Article:

Cohen, Dale, Blanc-Goldhammer, Daryn and Quinlan, Philip Thomas orcid.org/0000-0002-8847-6390 (2018) A mathematical model of how people solve most variants of the number-line task. COGNITIVE SCIENCE. pp. 2621-2647. ISSN 0364-0213

https://doi.org/10.1111/cogs.12698

A mathematical model of how people solve most variants of the number-line task

Dale J. Cohen[1*], Daryn Blanc-Goldhammer[1]

and

Philip T. Quinlan[2]

[1]Department of Psychology, University of North Carolina Wilmington

[2]Department of Psychology, The University of York

Keywords: Numerical Cognition, Numerical Architecture, Number processing, Number-line

*Correspondence to Dale J. Cohen, PhD., Department of Psychology, University of North Carolina Wilmington, 601 South College Road, Wilmington, NC 28403-5612. Email: cohend@uncw.edu. Phone: 910.962.3917. Fax: 910.962.7010. This work supported by NIH grant RO1HD047796 awarded to Dale J. Cohen.

Abstract

Current understanding of the development of quantity representations is based primarily on performance in the number line task. We posit that the data from number line tasks reflect the observer's underlying representation of quantity, together with the cognitive strategies and skills required to equate line length and quantity. Here, we specify a unified theory linking the underlying psychological representation of quantity and the associated strategies in four variations of the number-line task: the production and estimation variations of the bounded and unbounded number-line tasks. Comparison of performance in the bounded and unbounded number-line tasks provides a unique and direct way to assess the role of strategy in number-line completion. Each task produces a distinct pattern of data, yet each pattern is hypothesized to arise, at least in part, from the same underlying psychological representation of quantity. Our model predicts that the estimated biases from each task should be equivalent if the different completion strategies are modelled appropriately and no other influences are at play. We test this equivalence hypothesis in two experiments. The data reveal all variations of the number-line task produce equivalent biases except for one: the estimation variation of the bounded number-line task. We discuss the important implications of these findings.

1 A mathematical model of how people solve most variants of the number-line task

Of the many different experimental paradigms that have been used to examine the human number system one of the simplest is the number-line task. In the typical number-line task the participant is presented with a horizontal line whose length represents a given quantity. The left and right bounds of the line are, respectively, labeled with the minimum and maximum values of the range. On every trial, the participant's task is to mark, on the line, the point that corresponds to a probed number. We term this the *bounded number-line task*. Recently, there has been spirited debate concerning how best to characterize the psychological representations and processes that underpin performance on the bounded number-line task (see Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Link, Nuerk & Moeller, 2014; Opfer, Thompson, & Kim, 2016; Rouder, & Geary, 2014; Slusser, Santiago, & Barth, 2013). On one side of the debate are those who contend that the participant's responses reflect, primarily, the *mental representation* of quantity (e.g., Dehaene, Izard, Spelke & Pica, 2008; Opfer et al., 2016; Siegler & Booth, 2004; Siegler & Opfer, 2003). On the opposing side of the debate are those who contend that the participant's responses must be understood both in terms of a specific subtraction/division strategy as well as the nature of the mental representation of quantity (Barth, & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Rouder, & Geary, 2014; Slusser, Santiago & Barth, 2013). Whereas this debate continues in the developmental literature (e.g., Barth, & Paladino, 2011; Opfer et al., 2016; Rouder, & Geary, 2014; Slusser, et al., 2013), there appears to be a consensus forming that adults primarily use the subtraction/division strategy when completing this task (e.g., Barth, & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Rouder, & Geary, 2014; Slusser, et al., 2013). Although there has been some empirical examination of the cognitive processes that underpin the subtraction/division strategy and how they may interact with the mental representation of quantity when adults complete the number-line task (e.g., Barth, & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Rouder, & Geary, 2014; Peeters, Sekeris, Verschaffel, & Luwel, 2017; Peeters, Vershaeffel, Luwel, 2017), there has been no precise description or modeling of these processes. Here, we present and test a mathematical model of these processes across a range of different kinds of number-line tasks.

1.1 The Mental Representation of Numbers

Central to many accounts of the psychological representation of quantities is the hypothesis that psychological representations of quantities are abstract and lie on an internal continuum that is sometimes referred to as the *mental number line* (Dehaene, Dupoux & Mehler, 1990; Gallistel & Gellman, 1992). Due to various sources of noise, there is imprecision in mapping quantities onto the continuum, and, this is reflected in positing that each number maps onto its own quantity distribution: Each number's quantity representation is captured by a distribution of values on the continuum. Different theorists have posited competing accounts of how best to describe the relation between the distributions representing the quantities of successive numbers. The critical differences are with respect to (i) the mean placement of the distributions on the continuum, and, (ii) the relative variances of the distributions. All accounts agree, though, that successive quantity representations are rank ordered and that the corresponding quantity distributions are overlapping.

