

# UC Irvine

## UC Irvine Previously Published Works

### Title

Children's number-line estimation shows development of measurement skills (not number representations).

### Permalink

<https://escholarship.org/uc/item/8nf2s2hx>

### Journal

Developmental Psychology, 50(6)

### ISSN

1939-0599 0012-1649

### Authors

Cohen, Dale J  
Sarnecka, Barbara W

### Publication Date

2014

### DOI

10.1037/a0035901

Peer reviewed

Running Head: CHILDREN'S NUMBER-LINE ESTIMATION

Children's Number-Line Estimation Shows Development of Measurement Skills (Not Number  
Representations)

Word Count: 8,336

This work was supported by NIH grant RO1HD047796 to the first author, and by NSF grant DRL 0953521 to the second author. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of the National Institutes of Health or the National Science Foundation. We thank the children, families and staff of UCI Child Care Services for their participation in this study; thanks also to research assistants Annie Ditta and Ryan Galli for help with data collection.

## Abstract

Children's understanding of numbers is often assessed using a number-line task, where the child is shown a line labeled with 0 at one end, and a higher number (e.g., 100) at the other end. The child is then asked where on the line some intermediate number (e.g., 70) should go.

Performance on this task changes predictably during childhood, and this has often been interpreted as evidence of a change in the child's psychological representation of integer quantities. The present paper presents theoretical and empirical evidence that the change in number-line performance actually reflects the development of measurement skills used in the task. We compare two versions of the number-line task: the bounded version used in the literature and a new, unbounded version. Results indicate that it is only children's performance on the bounded task (which requires subtraction or division) that changes markedly with age. In contrast, children's performance on the unbounded task (which requires only addition) remains fairly constant as they get older. Thus, developmental changes in performance on the traditional, bounded number-line task likely reflect the growth of task-specific measurement skills, rather than changes in the child's understanding of numerical quantities.

## Children's Number-Line Estimation Shows Development of Measurement Skills (Not Number Representations)

A central question in the development of numerical cognition is how children's understanding of numerical magnitudes changes with age. One of the tasks most commonly used to study this is the number-line task (e.g., Booth & Siegler, 2006, 2008; Geary, 2004; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Geary, Hoard, Nugent, & Byrd-Craven, 2008; Opfer & Siegler, 2004, 2007; Siegler & Booth, 2004, Siegler & Opfer, 2003). Here we present mathematical, theoretical, and empirical evidence that the developmental changes in performance on the traditional number-line task reflect improvements in the mathematical skill needed to scale numbers to the line (termed *mensuration skills*) rather than changes in children's mental representations of numerical quantities. We argue that children's understanding of numbers should be assessed using a more valid, less mathematically demanding version of the number-line task: the unbounded number-line task. Data from this task provide strong evidence that children represent numerical quantities in the same way adults do.

### **Developmental Changes in Number-Line Estimation**

In the number-line task, participants must convert a numeral into a line length, or vice versa. Most commonly, the participant is given a horizontal line with two labeled endpoints (e.g. 0-100; see Figure 1). On each trial, the participant is given a target number and must make a mark on the line where the number should go. This type of task is a *bounded* number-line task, because all the target values fall between the upper and lower bounds (e.g., between 0 and 100).

Using the bounded number-line task, Siegler and Opfer (2003) showed that younger children produced a negatively accelerating error pattern (forming a logarithmic-looking curve), whereas older children and adults produced more linear patterns. To explain the data, the authors

argued that within each person, innate, nonverbal numerical cognition includes two separate and distinct systems for representing numerical magnitudes. The earlier, ‘logarithmic’ pattern of responding on the number-line task is taken as evidence that the participant is using the ‘logarithmic’ system; the later, ‘linear’ pattern is taken as evidence of a separate that the subject is using a ‘linear’ system. The shift in performance is said to reflect children’s increasing use of the linear system to solve the task (Siegler & Opfer, 2003; Siegler, Thompson & Opfer, 2009). This ‘logarithmic-to-linear shift’ has been found in participants of many different ages and mathematical abilities (e.g., Berteletti, et al., 2010; Booth & Siegler, 2006; Geary, et al., 2008; Laski & Siegler, 2007; Siegler & Booth, 2004).

Recently, it has been argued that ‘linear’ patterns of responses on the bounded number-line task only appear linear because researchers apply an inappropriate statistical model to the data (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011). Nonetheless, even when researchers analyze the number-line data with more appropriate statistical models, (1) adults and children still perform quite differently (see Cohen & Blanc-Goldhammer, 2011 for adults; Barth & Paladino 2011 for children); (2) there remains a distinct developmental trend in performance, with children performing better as they get older (Slusser, Santiago, & Barth, 2012); and (3) the youngest children’s performance still forms a logarithmic-like curve (Barth & Paladino, 2011). We argue below that all three of these effects arise because children develop the mathematical skills needed to scale the numbers to the line (i.e., their mensuration skills improve), not because they shift to a different system of nonverbal number representation. Below, we review the psychophysics of scaling in order to explain number-line responses from prior studies in terms of well-known psychological principles.

### **Mensuration: Scaling Numbers to Lines**

Number-line tasks are *cross-modality matching tasks* (e.g., Gescheider, 1988; Marks, 1974; Stevens, 1956) because they require participants to convert a numeral into a line length or vice versa. In a typical cross-modality matching task, a participant is given a standard stimulus in one modality, and a corresponding value in the second modality. This *Standard* links the scale of Mode A to that of Mode B. For example, the Standard may be a tone with an intensity of 80 DB and the corresponding value may be the number 100. The experimenter presents a series of probes (e.g., tones of different intensities) and asks the participant to scale these tones to that of the Standard. The participant might be instructed, “Listen to how loud this tone is. This loudness is assigned a value of 100. Now you’re going to hear another tone, and you have to say how loud it is. So, if it seems half as loud as the first tone, you say 50. If it seems three times louder than the first tone, you say 300.” Perceived loudness is thus scaled to perceived numerical quantities through the Standard.

Cross-modality matching tasks have been studied for decades (e.g., Algom & Marks, 1984, 1989; Cohen & Lecci, 2001; Marks, 1974; Stevens, 1956). Typically, bias<sup>1</sup> in the perception of the stimulus being measured (e.g., bias in the perception of loudness) is well described by the exponent in a power function (see online supplement). Exponents less than 1 mean that participants are more sensitive to changes at low stimulus values (e.g., soft tones) than at high stimulus values (e.g., loud tones). This can produce a pattern of data that looks similar to a logarithmic function. Exponents greater than 1 mean that participants are more sensitive to changes at high stimulus values than at low stimulus values. This can produce a pattern of data that looks similar to an exponential function.

The traditional version of the number-line task described above is a *bounded* number-line

task, because participants' answers cannot be less than the lower bound (usually 0) or more than the upper bound (e.g., 100). However, Cohen and Blanc-Goldhammer (2011) modified the task so that participants' responses were not restricted on the upper end. The authors called this an *unbounded* number-line task. In the unbounded task, the participant is given a horizontal line denoting the length of a single unit. (See Figure 1). The participant then has to indicate the position that a target number should occupy, somewhere to the right of the point labeled "1."

The estimated stimulus bias in a cross-modality matching task is subject to the influence of many factors, one of which is the range of responses. The bounded task restricts the range of possible responses by requiring them to fall between the upper and lower bounds. In contrast, the unbounded task allows the participant to make any response greater than one.

Di Lollo and Kirkham (1969) studied the influence of bounds on tasks like the bounded number-line task. Participants were shown a grid of black and white squares, and were asked to estimate the proportion of black squares. Possible responses range from 0% black to 100% black, just as possible responses on the bounded number-line task range from one end of the line to the other. Di Lollo and Kirkham reported that at lower target probes (few black squares), subjects tended to produce estimates that were much too large. When probes were larger (a high percentage of black squares), subjects revised their estimates downward to avoid going over 100%. This occurred in tasks "where the stimuli [were] unidimensional and where familiarity with mensuration techniques [was] low" (p. 525, Ross & Di Lollo, 1971).

In other words, this pattern of data (overestimation of the lower probe values and downward-revision of estimates for the upper probe values) happens whenever (a) there is only one dimension on which to judge the stimuli, *and* (b) participants are not familiar with the mathematics of scaling numbers to that dimension. Because the bounded number-line task is

unidimensional, subjects with low mensuration skills (e.g., children) are expected to produce a logarithmic-like pattern in the data (i.e., overestimation of low values and underestimation of large values). Because the bounded and unbounded number-line tasks require different mensuration techniques (i.e., different strategies for scaling numbers to line lengths), the influence of mensuration skill on this task is testable within participants.

In the bounded task, the simplest way to estimate the line length for any integer is to recursively estimate the target number's distance from the lower and upper bounds until the two distances appear consistent. For example, if the participant wants to place the number 70 on a line from 0 to 100, he or she chooses a point on the line, and then looks back and forth between that point and the ends of the line, adjusting the position until the distances between the point and the lower and upper bounds appear to be 70 and 30, respectively. This is the strategy that Cohen and others have argued is used by adults and older children on the bounded number-line task (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011).

The strategy was first described in detail by Spence (1990) as the way that people estimate proportions in a similar cross-modal task. Spence asked participants to estimate proportions between 0 and 1, and found a signature ogival (i.e., S-shaped) error pattern. Spence hypothesized that participants estimate the magnitude of both proportions (i.e., the distances to the right and left of the tick mark on the number line) and revise their estimates until these quantities sum to one. Based on this supposition, Spence derived a model of proportion estimation based on Stevens' Power Law (termed the Power Model, Hollands & Dyre, 2000; Hollands, Tanaka, & Dyre, 2002), which states that the psychological representation of proportion,  $\psi_p$ , is a function of the presented proportion,  $\Theta_p$ , and its inverse,  $1 - \Theta_p$ , summarized by the equation,

$$\psi_p = \Theta_p^\beta / [(\Theta_p^\beta) + (1 - \Theta_p)^\beta]. \quad (1)$$



When participants estimate proportions this way, their data form a signature ogive (an S-shaped curve around the accuracy line) or inverse ogive. The exponential term here,  $\beta$ , is the same exponent as that in Stevens' Power Law. Therefore, a  $\beta=1$  indicates accurate responding;  $\beta<1$  indicates a negatively accelerating bias; and  $\beta>1$  indicates a positively accelerating bias. Spence's Power Model fits estimated proportion data well (e.g., Begg, 1974; Brooke & Macrae, 1977; Shuford, 1961; Varey, Mellers, & Birnbaum, 1990).

