmicity across conditions (λ = .37 in forward, .28 in random, and .3 in backward). However, it is possible that the within-subject design, in which participants completed one condition after another, yielded *cross-conditional* dynamics and confounded with dynamic effects *within a condition*. Consistent with this postulation, the differences in logarithmicity were more pronounced when we analyzed median estimates from the first block, in which there were no previous blocks that could have affected the current block (Fig. 3A). As predicted by the D-MLLM, estimates were the most logarithmic in the forward condition (λ = .67) and the least logarithmic in the backward condition (λ = .17). Logarithmicity in the random condition fell in-between that of forward and backward conditions (λ = .22).

Dynamic updates in numerosity estimates were next examined at the individual level. When logarithmicity was computed for each individual, the averaged logarithmicity components were not different across conditions if block numbers were not accounted for (M_λ = .43 in forward, .36 in random, and .40 in backward, $F(2, 96)$ = 1.89, η_p^2 = .04, p = .16). However, when the first block was only considered, the difference in logarithmicity across conditions became more evident ($F(2, 46)$ = 3.02, η^2 = .12, p = .06). In line with the prediction of the D-MLLM, log compression of the first block was the greatest in the forward condition (M_λ = .57, SD_λ = .41), the smallest in the backward condition (M_λ = .29, SD_λ = .35), and in between that of forward and backward conditions in the random condition (M_λ = .34, SD_λ = .28; Fig. 3B).

The interaction between cross- and within-block dynamics (block order× stimulus order) was further examined in a linear mixed-effects regression with participants as random effects (random intercepts) and block number, condition, and their interaction as fixed-effects predictors of logarithmicity components. As shown in Table 1, compared to random order, logarithmicity substantially increased when to-be-estimated numbers were provided in forward order (b = .44, $p < .001$). The effects of forward order decreased as the forward condition was given in a later block (b = –.18, p = .001). However, the backward order neither affected logarithmic compression in estimates (b = .05, $p > .05$), nor showed an interaction with block number (b = –.01, $p > .05$). Together, numerical estimates dynamically changed with numbers previously encountered within a block and over three blocks. When block order was controlled, logarithmicity varied by stimulus order: as predicted in the D-MLLM simulations, adults, in particular, became more like children in the forward condition.[1]

[1] The D-MLLM also predicts the interaction between condition and block. Parameter values obtained from actual data in the random condition (a = 1.09, λ_1= .42, w = .06) were used to simulate estimates in all three conditions. In the stimulations, prior trials were always integrated into a current estimate regardless of block order. For example, the first trial in the second block was affected by the history of previous trials in the first block. As found in actual estimates, simulations presented greater logarithmicity in the forward condition than in the random condition (b = 1.23, $p < .001$), whereas estimates in the backward condition were not different from those in the random condition (b = -.07, p = .18). Block itself did not have significant effects (b = -.01, p = .49), but interacted with condition. The effect of forward order decreased as the condition was present in a later block (b = -.47, $p < .001$), whereas backward-order effects did not vary with block order (b = .01, p = .62). Although the predicted effects were a bit stronger than actual findings, the D-MLLM could describe the interaction between stimulus and block order, showing that block effects, in fact, reflect the dynamic effects of the history of previous trials.

Table 1.

Results of a mixed-effects regression model on logarithmicity (λ) in Study 1. The baseline is the random condition.

Predictor	*b*	*SE*	*df*	*t*	*p*
(Intercept)	.22	.08	138.41	2.65	.009**
Block number	.07	.04	114.56	.48	.07
Forward	.44	.12	111.19	1.81	.0002***
Backward	.05	.12	114.00	3.73	.63
Block number × Forward	–.18	.06	115.98	–3.28	.001**
Block number × Backward	–.01	.06	116.58	–.15	.88

Note. ** $p < .01$, *** $p < .001$.

3. Study 2

Study 1 demonstrated that adults' numerosity estimates change with previously encountered numbers and that dynamic effects could make adults become like children, producing more logarithmic estimates. In Study 2, we examined whether children could integrate prior trials into their estimates and become like adults, yielding more linear estimates. Although children seem unable to use dynamic information in the random condition, it is probably due to the limits in cognitive resources (D. Kim & Opfer, 2018). Children may become able to use priors in current responses if stimuli are presented in a systematic order, such that it is easy to track previous trials and predict what comes next.

3.1. Method

3.1.1. Participants

Ninety-three children were recruited at a children's science museum in Columbus, OH (47 girls, *M* = 8.13 years, *SD* = 2.38 years, age range = 4.11 to 15.17 years; 83% Caucasian, 9% African American, 4% Asian, 2% Native American, 1% Hispanic, and 1% Multiracial). Children's parents/guardians completed a consent form prior to testing and waited in a waiting area while children completed the number-line task in a testing room. All children were rewarded with a sticker for participating.