There are three main theories describing the psychological representation of quantities, namely, (i) the Linear account, (ii) the Logarithmic account, (iii) and the Scalar Variance account. The Linear account posits that successive quantity representations are rank ordered at equal intervals and that the different quantity distributions have the same variance (e.g., Cantlon, Cordes, Libertus & Brannon, 2009). The Logarithmic account posits that successive quantity representations are spaced on a logarithmic scale and that the different quantity distributions have the same variance (Dehaene 1992, 2003). As a consequence, the quantity distributions for smaller quantities are spaced relatively far apart, but as magnitudes increase, the distances between the corresponding distributions become increasingly compressed. The Scalar Variance account posits that the means of the ordered quantity distributions are spaced linearly but their variances scale linearly with quantity (see Church, Meck & Gibbon, 1983; Gallistel & Gelman, 1992; Meck & Church, 1983; Meck, Church & Gibbon, 1985). One important line of evidence, that has been used to try to adjudicate between these different accounts, comes about from responses on the number-line task.

Siegler and colleagues were the pioneers of the research that related the number-line task to the psychological representation of quantity (e.g., Booth & Siegler, 2006; Siegler & Booth, 2004; Siegler & Opfer, 2003). They set out evidence of an intriguing developmental trend in performance on the traditional number-line task. Central here is the notion of a response function that maps out numerical estimates of the probed numbers as a function of those numbers. Very young children's patterns of responses are well captured by a logarithmic response function. However, the children's responses change with age such that the developmental trajectory is one where the logarithmic pattern is replaced with a linear one. Elder children seem to map quantities on the number-line as defined by an equal interval scale. This developmental progression is captured by something known as the *log-to-linear shift hypothesis* (Siegler, Thompson, & Opfer, 2009). The hypothesis states that the underlying representation of quantity changes from one as described by the Logarithmic account to one as described by the Linear account. However, the hypothesis itself depends on the assumption that performance on the number-line task provides a direct window onto the psychological representation of quantities. We term this the *Assumption of Pure Numerical Estimation*. In what follows we examine the veracity of this assumption.

1.2 Strategies and the Number-line Task

The log-to-linear shift hypothesis has been extremely influential since it was first put forward (see Barth & Paladino, 2011), but recently, the validity of the Assumption of Pure Numerical Estimation has been questioned (see e.g., Link et al., 2014). Indeed, disquiet over the Assumption of Pure Numerical Estimation has resulted in accounts of task performance in which the emphasis is shifted from discussion of the mental representation of quantities to discussion of putative cognitive strategies that are deployed when participants engage with the number-line task (see e.g., Chesney & Matthews, 2013; Ebersbach, Luwel, Frick, Onghena & Verschaffel, 2008; Link et al., 2014; Nuerk, Weger, & Willmes, 2001; Peeters, Sekeris, Verschaffel & Luwel, 2017; Peeters et al., 2017). Perhaps the most widely discussed example is that described in the subtraction/division model put forward by Cohen and Blanc-Goldhammer (2011) and Cohen and Sarnecka (2014) (see also Barth & Paladino, 2011; and, Slusser, et al., 2013).

Central to this alternative approach are claims that the number-line task is influenced by (i) how a person represents integers, (ii) how a person perceives lines, and, (iii) the cognitive strategies required to equate line length and quantity and the mensuration skills necessary to successfully complete those strategies. In discussing these claims, Cohen and Blanc-Goldhammer (2011) considered performance on both the *bounded* and the *unbounded* versions of the task. The traditional number-line task is a bounded number-line task because all probe values fall on a line bounded by labeled minimum and maximum values. In contrast, in the unbounded number-line task, the participant is presented with a standard line segment denoting a value of 1 (i.e., its end bounds are labeled 0 and 1) and is asked to reproduce a target quantity with respect to this unit length. This is accomplished by stretching the line's length by the click-and-drag actions on a computer mouse.

Despite the salient differences across the two kinds of number-line tasks, it is important to be clear about what such differences mean at a psychological level. Indeed, it is of central import here to be explicit about exactly what performance on the two tasks reveals about common and distinctive properties of mathematical cognition. In this respect, the main motivation in developing a general computational account of performance on the tasks is to be explicit about how different cognitive strategies are combined with the same representational system to produce systematic patterns of performance across very different testing conditions. For example, with the bounded number-line task there is no objective method to judge the distance associated with a single unit without also judging the distance associated with the remaining units. If the end value of the number-line is 100, and the participant is asked to indicate the distance associated with a single unit, then the participant must judge the distance associated with both 1 (from the left side) and 99 (from the right side). Here, the participant iteratively adjusts the target position until size of the left-hand and right-hand portions of the number line appear to represent the target and the end value minus the target, respectively. This strategy is generalizable to all target values and depends on the participant being able to subtract the target number from the upper bound. Cohen and Sarnecka (2014) term this the *subtraction* strategy and it represents the lowest level mathematical skill necessary to complete the task. Assuming the participants have the necessary mensuration skills, the participants may also complete the task by dividing the target by the upper bound, thus producing the *target proportion*. The participant then iteratively adjusts the target position until size of the left-hand and right-hand portions of the number line appear to represent the target proportion and one minus the target proportion, respectively. Cohen and Sarnecka (2014) term this the *division* strategy. The subtraction and division strategies make similar predictions. Therefore, we will refer to them jointly as the *subtraction/division* strategy.