More recently, Hollands and Dyre (2000) developed a more generalized Power Model (termed the *Cyclic Power Model* - CPM) to accommodate data with multiple ogival cycles, which occur when participants use one or more additional reference points (such as the halfway point of the line).

When viewed closely, number-line data exhibit the signature ogival error pattern characteristic of Hollands and Dyre's (2000) CPM and are well fit by the model (e.g., Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011). The fact that the CPM fits better than either linear or logarithmic models indicates that older children and adults are using the strategy described above. Nevertheless, young children consistently produce negatively accelerating (i.e., logarithmic-looking) patterns of responses that are not well fit by the CPM (Barth & Paladino, 2011). Furthermore, older children whose bounded number-line performance is well described by the CPM for familiar (to them) number ranges (i.e., 1-20), still produce negatively accelerating patterns of responses for higher number ranges, which are less familiar to them (Slusser et al., 2012). The source of this negatively accelerating pattern has not yet been identified in the literature.

We argue that the negatively accelerating pattern is the result of poor mensuration skills. All successful strategies for the bounded number-line task require the participant to do some form of

subtraction or division. This is why young children (who have mastered neither subtraction nor division) are unable to perform the task accurately. In the terms used by Ross and Di Lollo (1971), young children have low familiarity with the required mensuration technique.

So, how do young children manage the task? According to Ross and Di Lollo (1971), they start by implicitly assigning a biased scale that overestimates the relation between line length and numerical quantity. For smaller target values, this results in overestimation. But for larger target values, these overestimates would soon stretch beyond the upper boundary of the line. Seeing this, the children shift their estimates downward to fit within the bounds. The resulting pattern of data shows overestimation of numbers at the low end, and a decelerating function for numbers at the high end—a pattern that can easily be confused for a logarithmic function.

To distinguish between a true logarithmic function and this ‘poor-mensuration’ function requires a formal model of the poor-mensuration function. We developed the *Subtraction Bias Cyclic Model* (SBCM) to formally model poor mensuration in the bounded number-line task. The SBCM is based on Holland and Dyre’s (2000) single-cycle Cyclic Power Model (CPM), but is modified to incorporate a subtraction bias. As described above, Spence (1990) proposed that participants scale a line length (or similar representation) to a proportion by estimating the distances from the right and left sides of the line. To do this, the participant must be able to subtract the target number from the value of the upper bound (e.g., to place the number 13 on a line stretching from 0 to 20, children must be able to subtract 13 from 20). If children’s estimates are poor because they have difficulty with this ‘subtract-and-compare’ process, a distinct measurement bias will be introduced. We modeled this bias with the following formula:

$$\psi_1 = \Theta_1^\beta / [(\Theta_1^\beta) + ((U - \Theta_1)^s)^\beta], \quad (2)$$

where  $\Theta_1$  is the target integer, the exponent  $s$  represents the bias associated with the subtraction, and  $U$  represents the value of the upper bound. Equation 2 is identical to Equation 1, with the addition of an exponent  $s$ , which captures the potential bias in subtraction. This function captures the pattern of overestimation for low target values and a shift downward for high target values. This model allows us to test whether the data are better fit by a true logarithmic function or by the SBCM. Furthermore, if the data are best fit by the SBCM, we can separate the perceptual<sup>2</sup> and subtraction biases, giving us a clearer idea of what children actually know about the numbers.

The true logarithmic pattern and the pattern resulting from poor mensuration techniques also differ in residuals around the prediction curve. Poor mensuration produces a pattern in which the upper bound interferes with the participant's intended responses. Thus, responses should be close to or at the upper bound even for numbers somewhat below the largest target number. These responses at or near the upper bound should also occur over several target numbers. The net effect is that the participant's responses cluster near the upper bound, resulting in heteroskedasticity when the data are fit with a logarithmic function. In contrast, a true logarithmic pattern is homoscedastic across the entire function.

In contrast to the bounded number-line task, scaling numbers to line lengths in the unbounded number-line task requires only addition. The participant merely needs to repeat the single line length, as many times as the target number. (For example, to estimate the position of the number 5, the participant must add five single-unit lengths together, or multiply the single unit by 5.) Cohen and Blanc-Goldhammer (2011) developed the Scalloped Power Model (SPM) to describe adults' performance in the unbounded number-line task. The Scalloped Power Model is a simple variant of Stevens' Power Law, describing the scalloped pattern that is created when

participants reach large numbers by repeatedly counting to a smaller number (termed the *working window*, e.g., to get to 20, viewers may estimate a line length of 5, and then repeat that length four times).

Because the implicit addition needed for the unbounded task is less mathematically sophisticated than the implicit subtraction needed for the bounded task, children should perform better on the unbounded task at a younger age. Indeed, because of the reduced mathematical constraints on the task, any child who knows the order of the numbers, and knows that numbers coming later in the list represent greater quantities, can perform quite well on the unbounded task. A failure of mensuration on the unbounded task would indicate a failure to understand this basic ordinality of numbers, resulting in random performance.

The present study examines young children's integer estimation using both the bounded and unbounded number-line tasks. For the bounded task, we compare the Cyclic Power Model (CPM), Subtraction Bias Cyclic Model (SBCM), linear and logarithmic models as predictors of children's performance. For the unbounded task, we compare the Scallop Power Model (SPM), linear and logarithmic models. By comparing data from the two different tasks, we see how the demands of each task influence children's responses. We hypothesize that younger children will complete the unbounded number-line task more accurately than the bounded number-line task, because the former requires lower mensuration skills than the latter.

## **Method**

### **Participants**

Sixty-two children (25 girls, 38 boys,  $M_{\text{age}} = 5$  years, 11 months, age range 3;6–8;0) were recruited from preschool and after-school programs located on a campus of the University of California. Children received a small toy (e.g., a stuffed animal or a ball) worth approximately

\$5 when they signed up to participate. No compensation was given at the time of testing.

Families were not asked about their income, race or ethnicity. Because the parents served by these centers are almost all university faculty or graduate students, parent education levels were high.

### **Apparatus and Stimuli**

All stimuli were presented on a 17-inch MacBook Pro laptop computer. The resolution of the monitor was 1920 by 1200 pixels. Participants sat approximately 30 inches away from the screen.

On each trial, participants were presented with a number line and target number (see Figure 1). The number line was centered on the y-axis of the screen. The target number was placed half an inch below the left boundary of the number line (in the position marked “X” on Fig. 1). The number line was constructed from 3-pixel-thick red lines. A 20-pixel-high vertical mark showed the start of the number line. This left boundary was labeled with the number "0." A similar vertical mark indicated the right end of the number line and was labeled with the number "20" in the bounded task, and with the number “1” in the unbounded task. The two vertical marks were connected at the bottom by a red horizontal line (i.e., the number line). The target numbers ranged from 2 to 19 and were chosen randomly from a uniform distribution from trial to trial.

To prevent participants from using reference points external to the number line (e.g., the left edge of the monitor, center of the monitor, etc.), we varied the location and physical length of the number line. For each trial the number line was randomly placed between 100 and 200 pixels from the left side of the screen. The length of the bounded number line was randomly varied from 200 to 600 pixels, in 20-pixel steps. The length of the unbounded number line was

randomly varied from 10 to 30 pixels, in 1-pixel steps. This equated the unit size of the two number-line conditions.

To keep children interested in the task, the experiment was programmed with visual reinforcement. After every trial, a cartoon image appeared in a random place on the screen for 1.5 seconds. On a variable ratio schedule (about every 5 trials), a unique fractal image gradually appeared on the screen and then gradually disappeared. Finally, on a slightly longer variable ratio schedule (about every 8 trials) an animation played. This combination of reinforcements was effective at maintaining most children's interest in the task. The reinforcement was given regardless of the values input by the child.

### **Procedure**

Children were tested individually, in a quiet room at their child-care center. Each child completed two sessions: one for the bounded task, another for the unbounded task. Because the children's performance on the bounded number-line task is relatively well understood, we had all participants complete the unbounded number-line task first. This ensured that children's unbounded-task performance was not influenced by prior experience with the bounded task<sup>3</sup>.

***Unbounded number-line task.*** The experimenter introduced the task by saying, "This is a little number line, see? Here's *zero*, and here's *one*. And all the other numbers go after one, right? So they'll go over *here* . . . (experimenter traces an imaginary line back and forth on the screen, extending rightward from the 1) . . . In this game, a number will show up *there* (experimenter points to the location on the screen where the target number will appear), and you have to drag this little mark (experimenter demonstrates how to move the cursor) to where the number should go."

To move the mark, the child pressed and held the button on the track pad (some of the

younger children needed the experimenter to hold the button down for them, so they could just move the mark) and then dragged the cursor over the right boundary (labeled “1”). At this point, a gray mark appeared over this boundary (see Figure 1), and a horizontal line connecting the grey mark to this boundary appeared.

Children could freely move the gray mark, dragging it back and forth, releasing it and dragging it again, without submitting a response. When the child was satisfied with the placement of the mark, he or she pressed the space bar to submit the response. (If the experimenter was helping to hold the track pad button down, he or she waited to release the button until after the child had pressed the space bar.) Children’s reaction times and accuracy (to the pixel) were recorded. Each child was presented with three practice trials and 40 experimental trials. Each session typically lasted less than 20 minutes.

***Bounded number-line task.*** This task was the same as the unbounded task, except that the experimenter introduced it by saying, “This is a number line, see? Here’s *zero*, and over here is *twenty*. In this game, a number will show up *there* (experimenter points to the location on the screen where the target number will appear) and you have to drag this little mark (experimenter demonstrates how to move the cursor) to where the number should go.”

In the bounded task, the gray mark appeared when the child moved the cursor over the left boundary (labeled “0”). The child then moved the gray mark to the estimated target location, somewhere on the line between the lower and upper boundaries.