3.1.2. Materials and procedure

The identical number-line task was used in Study 2, with two exceptions: the smaller number range and between-subject design. Children were randomly assigned to one of forward (*n* = 32, *M* = 8.04 years, *SD* = 2.51 years), backward (*n* = 30, *M* = 8.17 years, *SD* = 2.21 years), and random (*n* = 31, *M* = 8.19 years, *SD* = 2.49 years) conditions and completed the assigned condition only. We also used 0-30 number lines, on which young children previously failed to incorporate previous trials into current estimates when stimuli were given in random order (D. Kim & Opfer, 2018). Twenty to-be-estimated numbers were selected to evenly sample non-subitizable numbers: 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 19, 20, 21, 22, 23, 27, 28, and 29.

3.2. Results and discussion

Six children who experienced technical issues, could not complete the task, or only chose the two endpoints (i.e., 0 and 30) for all stimuli were excluded from analyses. We first computed trial-to-trial logarithmicity in estimates in the random condition. Children's numerosity estimates resembled those in the previous study (D. Kim & Opfer, 2018); there seemed to be neither solid memory for previous numbers ($w = -.017$, p = .73 from a bootstrapping significance test) nor log-to-linear calibration in estimates over trials ($b = -.008, p = .53$). Although logarithmicity on the first trial was moderately logarithmic (λ_1 = .24), overall estimates from the first to the last trials showed constant logarithmicity (M_λ = .47). In line with these findings in actual data, simulated estimates by the D-MLLM did not present a trial-to-trial decrease in logarithmicity ($b = -.002, p = .50$) and explained children's actual estimates adequately well ($R^2 = .94$).

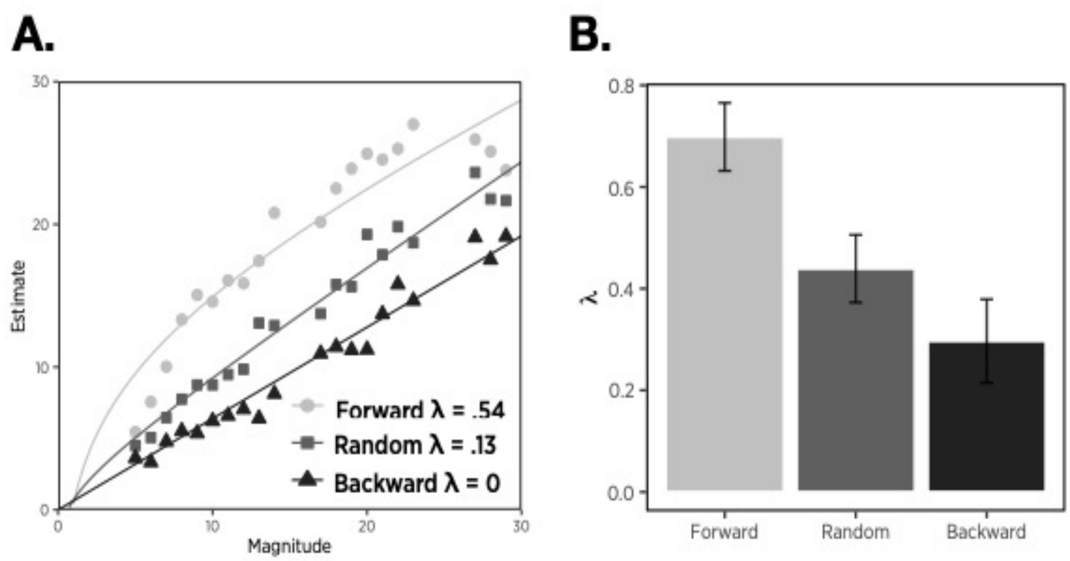

Figure 4. Median estimates (A) and mean logarithmicity values of individual children (B) in Study 2. Error bars indicate ± SEM.

Logarithmicity in estimates across conditions was next analyzed at the group level. Consistent with findings in adults in Study 1 and predictions by the D-MLLM, logarithmic compression in median estimates varied with stimulus order. Median estimates were the most logarithmic when stimuli were given in forward order ($\lambda = .54$) and less logarithmic when they were given in random order ($\lambda = .13$). More importantly, as hypothesized, children became more like adults, producing perfectly linear estimates, when numbers were provided in backward order ($\lambda = 0$; Fig. 4A).

In line with the group-level findings, logarithmic compression significantly changed with stimulus order at the individual level as well ($F(2,84) = 8.03$, $\eta^2 = .16$, $p < .001$). When logarithmicity was computed for each child, logarithmicity components in the forward condition were the greatest ($M_\lambda = .70$, $SD_\lambda = .37$), followed by those in the random condition ($M_\lambda = .44$, $SD_\lambda = .36$). As expected, logarithmicity components were the smallest in the backward condition ($M_\lambda = .30$, $SD_\lambda = .43$; Fig. 4B). In fact, the mean λ value in children's backward-order estimates was much smaller than that in adults' forward estimates on 0-500 number lines in Study 1 ($M_\lambda = .57$), indicating a developmental reversal by dynamic effects.