## Results

Below, we present analyses at both the group and individual levels. We first present an analysis of grouped data. This analysis is comparable to those in the literature that describe children’s performance as either logarithmic or linear, including the studies that posit a

logarithmic-to-linear shift (e.g., Barth & Paladino, 2011; Opfer & Siegler, 2007; Siegler & Opfer, 2003). We ran the group analysis to check that our bounded number-line data were comparable to data in the extant literature (i.e., revealing a logarithmic-looking pattern for young children and a more linear-looking, CPM pattern for older children). If the developmental trend in bounded number-line performance is, as we hypothesize, a result of changes in mensuration skills, then the same developmental trend should not be found on the unbounded number-line task.

All children's data were included in the group analyses, but only data from children who completed at least 20 trials were included in the individual analyses. Three children did not complete 20 trials in either of the two sessions. Another 15 children (5 in the bounded number line task; 10 children in the unbounded number line task) completed more than 20 trials, but did not complete the full session. Finally, six children dropped out between the first (unbounded task) and second (bounded task) sessions.

### **Group-level Analyses**

Data from all the children (including those who failed to complete one or both sessions) were included in the group analysis. Prior to analyzing the group data, we removed individual data points (i.e., individual observations) that were under 10 percent or over 600 percent of the target number (i.e., outliers)<sup>4</sup>. These constraints eliminated 7.6 percent of the data. We conducted separate analyses of the data from younger children (ages 3-6 years, N=48) and older children (ages 7-8 years, N=14). We separated children into these age groups because past research has shown a shift in response patterns between the ages of six and seven for familiar ranges in number line tasks (e.g., Barth & Paladino, 2011; Opfer & Siegler, 2007; Slusser et al., 2012). We calculated the mean estimate of each target number from all trials for each condition



(bounded and unbounded) by age group.

**Bounded Number Line.** Using generalized non-linear least squares (gnls) methods, we compared five models: (1) the linear Cyclic Power Model (CPM) with two reference points (the bounds of the number line); (2) the CPM with three reference points (the bounds and the midpoint of the number line); (3) the Subtraction Bias Cyclic Model (SBCM) with two reference points (the bounds of the number line); (4) the linear model; and (5) the logarithmic model. We included the CPMs because prior research has shown that participants use the subtraction strategy to complete the bounded number line task (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011). We included the linear and logarithmic models because these have been the models used most often in the number-line literature (e.g., Siegler & Opfer, 2003, Siegler, et al., 2009). Finally, we included the SBCM in order to test our hypothesis that poor mensuration skills account for the negatively accelerating pattern of data produced by younger children on the bounded task. Because the models have different numbers of parameters, we compare the relative fit of models using the BIC. The BIC is a measure of model fit that corrects for the number of parameters present (a lower BIC indicates a better fit). The model that had the lowest BIC was judged to fit the data best.

Data from the younger children as a group were best fit by the logarithmic model (see Figure 2), indicating that these children did not use the subtraction/division strategy captured by the CPM. The logarithmic model is a negatively accelerating function (slope = 4.78, SE = 0.25; Intercept = 1.5, SE = 0.57). Overall, the average error variance was rather low (about 4.8). There was a significant negative linear relation (slope = -0.08, SE = 0.03) between target number and -SD,  $F(1,16)=6.6$ ,  $p=.02$ ,  $r^2=.29$ . This negative relation indicates that the participants' error decreases as target number increases. This is exactly the pattern one would predict if upper

bound interfered with the participants' responses. Figure 2 shows participants' responses for each target number. The figure reveals that participants' estimates hit the upper bound for all target numbers 3 and higher. This pattern strongly suggests that the logarithmic pattern reported in the literature is an artifact of the bounded nature of the task, rather than an indicator of how participants' underlying quantity representations are organized. The individual analyses (below) revealed more about this pattern.

Data from the older children were best fit by the two-reference-point CPM (see Figure 2), indicating that these children did employ a subtraction/division strategy, and used only the two, labeled endpoints of the line for reference. The estimated bias was negatively accelerating ( $\beta = 0.76$ ,  $SE = 0.05$ ). The bias was significantly less than 1 ( $T > 4.0$ ). Figure 2 plots participants' responses as a function of target number. Overall, the average error variance was rather low (about 2.7). There is no linear relation between target number and SD,  $F(1,16)=1.2$ , ns.

Note that if only the logarithmic and linear models are compared, the linear model does fit these data better than the logarithmic one. In other words, our data replicate the well-known 'logarithmic-to-linear shift.' However as noted above, the logarithmic function results at least in part because participants' responses are inhibited by the upper bound, and the 'linear' pattern is not actually linear—it is better fit by the CPM (replicating the findings of Barth & Paladino, 2011, and Cohen & Blanc-Goldhammer, 2011)

**Unbounded Number Line.** Using generalized non-linear least squares methods, we assessed the fit of the linear model, the logarithmic model, and three Scalloped Power Models (SPMs). We included the SPMs because Cohen and Blanc-Goldhammer (2011) showed that participants use an addition strategy to complete the unbounded number line task. We included the linear and logarithmic models because these have been the models used most often in the

number-line literature (e.g., Siegler & Opfer, 2003; Siegler, et al., 2009.). The model that had the lowest BIC was judged to fit the data best.

Data from the younger children were best fit by the Multi-Scalloped Model (see Figure 3), indicating that these children used an addition strategy. The estimated bias was positively accelerating ( $\beta = 1.52$ ,  $SE = 0.01$ ). The bias was significantly greater than 1 ( $T > 40.0$ ) and resulted in substantial overestimation of the quantities associated with the target integers. The estimated size of the working window of numbers ( $d$ ) was 7.0 ( $SE=0.29$ ), meaning that these children estimated the distance to about 7 before starting from one again. Figure 3 plots participants' responses as a function of target number. Overall the average error variance was rather large (about 14.5). Variance in the data was scalar, meaning that there was a significant linear relation between target number and SD,  $F(1,16)=813$ ,  $p<.0001$ ,  $r^2=.98$ .

Data from the older children were also best fit by the Multi-Scalloped Model (see Figure 3), again indicating that these children too used an addition strategy. The estimated bias was positively accelerating ( $\beta = 1.42$ ,  $SE = 0.01$ ). The bias was significantly greater than 1 ( $T > 40.0$ ) and resulted in substantial overestimation of the quantities associated with the target integers. The older children's bias was significantly smaller than the younger children's bias ( $t > 5.0$ ), meaning that the older children estimated the positions of numbers more accurately than the younger children did. The estimated size of the working window of numbers ( $d$ ) was 6.3 ( $SE=0.63$ ), meaning that the older children estimated the distance to about 6 or 7 before starting from one again. Figure 3 plots participants' responses as a function of target number. Overall the average error variance was rather large (about 13.0). There was a significant linear relation between target number and SD,  $F(1,16)=134$ ,  $p<.0001$ ,  $r^2=.89$ , meaning that variance was scalar.

In sum, children from both age groups used the same strategy (addition) to complete the

unbounded number-line task, and they showed similar biases. However, older children produced more accurate estimates than younger children did.

### **Individual-level Analyses**

To investigate whether these patterns held true for individual children, we conducted the same model fits on individual participants' data as we had on the grouped data. Only data from children who completed at least 20 trials in a session were analyzed, because this was the minimum number of trials required to get a robust estimate of performance from an individual child. Below we summarize the results by age group: 4-year-olds ( $n=9$ ), 5-year olds ( $n=27$ ), 6-year-olds ( $n=11$ ), and 7-year-olds ( $n=10$ ). Because there was only one 3-year-old and one 8-year-old, we did not include them in these reports. The online supplement contains examples of data from individual children in each age group that were best fit by each of the assessed models.

**Bounded Number Line.** First, the overall fit of all the models was better for older children ( $r=0.87$ ,  $p<.001$ ; see Figure 4). Four-year-olds averaged an  $r^2=0.17$ , whereas seven-year-olds averaged an  $r^2=0.75$ . This indicates that younger participants produced noisier data than older participants, and that the model estimates were therefore more reliable for older children.

Figure 5 presents the proportion of children best fit by each model for each age group. Note that the results contradict some well-known findings from the extant literature. Developmental change on the bounded task has been described as a logarithmic-to-linear shift (e.g., Siegler et al., 2009). In contrast, we found that the proportion of children best fit by the linear model actually *decreased* with age. Moreover, the linear model fit the *fewest* children overall. Conversely, the proportion of children best fit by the logarithmic model actually *increased* with age.

These findings are undoubtedly driven by the inclusion of variations of the Cyclic Power

Model in the analysis. When it comes to describing the performance of children who are skilled at the mensuration techniques for the bounded task, the standard Cyclic Power Model (CPM) usually fits better than a simple linear function. Thus, many children whose performance would be described as ‘linear’ in the earlier studies are actually better described (and thus are described here) by the CPM.

Similarly, the Subtraction Bias Cyclic Model (SBCM) does a good job of describing the performance of children who are unskilled at the mensuration techniques needed for the bounded number-line task. In previous studies, these data would have been classified as ‘logarithmic,’ because a log function fits them better than a linear function.

In the present analysis, the proportion of children best fit by the standard CPM was relatively high across age groups. And although the fit of the CPM to the youngest children’s data was not as good as the fit for the older children, the CPM still fit better than the other models. The proportion of children best fit by the SBCM decreased with age. This is to be expected, because the SBCM describes the performance of children who have poor mensuration skills. As mensuration skills improve with age, the proportion of children who are best fit by this model gets smaller.

Figure 6 shows the estimated bias by age group for the children fit by the CPM. With the CPM model, a bias estimate of 1 indicates accurate performance. As can be seen in the figure, the estimates of children’s bias get closer to one (reflecting more accurate performance) as the children get older.

**Unbounded Number Line.** The unbounded tasks reveal a very different pattern than the bounded task. First, the overall fits of all the models were constant across ages ( $r=0.02$ , ns; see Figure 4). The average  $r^2=0.5$ . In other words, in the unbounded task, younger participants did

*not* produce noisier data than older participants; the reliability of the model estimates was equally good for both age groups. This is remarkable given that there is more inherent variability in the unbounded number-line data than in the bounded number-line data, due to scalar variance in the approximate number system. Despite this inherent variability, our models fit well to unbounded number-line data from individual children as young as four years old. Figure 4 reveals that  $r^2$  varied as a function of model fit. Specifically, of the best-fit models, the Log model accounted for the least variance and the SPM accounted for the most variance.

Figure 5 shows the proportion of children in each age group who were best fit by each model. Just as in the bounded task, the linear model failed to explain much. In fact, for all age groups except six-year-olds, the linear model explained the *fewest* children's data. The SPM model, on the other hand, explained the *most* children's data for all age groups except six-year-olds. Finally, unlike in the bounded task, the proportion of children best fit by the Log model decreased with age. This is important, because it shows that the logarithmic pattern of performance seen in the older children on the bounded task is not a reflection of their underlying integer representations (or else it would show up on the unbounded task too); the pattern is caused by the demands of the bounded task itself.

Figure 6 shows the results by age group for the children fit by the SPM. In contrast to the bounded task, estimation performance on the unbounded task remains constant as a function of age ( $M=1.47$ ). These results suggest that even the youngest children have the mensuration skills to complete the unbounded number-line task.

**Within-Subjects Comparison.** Forty-eight of our participants completed at least 20 trials in both the bounded and unbounded number-line tasks. This allowed us to look at within-child performance across tasks. We found that even for the same child, performance differed markedly

on the two tasks. Of the 48 children, only four produced a logarithmic pattern in both tasks; no child produced a linear pattern in both tasks; and 10 children who produced a CPM pattern in the bounded task produced a parallel SPM pattern in the unbounded task.

To explore these differences, we ordered the models in terms of the sophistication required (of the child) to implement each strategy: SBCM=0; Log=1; CPM=2; SPM=2; Linear=3. The SBCM represents the least sophisticated strategy for scaling line length to quantity; the linear model represents the most sophisticated strategy; the log and the power models fall in between. Such an ordering is not perfect, but it gives us a way of asking whether one task requires more sophisticated scaling than the other. And indeed, a paired-sample t-test reveals that children demonstrated more sophisticated scaling in the unbounded task than the bounded task (mean change = 0.67, SD=1.1,  $t(47)=4.3$ ,  $p<0.001$ ). This means that many children who failed at scaling the bounded number line still succeeded at scaling the unbounded number line.

The responses of the 4-year-olds demonstrate this effect. There were six 4-year-olds who completed at least 20 trials in both tasks. Figure 7 presents the data from each of these children in the bounded and unbounded number-line tasks. Note that even children who produced very noisy and non-ordinal responses in the bounded number-line task produced ordinally correct and often precise responses in the unbounded number-line task. This reflects the fact that the unbounded task requires less mathematical sophistication than the bounded task.

### **Discussion**

We propose that age-related changes in bounded number-line performance (including the purported ‘logarithmic-to-linear shift’) reflect changes in the participants’ mensuration skills, not their integer representations.

The bounded and unbounded number-line tasks are both cross-modal matching tasks in

which the participant must equate line length and quantity. But this is more difficult to do in the bounded task (where subtraction or division is required) than in the unbounded task (where addition or multiplication is required). We propose that the shift in performance long seen on the bounded task (logarithmic-like performance in less skilled participants; linear-like performance in more skilled participants) is evidence not of a change in integer representations, but of improvements in participants' ability to scale numbers to line lengths using subtraction or division.

We find evidence for this when we look closely at the data. Participants' use of subtraction or division on the bounded task produces an ogival (i.e., *S*-shaped) pattern of data—a pattern that is well described by the Cyclic Power Model, or CPM. (Hollands & Dyre, 2000). The majority of participants in our bounded number-line task were well fit either by the CPM (showing that they used subtraction or division to complete the task successfully), or by a variation of the CPM that we developed for this study, called the Subtraction Bias Cyclic Model, or SBCM. The SBCM is a model of how participants perform on the task when they *try* to use the appropriate mensuration techniques (involving subtraction or division) to scale the numbers to line lengths, but lack the mathematical skill to do so successfully.

When we model poor mensuration techniques on the bounded task using the SBCM, we find that this model describes the performance of younger children as a group better than the logarithmic model that has been used in the past. Furthermore, the fit of the SBCM gets worse as children get older-- as it should, because the SBCM is a model of *unskilled* performance.

For individual children whose bounded number-line data are best fit by the CPM or SBCM, our cross-sectional data suggest that estimation bias (the systematic inaccuracy of their number-to-line placement) changes with age. As has often been reported, very young children show a



negatively-accelerating pattern of bounded number-line estimation. But as children get older, their data begin to form a positively accelerating pattern (Barth & Paladino, 2011; Slusser et al., 2012). Similarly, Cohen and Blanc-Goldhammer (2011) showed that adults also display a positively accelerating pattern in the bounded number-line task.

A second line of evidence for our argument comes from another task that we gave to children—the unbounded number-line task. This task is similar to the bounded task in many ways (i.e., it requires children to scale numbers to line lengths), but does not require subtraction or division. We believe that the early, negatively accelerating pattern of data found on the bounded task reflects children’s poor subtraction and division skills. If this correct, then children should exhibit the more mature, positively accelerating pattern from the outset in the less-demanding, unbounded task.

Instead of giving children a Standard that is larger than any probe (i.e., a number line 20 units long, when the probe values range from 2 to 19), the unbounded number-line task gives children a Standard that is smaller than any probe: a number line 1 unit long. This forces participants to use addition or multiplication to convert between numbers and line lengths in the unbounded task. When participants use a strategy of addition or multiplication to do this, their data show up as a scalloped function that is well fit by the Scalloped Power Model, or SPM (Cohen & Blanc-Goldhammer, 2011). In the unbounded task, most of our participants’ data were best fit by the SPM, confirming that they did use addition or multiplication to complete this task.

For individual children whose unbounded number-line data were best fit by the SPM, the shape of their estimation bias (a positively or negatively accelerating curve) was unrelated to their age. In other words, there was no evidence of a shift (logarithmic-to-linear or otherwise), in how numbers are mentally represented. Such a shift, if it existed, should show up on any task

where children estimated numerical magnitudes, including the unbounded number-line task. But it did not. Instead, on the unbounded task, very young children showed the same positively accelerating (i.e., exponential-like) pattern of data as older children and adults (see Cohen & Blanc-Goldhammer, 2011 for data from adults).

We believe that young children perform like older children and adults in the unbounded task because they have the mathematical skill needed (addition or multiplication) to scale numbers on the unbounded number line. In this sense, data from the unbounded task data are a more accurate reflection of children's underlying quantity representations than are data from the bounded task.

Young children, older children and adults (see Cohen & Blanc-Goldhammer, 2011 for data from adults) all produce a positively accelerating pattern on the unbounded number-line task, but only older children and adults do so on the bounded task. Moreover, when we compared the same children's performance across tasks, we found that children performed better (i.e., estimated integer quantities more accurately) on the unbounded task than on the bounded task.

Given these facts, it seems clear that young children's early, negatively accelerating (i.e., logarithmic-like) pattern on the bounded task reflects their difficulty with the scaling required by the task itself, rather than the use of a different system of quantity representations. Our analysis of logarithmic performance found that about 40% of our participants were best fit by a true logarithmic function. However, this proportion was not related to age. In other words (in contrast to previous accounts), we found no evidence that young children are more likely to produce logarithmic patterns of data than older children.

There are several reasons why we might find only 40% of children producing a logarithmic pattern, whereas previous studies have found that virtually all young children do. First, previous

researchers (e.g., Booth & Siegler, 2006, 2008; Opfer & Siegler, 2007; Siegler & Booth, 2004) grouped participants' data, and this grouping itself results in a logarithmic pattern (see Figure 1). Second, unlike previous studies, we included the SBCM to capture the pattern caused by a biased subtraction strategy—a pattern that can easily be confused for a logarithmic function. Many children whose data would in previous studies have been called logarithmic were actually better fit by our SBCM. When these children's data are removed, the relation between age and logarithmic patterns (i.e., the trend for younger children to produce more-logarithmic patterns) disappears.

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, but we do not. In fact, the logarithmic pattern of responses was relatively rare in the unbounded task. Furthermore, analyses converged on the logarithmic model only for the noisiest data in the unbounded task, meaning that we cannot be as confident in describing these patterns as 'logarithmic' as we are in describing other patterns found in the data.

In spite of our hesitation to describe these patterns as truly logarithmic, they are intriguing in their own right. For example, the endpoints of the bounded number line may actually bias the errors of participants who are *not* using an implicit subtraction strategy, pushing them toward a negatively accelerating (i.e., logarithmic-like) function. Di Lollo and Kirkham (1969) showed that in a bounded cross-modal task, participants with poor mensuration skills exhibit a negatively accelerating pattern similar to the logarithmic pattern. (Di Lollo and Kirkham did not look specifically for the logarithmic pattern.) The authors hypothesized that this negatively accelerating pattern results when participants overestimate smaller values, but recognize that

they are running out of room at the high end. This recognition results in a shift to underestimation at larger values.

The data from our unbounded task give a picture of mental number representations that is consistent with the existing literature on nonverbal number representations in children and adults. (In the developmental literature, this is often called the *approximate number system* or ANS). But it is unwise to assume that responses on a number-line task provide a *direct* window onto participants' underlying quantity representations. Instead, the responses are a function of the participant's underlying quantity representation, their perception of the line length, their strategy for equating line length to quantity, and their mensuration skills.

In our data, most participants' mensuration skills were adequate to scale numbers to line lengths in the unbounded task. Therefore, if appropriate models are applied to the data, the estimated bias parameter ( $\beta$ ) will capture the biases resulting from (a) the underlying quantity representation and (b) the perception of line length. If we set aside the bias associated with line-length perception for a moment, we can interpret  $\beta$  as providing information about the psychological quantity representation.