The dynamic effects were also well captured in a regression model with children's age and condition as predictors of logarithmicity components (Table 2). Age was significantly predictive of children's logarithmicity ($b = -.14$, $p < .001$). Replicating previous work (D. Kim & Opfer, 2018; Opfer & Siegler, 2007; Siegler & Opfer, 2003), children's estimates were more linear with age, exhibiting the log-to-linear shift in development. Logarithmicity also changed with stimulus order; Compared to the random condition, logarithmicity components were significantly greater in the forward condition ($b = .23$, $p < .05$). Children's logarithmicity decreased with the descending order of stimuli although it was marginally significant ($b = -.16$, $p < .10$). We also conducted another regression analysis to examine interactions between age and condition. The results remained almost identical for age and condition, and the age × forward and age × backward interaction terms were insignificant ($p > .05$), indicating the developmental decrease in logarithmicity was consistent across conditions.

Table 2.

Results of a regression model on logarithmicity (λ) in Study 2. The baseline is the random condition.

Predictor	*b*	*SE*	*t*	*p*
(Intercept)	.45	.07	6.90	.0001***
Age	–.14	.04	–3.70	.0004***
Forward	.23	.09	2.51	.014*
Backward	–.16	.10	–1.67	.098†

Note. $^{\dagger}p < .10$, * $p < .05$, *** $p < .001$.

Moreover, this model with additional interaction terms did not explain data better than the simpler regression model without interaction (ΔBIC = 7.89). Taken together, the log-to-linear improvement in development persisted regardless of stimulus order. More importantly, we found that children became capable of using previous trials in current estimates when stimuli changed systematically, producing more logarithmic estimates in the forward condition and more adult-like, linear estimates in the backward condition.

4. Study 3

To compare children's and adults' performance better, we carried out Study 3. In this Study, adult participants were asked to complete the forward condition that was identical to that in Study 2. We hypothesized that the stimuli ascending over trials would increase logarithmicity, leading to child-like estimates in adults.

4.1. Method

4.1.1. Participants

Thirty adults were recruited from Amazon Mechanical Turk (11 females, *M* = 37.95 years, *SD* = 10.99 years, age range = 25.67 to 65.33 years; 60% Caucasian, 13% Native American, 10% African American, 7% Asian, 3% Hispanic, and 7% no answer). Prior to proceeding to the study, all participants were asked to provide consent for their participation. The study took less than 5 minutes, and participants received $1.5 in return for their participation.

4.1.2. Materials and procedure

Participants were asked to estimate numerosities given in forward order on 0-30 number lines. The task was the same as the forward condition in Study 2.

4.2. Results and discussion

One participant who clicked at random places too quickly (mean RT < 500ms) was removed from analyses.

First, we examined the degree of logarithmic compression in median estimates. The median estimates were pretty logarithmic for adults ($\lambda = .44$, Fig. 5). The logarithmicity component was greater in value than those in previous research where adults estimated numerosities given in random order on the same numerical scales (e.g., $\lambda = 0$ to .11 in D. Kim & Opfer, 2018). More importantly, the median estimates by adults were more logarithmic than those by children in the backward and random conditions in Study 2.

Next, logarithmicity was computed for each individual. Individual adults produced quite logarithmic estimates ($M_{\lambda} = .58$, $SD_{\lambda} = .40$). Compared to children in Study 2, adults in Study 3 were indeed more logarithmic than children in the backward condition ($b = -.29$, $p < .01$). They appeared to be less logarithmic than children in the forward condition, but the difference was not significant ($b = .11$, $p = .25$). Together, even when the number scale was controlled to be the same, adults' estimates in the forward condition were more logarithmic than children's estimates in the backward condition at both group and individual levels. Together, these results provide supportive evidence for a developmental reversal created by stimulus order.

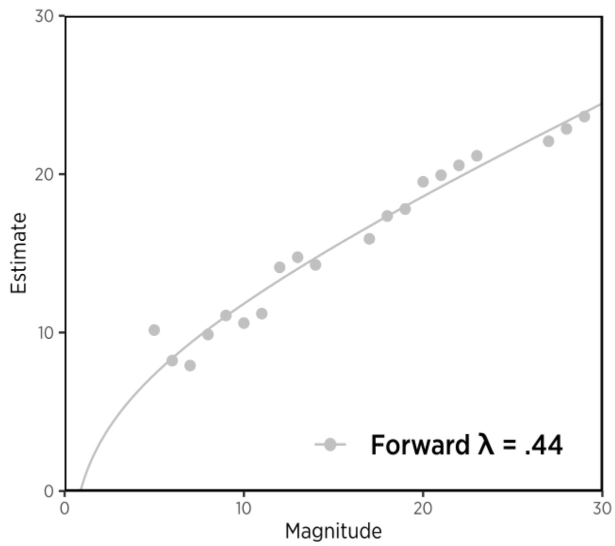

Figure 5. Median estimates (A) in Study 3.

5. General Discussion

In the present paper, we sought to test the validity of the D-MLLM of number-line estimation by testing a unique and unexpected implication. Specifically, the D-MLLM predicts a particular developmental reversal as a function of stimulus order. In line with the model predictions, there was a clear developmental reversal in the current Studies. Adults in Study 1 and 3 made more logarithmic estimates when stimuli were ascending over trials. On the other hand, children in Study 2 produced more linear estimates when stimuli were descending over trials. Furthermore, as predicted by the D-MLLM, the inflation and the deflation of logarithmicity by stimulus order were present within age groups. Both adults and children, at the group and individual levels, were more logarithmic in the forward condition and more linear in the backward condition, while their logarithmicity in the random condition fell in-between that of forward and backward conditions. Altogether, these findings provide experimental evidence that the D-MLLM is a model that adequately captures the underlying processes of number-line estimation.

The context effects by stimulus order in the current Studies also provide strong evidence against an argument that log compression in numerical estimation comes from a central tendency of judgment (Cicchini et al., 2014). For this account, a given numerosity is represented rather linearly, but its estimate is corrupted by the prior, the history of previous trials. The influence of previous trials creates certain biases in current estimates: a current numerosity is underestimated after small-number trials and overestimated after large-number trials, resulting in log-like compression in overall estimates. Although the dependence on previous trials may lead to more logarithmic estimation (D. Kim & Opfer, 2018), however, this sequential mechanism alone cannot account for the order effects observed in the present Studies.

For the central tendency account, constant underestimation would be expected in the forward condition, where previous trials were always smaller than a current numerosity. On the other hand, stimuli should have been overestimated in the backward condition, in which a current numerosity was always smaller than numerosities on previous trials. This under- or overestimation does not necessarily mean that overall estimates would be compressive or expansive. Estimates can be very linear with an overall tendency of over- or underestimation. When estimates were simulated using the central tendency model by Cicchini et al. (2014), they were perfectly linear in both forward and backward conditions for adults in Study 1 and children in Study 2, with poor to moderate fittings (λs = 0, R^2= .24 to .77). Inconsistent with these predictions, however, actual estimates by both children and adults were more logarithmic with a tendency of overestimation in the forward condition, while more linear with an underestimation tendency in the backward condition. For adults in Study 3, the model failed to generate estimates that describe the actual data (R^2= 0). In fact, the model without dynamic update, which is equal to a linear model, appeared to explain the actual data better (R^2= .87). Together, this suggests that the simple sequential effects are not the sole driving force of logarithmic or linear estimation. Instead, log and linear representations are employed in numerical estimation, and the reliance on each representation is determined dynamically over trials.

One may argue that spatial restriction or adaptation effects might have elicited systematic distortion found in the current Studies. More specifically, there might have been growing concerns about the limited space in forward and backward conditions as a task proceeded. If so, estimates for numbers on later trials would have been crammed into the right side of a number line in the forward and the left side in the backward condition. Alternatively, participants might have adapted to small or large numerosities in the beginning of a task—i.e., small numbers in the forward and large numbers in the backward condition. The adaptation effects could have led to overestimation for large numbers on later trials in the forward condition and underestimation for small numbers in the backward condition. In both cases, estimates were expected to be compressive for forward order and to be expansive (rather than linear) for the backward order.

We examined this possibility using Stevens' power model ($y = a \bullet x^{\beta}$). In this model, parameter β is 1 when estimates are perfectly linear, smaller than 1 when estimates are compressive, and larger than 1 when estimates are expansive. For Study 1, median estimates in the first block were compressive regardless of stimulus order (βs = .37 to .76). Importantly, the β value in the backward condition was significantly smaller than 1 (β = .76, 95% CI = [.68, .87]), indicating that backward-order estimates were considerably compressive rather than linear or expansive. This is also in line with our findings, where estimates were found logarithmic for all conditions. In Study 2, children's median estimates were compressive in forward and random conditions (β =

.62 and .88 respectively), and expansive in the backward condition (β = 1.10). However, the β value of the backward condition was not different from 1 (95% CI = [.92, 1.25]), meaning that backward estimates were not significantly different from linear estimates. This is consistent with our findings using the MLLM, in which estimates were logarithmic in forward and random conditions and linear in the backward condition. Together, log and linear patterns in estimates in the current Studies are not accounted for by spatial restrictions or perceptual adaptation.

What made it possible for children to use previous trials in the forward and backward conditions? Replicating the previous finding (D. Kim & Opfer, 2018), children in Study 2 failed to incorporate previous numerosities into current estimates in the random condition. Considering that the current Study involved older children (the highest age = 15.17 years; *M* = 8.13 years) than the previous work (the highest age = 6.99 years; *M* = 6.02 years), this result suggests that the ability for dynamic update may emerge slowly in later development. When stimuli systematically increased or decreased, however, children became able to use previous trials to adjust their current responses. The systematic change in stimuli might alleviate memory loads, enabling children to remember and exploit previous trials better – i.e., significant w of the D-MLLM. In addition to memory, conscious registration for previous trials could be promoted in the forward and backward conditions, enhancing the expectancy for current stimuli. Recent work by (S. Kim et al., 2020) has demonstrated that although the prior may affect current responses at a low sensory level, the perceptual history of previous trials is constructed at a high level that involves conscious perception of stimuli. Aligned with this, most children in the forward and backward conditions in Study 2 showed the explicit awareness of the systematic change in stimuli, correctly pointing out the ascending or descending order of to-be-estimated numerosities.