Treating our data in this way, we find that children's perception of integer quantities shows scalar variance (a common signature of ANS representations across tasks and species, e.g., Cantlon, Cordes, Libertus, & Brannon, 2009; Cohen & Blanc-Goldhammer, 2011; Gallistel & Gelman, 2000; Gibbon, Church, & Meck, 1984). Furthermore, the positively accelerating bias indicates that the distance between the mean perceived quantity of integers increases as the quantity denoted by the integer increases. This combination of positively accelerating means and scalar variance is consistent with the signatures of the ANS as reported in the literature.

In sum, the bounded and unbounded number-line tasks are both cross-modal matching tasks.

However, the bounded number-line task requires more advanced mensuration skills. When these mensuration skills are lacking, participants produce a negatively accelerating pattern of data. In contrast, the unbounded number-line task requires less advanced mensuration skills, allowing it to reveal participants' quantity representations more accurately. This is especially important when the participants are children, who often lack the subtraction skills needed to do well on the bounded task.

In general, research using number-line estimation tasks must recognize that these, like all cross-modal matching tasks, are influenced by both stimulus and response biases. Analyses that ignore these biases and treat the tasks as direct measures of observers' quantity representations are flawed.

In particular, number-line estimation data (when properly analyzed) actually provide no support for the idea that children's representations of integer magnitude shift from a logarithmically-organized system to a linearly-organized system during development. Instead, children's performance-- on the bounded task only-- shifts from a negatively-accelerating pattern to a positively-accelerating pattern as children master the subtraction skills necessary to make the line lengths to the right and left of the tick mark sum to one.

When these same children are tested on the unbounded number-line task (where scaling numbers to lines requires only addition), they produce a positively accelerating pattern, just as older children and adults do. Together, data from both versions of the number-line task indicate that changes in performance come from children's growing ability to scale numbers to line lengths on the bounded task-- not a shift in the underlying quantity representations themselves.

## References

- Algom, D., & Marks, L. E. (1984). Individual-Differences in Loudness Processing and Loudness Scales. *Journal of Experimental Psychology-General*, *113*(4), 571-593.
- Algom, D., & Marks, L. E. (1989). Memory Psychophysics for Taste. *Bulletin of the Psychonomic Society*, *27*(3), 257-259.
- Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. [Research Support, Non-U.S. Gov't]. *Developmental science*, *14*(1), 125-135. doi: 10.1111/j.1467-7687.2010.00962.x
- Begg, I. (1974). Estimation of Word Frequency in Continuous and Discrete Tasks. *Journal of experimental psychology*, *102*(6), 1046-1052.
- Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. [Research Support, Non-U.S. Gov't]. *Developmental psychology*, *46*(2), 545-551. doi: 10.1037/a0017887
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. [Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S.]. *Dev Psychol*, *42*(1), 189-201. doi: 10.1037/0012-1649.41.6.189
- Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. [Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S.]. *Child development*, *79*(4), 1016-1031. doi: 10.1111/j.1467-8624.2008.01173.x
- Brooke, J. B., & Macrae, A. W. (1977). Error Patterns in Judgment and Production of Numerical Proportions. *Perception & psychophysics*, *21*(4), 336-340.
- Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on "Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene

- cultures". [Comment]. *Science*, 323(5910), 38; author reply 38. doi: 10.1126/science.1164773
- Chen, Q., & Verguts, T. (2010). Beyond the mental number line: A neural network model of number-space interactions. [Research Support, Non-U.S. Gov't]. *Cognitive psychology*, 60(3), 218-240. doi: 10.1016/j.cogpsych.2010.01.001
- Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number-line tasks. *Psychonomic bulletin & review*. doi: 10.3758/s13423-011-0059-z
- Cohen, D. J., & Lecci, L. (2001). Using magnitude estimation to investigate the perceptual components of signal detection theory. *Psychonomic bulletin & review*, 8(2), 284-293.
- Di Lollo, V., & Kirkham, R. (1969). Judgmental contrast effects in relation to range of stimulus values. *Journal of Experimental Psychology*, 81(3), 421.
- Gallistel, C. R., & Gelman, I. I. (2000). Non-verbal numerical cognition: from reals to integers. *Trends in cognitive sciences*, 4(2), 59-65.
- Geary, D. C. (2004). Mathematics and learning disabilities. [Research Support, U.S. Gov't, P.H.S.]. *Journal of Learning Disabilities*, 37(1), 4-15.
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. [Research Support, N.I.H., Extramural]. *Child development*, 78(4), 1343-1359. doi: 10.1111/j.1467-8624.2007.01069.x
- Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. [Research Support, N.I.H., Extramural]. *Developmental neuropsychology*, 33(3), 277-299. doi: 10.1080/87565640801982361

- Gescheider, G. A. (1988). Psychophysical scaling. [Review]. *Annual review of psychology*, 39, 169-200. doi: 10.1146/annurev.ps.39.020188.001125
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, 423, 52-77.
- Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological review*, 107(3), 500-524.
- Hollands, J. G., Tanaka, T., & Dyre, B. P. (2002). Understanding bias in proportion production. *Journal of Experimental Psychology-Human Perception and Performance*, 28(3), 563-574. doi: Doi 10.1037//0096-1523.28.3.563
- 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. [Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S.]. *Child development*, 78(6), 1723-1743. doi: 10.1111/j.1467-8624.2007.01087.x
- Marks, L. E. (1974). *Sensory processes: the new psychophysics*. New York,: Academic Press.
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. [Research Support, U.S. Gov't, Non-P.H.S.]. *Journal of experimental psychology. Animal behavior processes*, 9(3), 320-334.
- Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: a microgenetic analysis of conceptual change in basic biology. [Research Support, U.S. Gov't, P.H.S.]. *Cogn Psychol*, 49(4), 301-332. doi: 10.1016/j.cogpsych.2004.01.002
- Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cogn Psychol*, 55(3), 169-195. doi: 10.1016/j.cogpsych.2006.09.002
- Osborne, J. W. & Overbay,A. (2004). The power of outliers (and why researchers should always



- check for them). *Practical Assessment, Research & Evaluation*, 9(6).
- Ross, J., & Di Lollo, V. (1971). Judgment and response in magnitude estimation. *Psychological Review*, 78(6), 515.
- Shuford, E. H. (1961). Percentage Estimation of Proportion as a Function of Element Type, Exposure Time, and Task. *Journal of experimental psychology*, 61(5), 430-&.
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child development*, 75(2), 428-444. doi: 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. [Research Support, U.S. Gov't, P.H.S.]. *Psychological science*, 14(3), 237-243.
- Siegler, R.S., Thompson, C.A. & Opfer, J.E. (2009) The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain and Education*, 3, 143-150.
- Slusser, E. B., Santiago, R. T., & Barth, H. C. (2012). Developmental Change in Numerical Estimation. *J Exp Psychol Gen*. doi: 10.1037/a0028560
- Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 683-692.
- Stevens, S. S. (1956). The direct estimation of sensory magnitudes-loudness. *The American journal of psychology*, 69(1), 1-25.
- Varey, C. A., Mellers, B. A., & Birnbaum, M. H. (1990). Judgments of Proportions. *Journal of Experimental Psychology-Human Perception and Performance*, 16(3), 613-625.
- Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. [Research Support, Non-U.S. Gov't]. *Psychonomic bulletin & review*, 12(1), 66-80.

### Figure Captions

Figure 1: Illustration of the bounded and unbounded number-line tasks. The X shows where the target number (e.g., 13) appears on each trial. The dashed arrow indicates the movement of the tick mark (when dragged by the participant) from its starting position to the target position chosen as a response.

Figure 2: Children's average bounded number-line estimates by target number (left) and raw responses by target number (right). Top row shows data from four- to six-year-olds; bottom row shows data from seven- and eight-year-olds. Note that the left-hand figures replicate the 'log-to-linear shift'—the developmental change in performance, about which much has been written.

Figure 3: Children's average unbounded number-line estimates by target number (left) and raw responses by target number (right). Top row shows data from four- to six-year-olds; bottom row shows data from seven- and eight-year-olds. Note that the general shape of the response curve is the same for both groups.

Figure 4: Average  $R^2$  by model fit by age for the bounded (top) and unbounded (bottom) number-line task.

Figure 5: The proportion of children in each age group who are best fit by each model in the bounded (top) and unbounded (bottom) number-line task

Figure 6: Average response bias for each age group (4-7) by Task (bounded and unbounded). A response bias of  $\beta=1$  represents accurate responding. Note that older children responded more accurately than younger children on the bounded task, but no such age-related difference was seen on the unbounded task.

Figure 7: Within-child comparison of data from 6 four-year-olds who completed both the bounded (left) and unbounded (right) task.