In Study 1, we also showed that the numerosity size might affect dynamic effects in adults. As found in the previous research using 0-30 number lines (D. Kim & Opfer, 2018), adults in the current Study incorporated previous trials into current estimates even on larger number lines – i.e., 0 to 500. However, compared to estimates on 0-30 number lines, adults' estimates of large numerosities exhibited weaker dependence on previous trials. The trial-by-trial logarithmicity in the current Study did not vary too much and remained constantly high, whereas the estimates of small numerosities on 0-30 number lines were logarithmic on the first trial but became very linear after few trials (D. Kim & Opfer, 2018). This finding insinuates that along with the duration of stimulus display (Crawford et al., 2000) and multi-task paradigms (Anobile et al., 2012; Cicchini et al., 2014), the magnitude of stimuli may change dynamic processes in adults. Previous research has shown that larger numerosity leads to greater logarithmicity in estimates collapsed across all trials (Lee et al., 2019), but nothing is known about trial-to-trial logarithmicity in small vs. large number estimates. The role of the numerical magnitude in dynamics and internal scaling could be an interesting venue for future research.

Importantly, the results from the current Studies call attention to controlling for experimental parameters. Even though dynamics effects are evident in number line tasks, researchers often fail to consider them by using ascending stimulus order (Gross et al., 2018), fixed order for every individual (Moeller et al., 2009; Weijdena et al., 2018; Yuan et al., 2019), or non-counterbalanced or non-randomized condition order (Cohen & Sarnecka, 2014; Link et al., 2014; Regina M. Reinert et al., 2015; Regina Miriam Reinert et al., 2019; Sella et al., 2015; Weijdena et al., 2018). The current study also showed that block order needs to be controlled for especially when blocks can create different contexts. In the current study, a previous block with ascending or descending stimuli appeared to form a unique context that affects estimates in the following blocks, whereas block effects were not evident in previous work, where numbers were always randomly presented (D. Kim & Opfer, 2018).

In addition, trial-to-trial dynamics found in the current Studies may not be limited in number line tasks, but bias other tasks tapping into numerical representations. For example, inter-trial dynamics may affect children's performance in a give-a-number task (GAN), a number recall task, or number comparison, all of which show strong associations with number-line estimation (Laski & Siegler, 2007; Opfer et al., 2019; Thompson et al., 2017). Supporting this idea, Odic et al. (2014) found that children compared numbers more accurately after easy trials than hard trials. This perceptual hysteresis—i.e., the dependence on previous trials—required explicit feedback on children's performance regardless of whether feedback was accurate or inverted. The authors concluded that feedback helped children to form their internal confidence in numerical comparison and that no matter whether the confidence is low or high, a stable confidence status might cause the hysteresis effects. In the current study, previous trials affected children's estimation even in the absence of feedback, but only if stimuli were presented in salient, systematic order. This suggests that the forward or backward order of stimuli may be sufficient for children to establish internal confidence in the absence of feedback, such that children are encouraged to remember and rely on trials prior to a current one. Consistent with this, the accuracy of number-line estimation is strongly associated with confidence in the estimation (Fitzsimmons et al., 2020; Rivers et al., 2020). Together, the findings in the present paper suggest that it is important to carefully control for dynamic effects across conditions as well as over trials in numerical tasks. Without control over these parameters, adults can be made to look like children (and vice-versa).

We would like to note the limitation of the current Studies that dynamic effects were examined only in non-symbolic, numerosity estimation. Previous work showed that there were no dynamic updates in estimates of symbolic numbers (i.e., Arabic numerals) in both children and adults (D. Kim & Opfer, 2018). Despite the absence of dynamic effects, age-related differences were evident in symbolic number estimates from the very first trial: from the first to the last trials, children were steadily logarithmic, whereas adults were completely linear. This suggests that the true developmental change in numerical representation may be better captured in symbolic number estimation, which is not susceptible to trial-by-trial dynamics. What is unknown is whether symbolic number estimates are still free from strong priors, like ascending or descending stimuli, that predict next trials. Future studies may examine the effects of stimulus order and dynamic encoding on symbolic number estimation in comparison with numerosity estimation.

In sum, we experimentally created situations, in which children became more like adults and adults became more like children in number line estimation. As predicted by the D-MLLM, task contexts changed logarithmicity in children's and adults' estimates. Therefore, the D-MLLM accurately predicts dynamic processes and developmental changes in number line estimation.

Acknowledgments

This study was supported by grant R305A160295 from the Institute of Education Sciences (IES). We would like to thank the children and their parents for participating in the current study, as well as Katie Kuhlwein and Wei Fang for their help in data collection. We also thank Jike Qin for her review and Elida Laski and Bob Siegler for pointing out the similar phenomenon observed in symbolic number-line estimation.

References

Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*(3), 454–459.

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology, 111*(2), 246-267.

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*(2), 545.

Birnbaum, M. H. (1999). How to show that 9> 221: Collect judgments in a between-subjects design. *Psychological Methods, 4*(3), 243.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*(1), 189–201.

Chesney, D. L., & Matthews, P. G. (2018). Task constraints affect mapping from Approximate Number System estimates to symbolic numbers. *Frontiers in Psychology, 9*, 1801.

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*, *111*(21), 7867–7872.

Ciesielski, K. T., Lesnik, P. G., Savoy, R. L., Grant, E. P., & Ahlfors, S. P. (2006). Developmental neural networks in children performing a categorical n-back task. *Neuroimage*, *33*(3), 980–990.