Figure 1

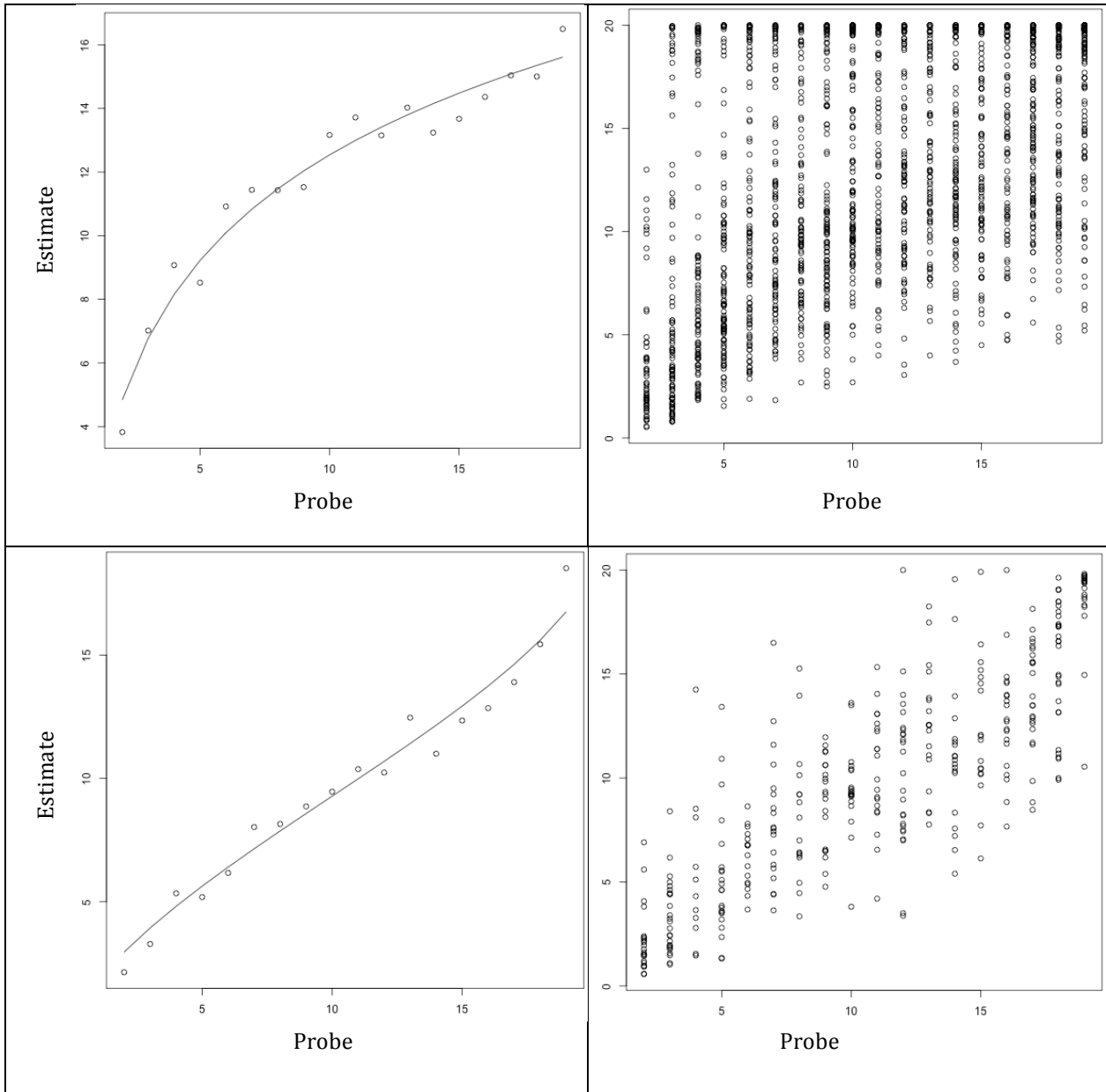


Figure 2

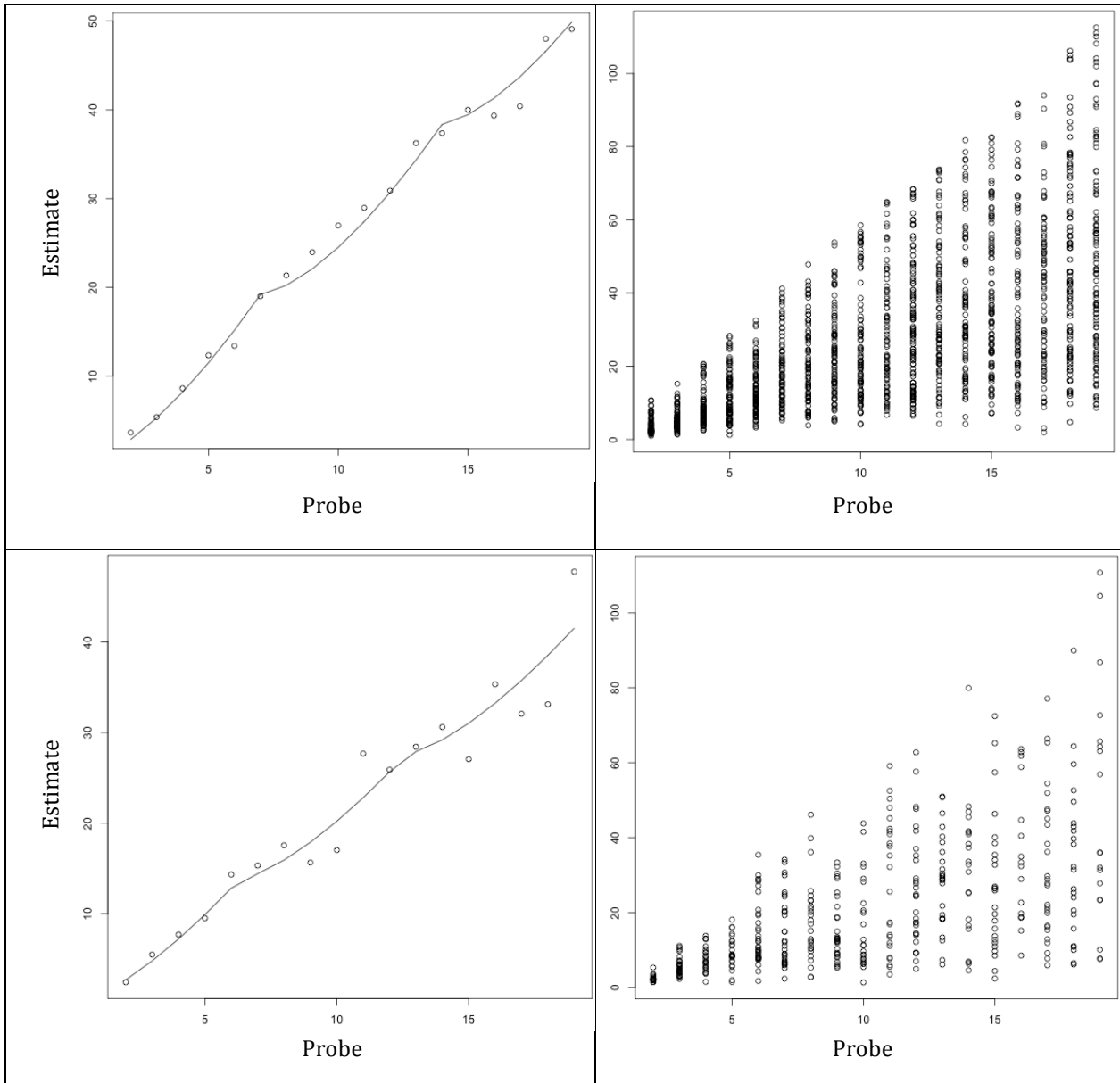


Figure 3

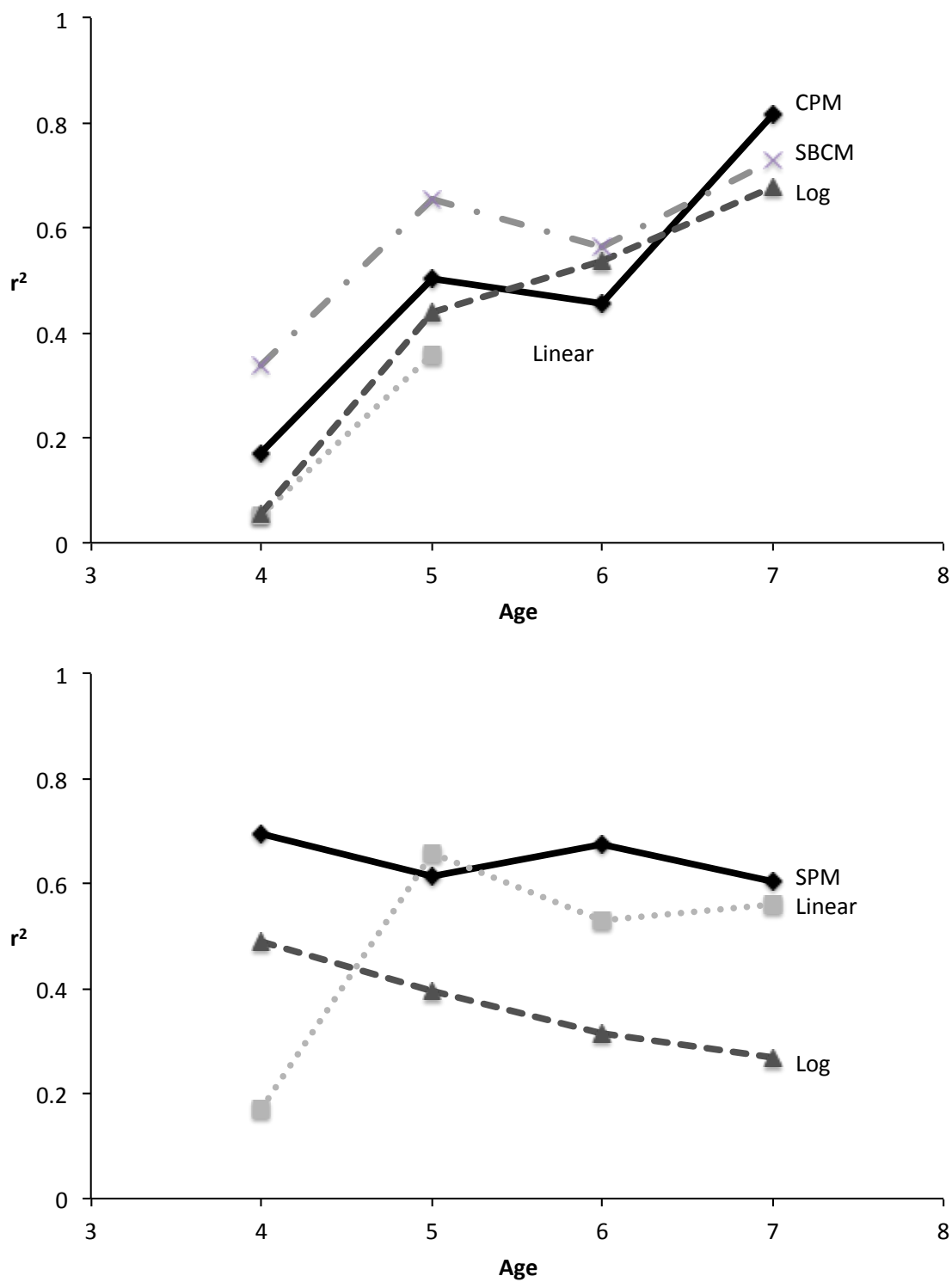


Figure 4

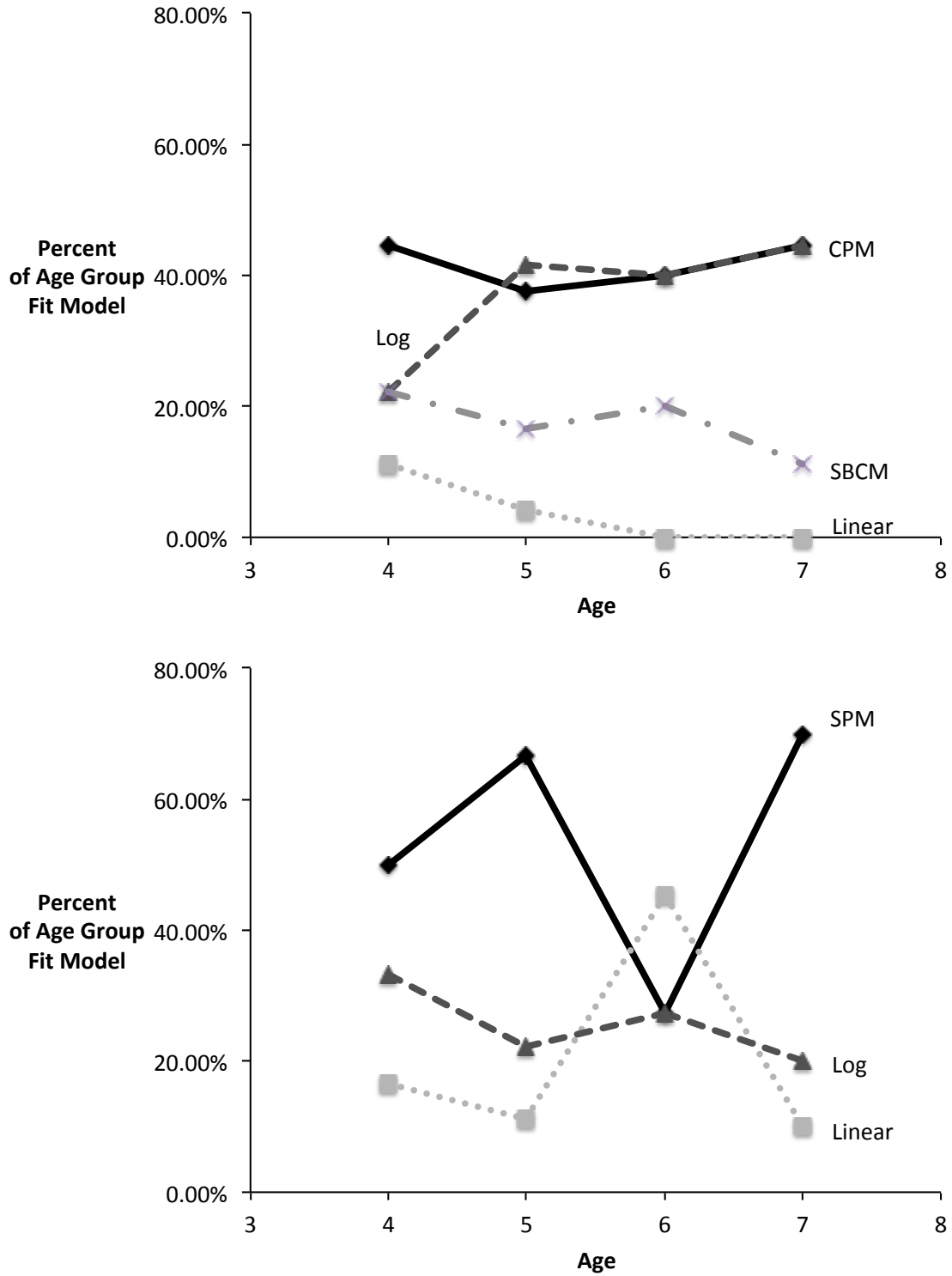


Figure 5

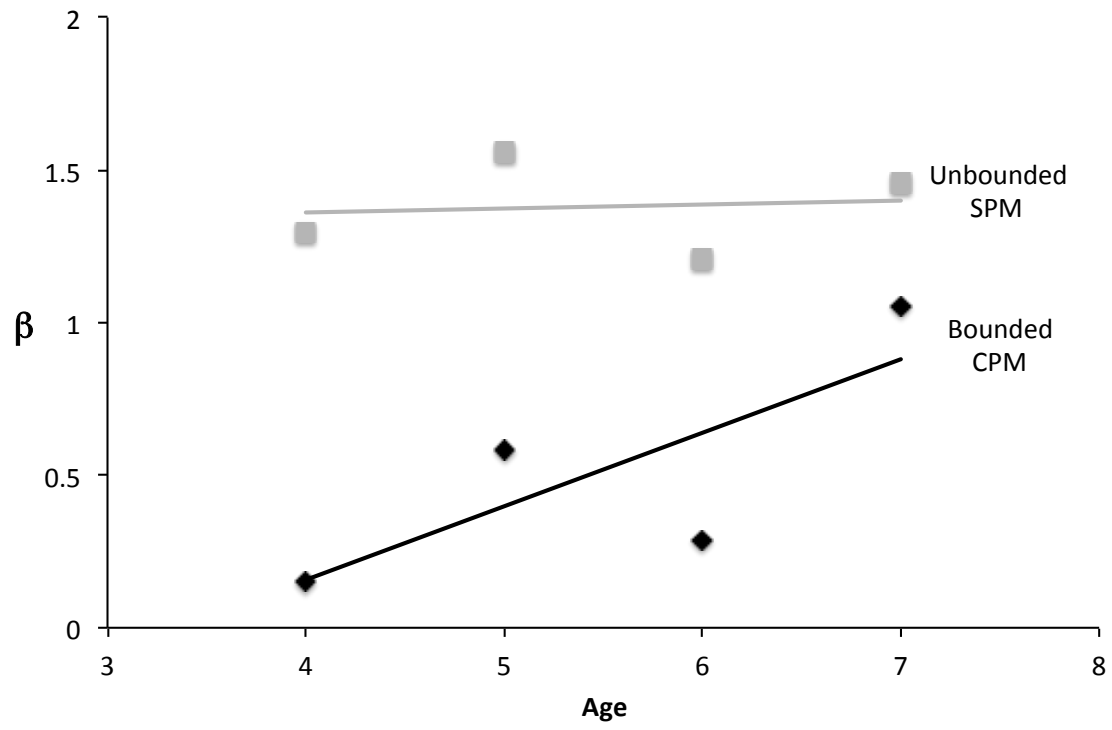


Figure 6



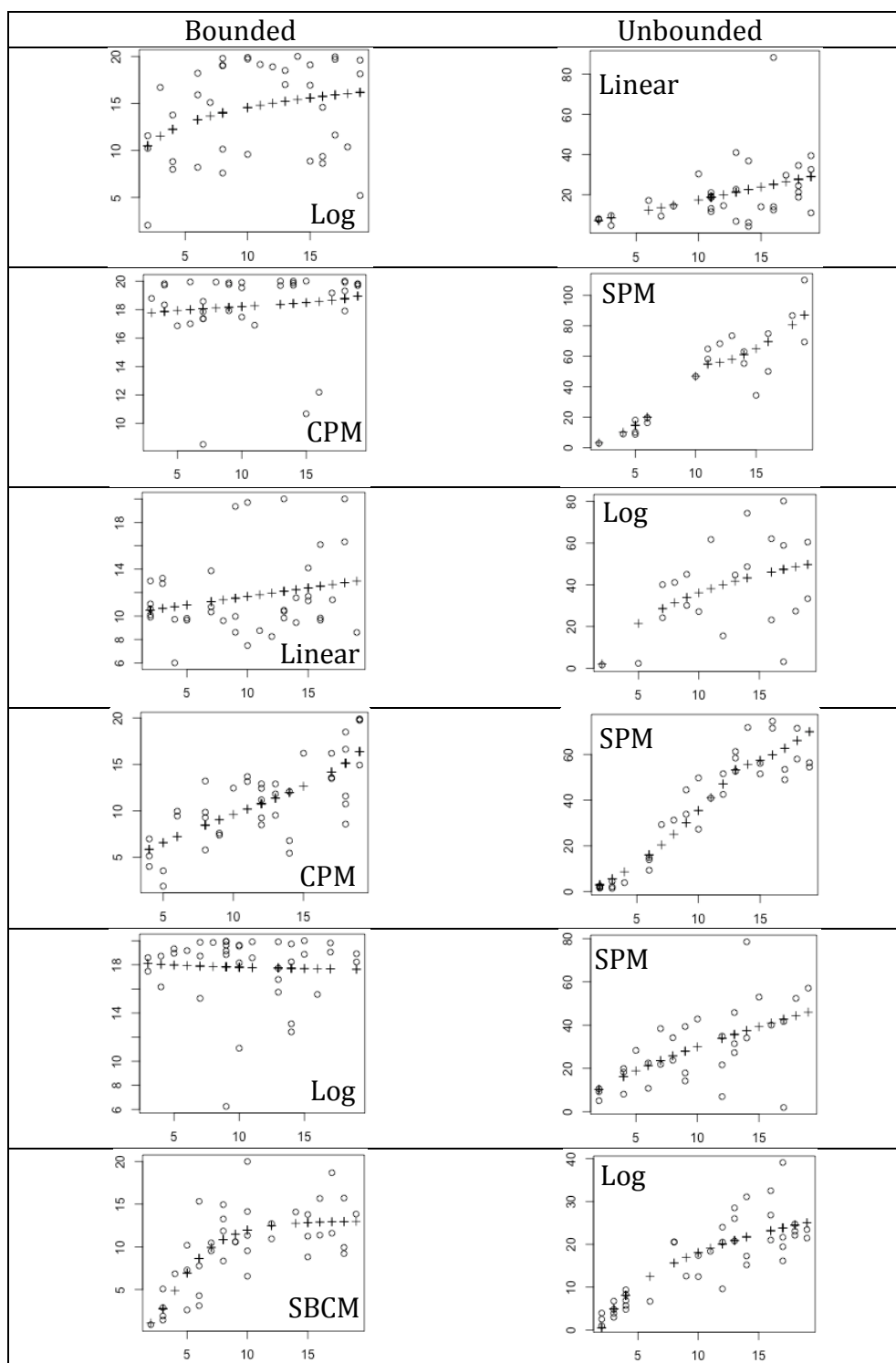
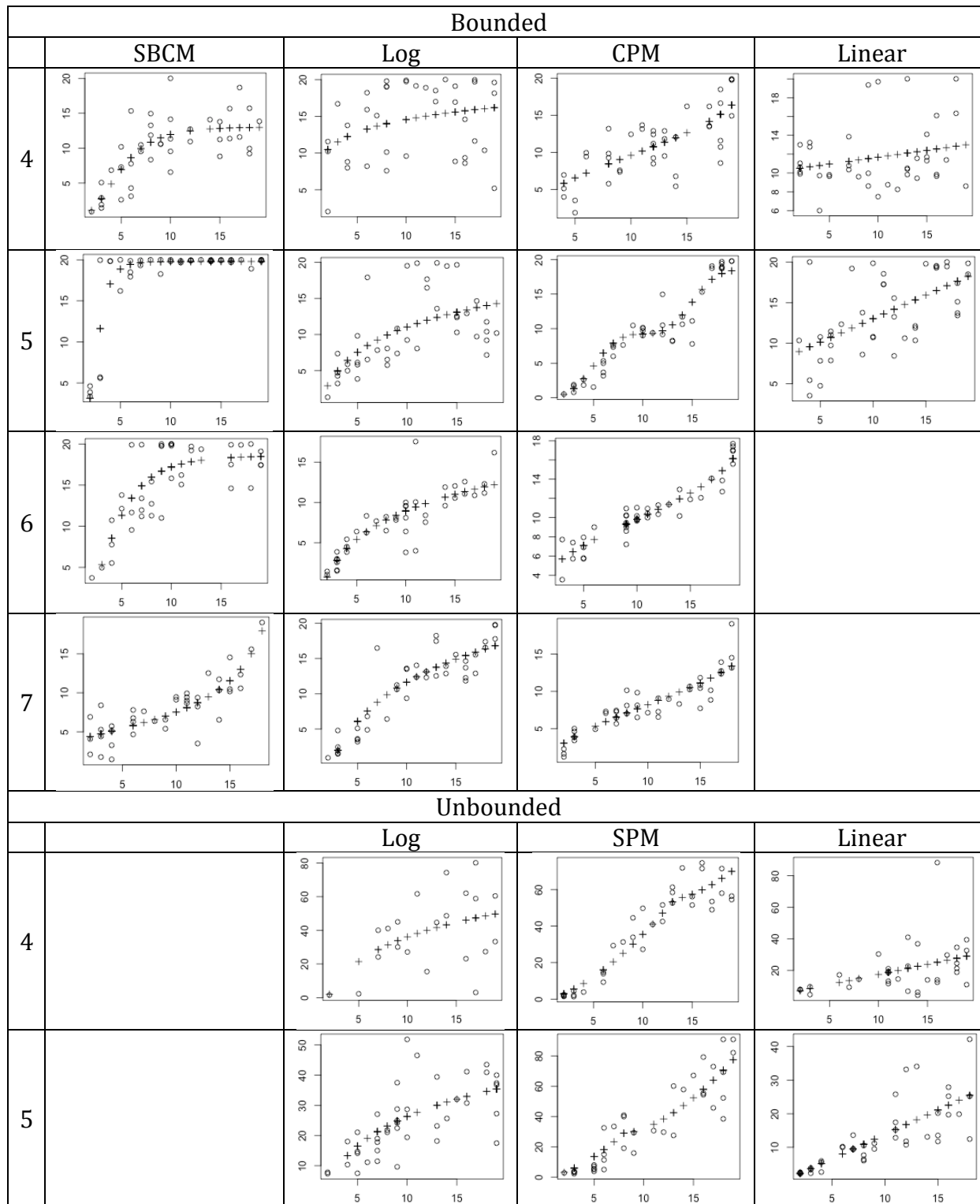


Figure 7

## Online Supplement



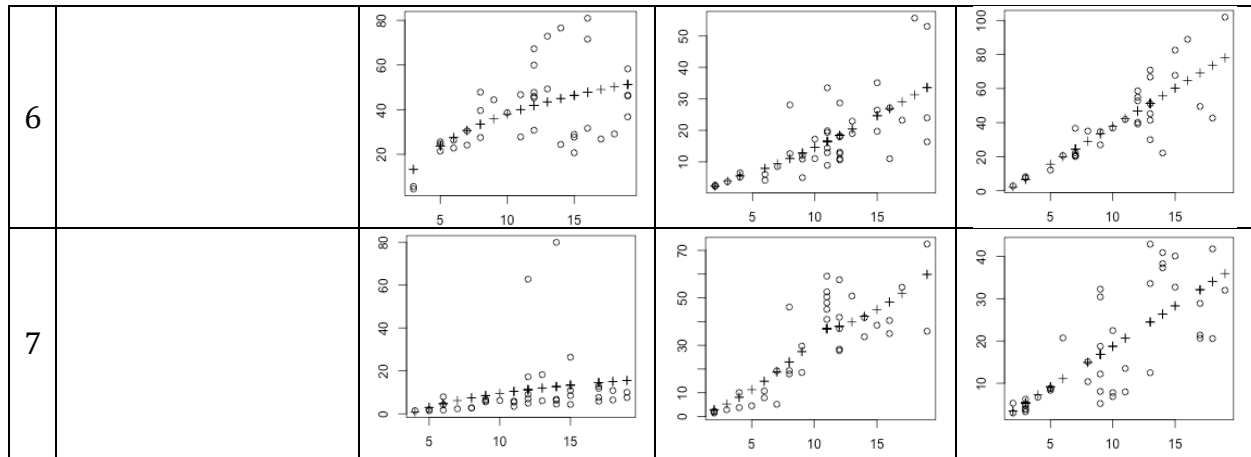


Figure 1: Data from example children in each age group (ages 4, 5, 6 and 7) who were best fit by each of the models in the bounded and unbounded task.

## Appendix

Cross-modality matching tasks require the participant to translate perceived variation in one stimulus class (e.g., numbers, specifically integer magnitudes) into variation in another stimulus class (e.g., line lengths). In the case of the number-line task, the participant must convert an integer,  $I$ , into a psychological representation of quantity,  $\psi_q$ . Assuming this conversion is not perfect, there is a stimulus transformation function ( $S_I$ ) that describes the conversion bias. This can be described by the following formula,

$$\psi_q = S_I(I) + e_I, \quad (1)$$

where  $e_I$  is trial-by-trial variation or error. Note that the participant's line-length response is based on  $\psi_q$ .

Any response that requires the participant to convert a psychological representation into a physical response (in this case, to convert an integer magnitude into line length) also undergoes a conversion, termed a *response transformation*. Because this transformation converts the psychological quantity ( $\psi_q$ ) into a line, it is the inverse of the stimulus transformation function of a line (which converts line length into a psychological quantity), or

$$Y_L = S_L(\psi_q)^{-1} + e_L, \quad (2)$$