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review, 18*(2), 331–338. http://doi.org/10.3758/s13423-011-0059-z

Cohen, D. J., Blanc-Goldhammer, D., & Quinlan, P. T. (2018). A mathematical model of how people solve most variants of the number-line task. *Cognitive Science*, *42*(8), 2621-2647. doi:10.1111/cogs.12698

Cohen, D. J., & Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology*, *50*(6), 1640.

Crawford, L. E., Huttenlocher, J., & Engebretson, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science*, *11*(4), 280–284.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.

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*(5880), 1217–1220.

DeVries, J. M., Kuhn, J. T., & Gebhardt, M. (2020). What applying growth mixture modeling can tell us about predictors of number line estimation. *Journal of Numerical Cognition*, *6*(1), 66-82.

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), 1-17.

Fitzsimmons, C. J., Thompson, C. A., & Sidney, P. G. (2020). Confident or familiar? The role of familiarity ratings in adults' confidence judgments when estimating fraction magnitudes. *Metacognition & Learning*, *15*(2).

Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., & Van Lieshout, E. C. D. M. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology*, *134*, 12–29. http://doi.org/10.1016/j.jecp.2015.02.002

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/f noise in human cognition. *Science*, *267*(5205), 1837–1839.

Gross, S. I., Gross, C. A., Kim, D., Lukowski, S. L., Thompson, L. A., & Petrill, S. A. (2018). A comparison of methods for assessing performance on the number line estimation task. *Journal of Numerical Cognition*, *4*(3), 554–571.

Helson, H., Michels, W. C., & Sturgeon, A. (1954). The use of comparative rating scales for the evaluation of psychophysical data. *The American Journal of Psychology*, *67*(2), 321–326.

Heine, A., Thaler, V., Tamm, S., Hawelka, S., Schneider, M., Torbeyns, J., De Smedt, B., Verschaffel, L., Stern, E., & Jacobs, A. (2010). What the eyes already "know": using eye movement measurement to tap into children's implicit numerical magnitude representations. *Infant and Child Development*, *19*(2), 175–186. http://doi.org/10.1002/icd.640

Hoffmann, D., Hornung, C., Martin, R., & Schiltz, C. (2013). Developing number–space associations: SNARC effects using a color discrimination task in 5-year-olds. *Journal of Experimental Child Psychology*, *116*(4), 775-791.

Hurst, M., Leigh Monahan, K., Heller, E., & Cordes, S. (2014). 123s and ABC s: developmental shifts in logarithmic-to-linear responding reflect fluency with sequence values. *Developmental Science*, *17*(6), 892-904.

Jung, S., Roesch, S., Klein, E., Dackermann, T., Heller, J., & Moeller, K. (2020). The strategy matters: Bounded and unbounded number line estimation in secondary school children. *Cognitive Development*, *53*, 100839. doi:10.1016/j.cogdev.2019.100839

Kim, D., & Opfer, J. E. (2017). A unified framework for bounded and unbounded numerical estimation. *Developmental Psychology*, *53*(6), 1088.

Kim, D., & Opfer, J. E. (2018). Dynamics and development in number-to-space mapping. *Cognitive Psychology*, *107*, 44–66.

Kim, D., & Opfer, J. E. (2020). Compression is evident in children's unbounded and bounded numerical estimation: Reply to Cohen and Ray. *Developmental*

Psychology, *56*(4), 853.

Kim, S., Burr, D., Cicchini, G. M., & Alais, D. (2020). Serial dependence in perception requires conscious awareness. *Current Biology*, *30*(6), R257–R258.

Kwon, H., Reiss, A. L., & Menon, V. (2002). Neural basis of protracted developmental changes in visuo-spatial working memory. *Proceedings of the National Academy of Sciences*, *99*(20), 13336–13341.

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

Laski, E. V., & Yu, Q. (2014). Number line estimation and mental addition: Examining the potential roles of language and education. *Journal of Experimental Child Psychology*, *117*, 29-44. http://doi.org/10.1016/j.jecp.2013.08.007

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive Science*, *37*(5), 775–799.

Lee, S., Kim, D., Opfer, J. E., Pitt, M. A., & Myung, I. J. (2019). *Active learning for a number-line task with two design variables*. The 41st Annual Meeting of the Cognitive Science Society.

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.

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology*, *103*(4), 503–515.

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General, 143*(1), 255.

Opfer, J. E., Kim, D., Young, C. J., & Marciani, F. (2019). Linear spatial-numeric associations aid memory for single numbers. *Frontiers in Psychology*, *10*, 146.

Opfer, J. E., & Martens, M. A. (2012). Learning without representational change: development of numerical estimation in individuals with Williams syndrome. *Developmental Science, 15*(6), 863–875. http://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*(3), 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*(5), 1194–1204. http://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*(3), 788–804.

Opfer, J. E., Thompson, C. A., & Kim, D. (2016). Free versus anchored numerical estimation: A unified approach. *Cognition*, *149*, 11–17.

Parducci, A. (1963). Range-frequency compromise in judgment. *Psychological Monographs: General and Applied*, *77*(2), 1.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*(6), 407.

Petrov, A. A., & Anderson, J. R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review*, *112*(2), 383.

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*(5), 285-293.

Reinert, R. M., Hartmann, M., Huber, S., & Moeller, K. (2019). Unbounded number line estimation as a measure of numerical estimation. *PloS ONE*, *14*(3).

Reinert, R. M., Huber, S., Nuerk, H.-C., & Moeller, K. (2017). Sex differences in number line estimation: The role of numerical estimation. *British Journal of Psychology*, *108*(2), 334–350. http://doi.org/10.1111/bjop.12203

Reinert, R. M., Huber, S., Nuerk, H.-C., & Moeller, K. (2015). Strategies in unbounded number line estimation? Evidence from eye-tracking. *Cognitive Processing*, *16*(1), 359–363.

Rivers, M. L., Fitzsimmons, C. J., Fisk, S. R., Dunlosky, J., & Thompson, C. A. (2021). Gender differences in confidence during number-line estimation. *Metacognition and Learning*, *16*(1), 157-178.

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, *30*(2), 344–357. http://doi.org/10.1111/j.2044-835X.2011.02048.x

Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2015). Varieties of quantity estimation in children. *Developmental Psychology*, *51*(6), 758.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237–250.

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*(3), 143–150.

Sjöberg, L., & Thorslund, C. (1979). A classificatory theory of similarity. *Psychological Research*, *40*(3), 223–247.

Thompson, C. A., Morris, B. J., & Sidney, P. G. (2017). Are books like number lines? Children spontaneously encode spatial-numeric relationships in a novel spatial estimation task. *Frontiers in Psychology, 8*, 2242.

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*(1), 20–51. http://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*(6), 1768–1786.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327.

van der Weijdena, F. A., Kamphorsta, E., Willemsena, R. H., Kroesbergenab, E. H., & van Hoogmoed, A. H. (2018). Strategy use on bounded and unbounded number lines in typically developing adults and adults with dyscalculia: An eye-tracking study. *Journal of Numerical Cognition*, *4*(2), 337–359.

White, S. L., & Szűcs, D. (2012). Representational change and strategy use in children's number line estimation during the first years of primary school. *Behavioral and Brain Functions*, *8*(1), 1. http://doi.org/10.1186/1744-9081-8-1

Yuan, L., Prather, R., Mix, K. S., & Smith, L. B. (2019). Number representations drive number-line estimates. *Child Development*.

Zax, A., Slusser, E., & Barth, H. (2019). Spontaneous partitioning and proportion estimation in children's numerical judgments. *Journal of Experimental Child Psychology*, *185*, 71-94.

Appendix

Table A1

Summary of previous studies examining number-line estimation.

Study	Participants	Number format	Number Scale	Models Applied	Best-fitting Model
Siegler & Opfer (2003)	Children aged 7 - 11 years and	Symbolic	0-100, 0-1,000	Log, linear	For young children, linear in 0-100 and log in 0-1000 tasks; for older children and adults, linear in both
Laski & Siegler (2007)	Children aged 5-8 years	Symbolic	0-100	Log, linear	Log for young and linear for older children in pretest; more linear in posttest after training
Opfer & Siegler (2007)	Children aged 8 years	Symbolic	0-1,000	Log, linear	Log before training, linear after training
Dehaene et al. (2008)	Mundurucu adults and children aged 7-17 years,	Non-symbolic	1-10, 10-100	Log, linear	For Mundurucu participants, log in both; for American adults, log in 10-100 and linear in 1-10
Ebersbach et al. (2008)	Children in K-3rd grades	Symbolic	1-100 for K-2nd, 1-1,000 for	Log, linear, 2-phase linear	2-phase linear for all grades
Opfer & Thompson (2008)	Children aged 7 years	Symbolic	0-1,000	Log, linear	Log before and linear after training
Thompson & Opfer (2008)	Children aged 7-9 years	Symbolic	0-1,000	Log, linear	Log for young and linear for older children in pretest; more linear in posttest after training
Berteletti et al. (2010)	Children aged 3.5-6.5 years	Symbolic	0-100, 1-10, 1-20	Log, linear	Log for young and linear for older children in 1-10 and 1-20; log for all in 0-100
Heine et al. (2010)	Children in 1st-3rd grades	Symbolic	0-100	Log, linear	Log for 1st graders; linear for 2nd-3rd graders
Cohen & Blanc-Goldhammer	Adults	Symbolic	0-26, 0-1 (stimuli < 26)	Linear, 1-2CPMs, 1-multiple	One of CPMs in 0-26 and one of SPMs in 0-1
Opfer et al. (2011)	Children in K-4th grades	Symbolic	0-100 for K-2nd, 0-1,000 for 1st-4th	Log, linear, 1-2CPMs	Log for younger children on both scales, linear for older children on both scales