where  $Y_L$  is the line length response. Using Equation 1 and replacing  $\psi_q$  with  $S_I(I)$ , this results in the following formula, which relates an integer to the line-length produced by a subject,

$$Y_L = (S_L(S_I(I) + e_I) + e_L)^{-1}. \quad (3)$$

Although many previous studies have tried to describe  $S_I$  (i.e., the subject's representation of integer magnitudes) most have ignored  $S_L$  (the response transformation associated with changing a mental magnitude into a line length). In other words number-line

studies have typically assumed that people's estimates of line length are direct and unbiased measures of their mental representations of integer magnitudes. This is a mistake.

It has been reported in the literature that when the bounded number-line task is reversed (i.e., when a participant is shown a position on a number line and asked what number should go there), the resulting response curve is exponential rather than logarithmic (Siegler & Opfer, 2003). Equation 3 explains why this occurs. On the reversed (position-to-numeral) task, the participant converts the line length into a psychological quantity, and then converts that psychological quantity into a number. This is described by the reverse of Equation 4, or

$$Y_L = (S_I(S_L(L) + e_L) + e_I)^{-1}. \quad (4)$$

Typically, stimulus biases ( $S_I$  and  $S_L$ ) revealed through cross-modality matching tasks ( $S_L$  in the equations above) follow Stevens' Power Law (Stevens, 1956). Stevens' Power Law states that the psychological representation of a stimulus ( $\Psi$ ) is a function of the physical stimulus ( $\phi$ ) taken to a power,  $\beta$ ,

$$\Psi = k\phi^\beta, \quad (7)$$

where  $k$  is a function of the units of measurement and  $\beta$  is the characteristic exponent that describes the perceptual bias (Stevens, 1956). A  $\beta > 1$  indicates a positively accelerating (i.e., exponential) bias;  $\beta < 1$  indicates a negatively accelerating (i.e., logarithmic) bias. Equation 1 in the text is derived directly from this Equation 7 in the Appendix (see Spence (1990)).