Anobile et al. (2012)	Adults	Non-symbolic	1-10, 1-30, 1-100 with or without an additional attentional	MLLM, Central-tendency Bayesian model	Both; more log compression with the attentionally-demanding task
Ashcraft & Moore (2012)	Children in 1st-5th grades and adults	Symbolic	0-100 for 1st-2nd, and 0-100 and 0-1,000 for	Log, linear, exponential	With grade, a linear model fits better
Opfer & Martens (2012)	Children aged 6-17 years and adults with or without Williams syndrome	Symbolic	0-1,000	Log, linear	Log for children and adults with WS (even after training) and young children without WS; linear for older children and adults without WS
Sasanguie et al. (2012)	Children in K-2nd, and 6th grades	Symbolic & Non-symbolic	0-10 for K-1st, 0-100 for 1st-2nd,	Log, linear	Linear for all in 0-10, and more log for young children and linear for older children in 0-100 regardless of number formats
White & Szucs (2012)	Children in 1st-3rd grades	Symbolic	0-20	Log, linear, 1-2CPMs	No model differences for 1st graders; linear for 2nd and 3rd graders
Hoffmann et al. (2013)	Children in K	Symbolic	0-20	Log, linear	Log for first-term children; no model differences for second-term
Cicchini et al. (2014)	Adults	Non-symbolic	0-100	MLLM, Bayesian integration model	Bayesian integration model especially when attention was deprived
Cohen & Sarnecka (2014)	Children aged 3-8 years	Symbolic	0-20, 0-1 (stimuli < 20)	Log, linear, SBCM, 1-2CPMs, 1-multiple SPMs	SBCM or CPMs in bounded tasks and one of SPMs in unbounded tasks
Hurst et al. (2014)	Adults	Symbolic	0-1,258, 2,000-3,000,	Log, linear	linear in standard endpoint conditions (0-1,258, 2,000-3,000) and log in the non-standard
Laski & Yu (2014)	Chinese and Chinese American children in K and 2nd	Symbolic	0-100, 0-1,000	Log, linear	More log for Chinese American K, and linear for Chinese K and all 2nd graders in 0-100; log for Chinese and Chinese American K, and linear for Chinese 2nd graders

Link et al. (2014)	Children in 1st-4th grades and adults	Symbolic	0-10 for 1st, 0-20 for 2nd, 0-100 for 3rd, 0-1000 for 4th graders, 0-10,000 for adults,	Linear, 0-2CPMs, 1-multiple SPMs	Linear for 1st-2nd graders, and one of CPMs for 3rd-4th graders and adults when tasks were bounded; 1SPM*** when tasks were unbounded
Friso-van den Bos et al. (2015)	Children aged 5-8 years	Symbolic	1-100	Log, linear 0-2CPMs	0CPM for 5- to 6-year-olds, and 1CPM for 7- to 8-year-olds
Sella et al. (2015)	Children in K, 1st, 3rd grades	Symbolic & Non-symbolic	0-100	Log, linear	More log for K and 1st graders than for 3rd graders regardless of number formats
Opfer et al. (2016)	Children aged 6-8 years	Symbolic	0-1,000	Log vs. 2CPM, MLLM vs. MCPM	2CPM with but log without midpoint-anchoring instruction; MLLM for all conditions
Kim & Opfer (2017)	Children in K-2nd grades	Symbolic	0-30 for K, 0-100 for 1st, 0-1,000 for 2nd graders, and 0-1 (stimuli < 30 for K, <	MLLM, MCPM1, MCPM2, MSPM	MLLM
Reinert et al. (2017)	Adults	Symbolic	0-50, and variable unit (0-1 to 0-10;	Linear, 1-2CPMs, SPMs	One of CPMs in bounded tasks and 1SPM in unbounded tasks
Chesney & Matthews (2018)	Adults	Non-symbolic	1-300	Log, linear, 0-1CPMs	Linear and 0-1CPMs fit equally well
Cohen et al. (2018)	Adults	Symbolic	0-22, 0-1 (stimuli < 22)	0-2CPMs, MCPM, 1-multiple SPMs	One of CPMs in bounded tasks and 1SPM in unbounded tasks
Kim & Opfer (2018)	Children aged 5-6 years and adults	Symbolic & Non-symbolic	0-30	MLLM, dynamic MLLM, Bayesian integration model	MLLM or dynamic MLLM

Opfer et al. (2019)	Children aged 3-5 years	Symbolic	0-20	MLLM, MCPM1, MCPM2	MLLM
Yuan et al. (2019)	Children aged 4-6 years	Symbolic & Non-symbolic	0-1,000	Log, linear, MLLM, 1-2CPMs	Log or MLLM
Zax, Slusser, & Barth (2019)	Children aged 6-8 years	Symbolic	0-100	0-2CPMs, MCPM	MCPM for median estimates
Devries, Kuhn, & Gebhardt	Children in 2nd grade	Symbolic	0-20, 0-100	Log, linear, 1CPM	Linear in 0-20 and log in 0-100
Jung et al. (2020)	Children in 5th-7th grade	Symbolic	0-10,000, 0-1 (stimuli <	Linear, 1-2CPMs, 1-multiple	Linear
Kim & Opfer (2020)	Children aged 4-12 years	Symbolic	0-538, 0-1 (stimuli < 58, 132, or	1SPM, MLLM	MLLM in both bounded and unbounded tasks

Note. CPM = cyclic power model; MCPM = mixed cyclic power model; SPM = scallop power model; MSPM = mixed scallop power model; MLLM = mixed log-liner model.