## Footnotes

---

<sup>1</sup> Here, we use the term *bias* in the psychophysical sense. Specifically, bias is the difference between an observer's perception of a stimulus (or concept) and the objectively accurate description of the stimulus (or concept).

<sup>2</sup> Here, we use the term *Perceptual bias* to mean both the bias associated with matching a quantity to a number (often called the numerical bias) and the bias associated with perceiving the length of a line.

<sup>3</sup> Although we can never completely rule out that performance on the unbounded task may influence performance on the bounded task our data strongly suggest this is not the case. First, our participants perform the same on the unbounded task as participants in the extant literature (who had no experience with the unbounded task). Furthermore, we found far less within participant similarity in biases across tasks than we anticipated. If performance in the unbounded task influenced performance in the bounded task, we would expect much more within participant similarity.

<sup>4</sup> The removal of outliers is beneficial when individual data points have large leverage on a fitted function (Osborne & Overbay, 2004). In the present dataset, we fit multiple non-linear models to each individual participant's data. Because there were relatively few data points per participant (as few as 20 in some cases), the removal of outliers facilitated the convergence of the statistical models for a subset of participants. Rather than remove these participant's entire dataset, we created an a priori formula for the removal of outlier data points that we applied to all participants. Removal of these outliers does not favor one model over another.