In the subtraction/division strategy, the end points can be considered reference points. Therefore, this strategy generalizes to any use of reference points in the bounded number line task. For example, one can have explicit (Peeters, Sekeris, Verschaffel & Luwel, 2017; Peeters et al., 2017) or implicit reference points (as discussed by Cohen & Blanc-Goldhammer, 2011) at, say, the 25%, 50%, etc., points on the number-line. When a reference point other than the end point is used, the reference point is substituted for the end point in the description of the strategy (we detail this mathematically below). Thus, the number of reference points, and their placement, are parameters of the subtraction/division strategy.

The subtraction/division strategy with two reference points has been modeled by Spence (1990) in his description of how observers estimate proportions. Spence's (1990) model has been generalized to multiple reference points by Hollands and Dyre (2000). Significant evidence has accumulated to support the use of this strategy in both children and adults (Barth, & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Rouder, & Geary, 2014; Peeters, Sekeris, Verschaffel & Luwel, 2017; Peeters et al., 2017) when they engage with the bounded number-line task.

In contrast, the successful completion of the unbounded number-line task does not require the participant to complete any mathematics to identify the length of a single unit: This is provided by the length of the line itself. Essentially, what the participant must do is *add* the target number to the lower bound. Hence in deploying an *addition* or "*dead reckoning*" strategy, participants iteratively increase the length of the line by some fixed interval - known as *the* *working window*. For some participants, the working window is equivalent to 1 unit, but for others, larger values are used. For instance, in producing a line of “12” participants might first estimate a value for “10” and then increment this by a further “2” units. As a consequence, the size of the working window is an estimable parameter of the dead reckoning strategy (see Equations 7 and 8 below).

In sum, the two variations of the number-line task (bounded and unbounded) each produce different patterns of data. The contrasting patterns of data are consistent with the view that the two tasks possess very different task constraints. These task constraints necessitate different cognitive strategies to equate line length to quantity and each cognitive strategy requires different mathematical skills to complete. Nevertheless, Cohen and Blanc-Goldhammer (2011) hypothesized that both the bounded and unbounded number-line tasks access the same psychological representation of quantity. So, if the cognitive strategies are modelled correctly, it should be possible to extract a common estimate of the psychological representation of quantity from each task. Systematic deviations in accuracy between the psychological representation of quantity and the actual quantity is termed *bias*. A key hypothesis of the current account is (i) if the cognitive strategies required to complete each version of the number-line task are modeled correctly, and (ii) participants have the necessary mensuration skills to successfully deploy those strategies, then the same amount of bias should be found across the bounded and unbounded versions of the number-line task.

Estimates of bias from both tasks should diverge, however, when the mensuration skills of participants are either undeveloped or compromised in some way. This is exactly what Cohen and Sarnecka (2014) recently found. Cohen and Sarnecka (2014) asked children to complete both kinds of number-line tasks. They found that the bias estimated from performance on the bounded number-line task was related to age of the participant (similar to other published results; Berteletti, Lucangeli, Piazza, Dehaene & Zorzi, 2010; Booth & Siegler, 2006; Geary, Hoard, Nugent & Byrd-Craven, 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003; Slusser, et al., 2013; Thompson & Opfer, 2008). In contrast, the bias estimated by the unbounded number-line task was unrelated to age of the participant. Cohen and Sarnecka (2014) concluded that performance on the bounded number-line reflected the development of advanced mensuration skills. That is, successful completion of the bounded task requires mastery of subtraction and/or division, and very young children (e.g., those less than 6 years old or so) do not possess these mathematical skills. The authors argued that the developmental changes in performing the bounded number-line task reflect emerging skills in the strategies that are deployed to complete task successfully. This is in contrast to claims about evidence for changes to the underlying quantity representations (Siegler et al., 2009) .

2 A Model of Number-Line Task Completion

Here we present a computational model that describes the processes involved in number-line estimation and the role of strategy in the completion of these tasks. The model describes how the patterns of data produced in the unbounded and bounded number-line tasks are different, but predictable from their reliance on the processes involved in the perception of lines and quantities. We term the overall bias associated with these perceptions the *line/quantity bias*. We claim that the line/quantity bias reflects the fundamental perceptual event underlying the patterns of the data produced in both the bounded and unbounded number-line tasks. We show how participants use different strategies to complete the bounded and unbounded number-line tasks and we model these strategies in Equations 5-8. Finally, in our model, the line/quantity bias is recoverable. Specifically, the line/quantity bias is estimated in the beta parameter (β) of Equations 5-8.

Figure 1 presents a schematic of the processes that we propose are involved in successful completion of the bounded and unbounded number-line tasks. Broadly speaking, there are four major cognitive components involving (i) encoding of integers, (ii) encoding of lines, (iii) various mensuration skills required to implement the discussed strategies, and, (iv) the mental representation of quantity, Ψ_q. We consider each separately below.

We assume that to complete the task successfully, the observer must accurately encode the presented integers and lines. We acknowledge, however, that integer encoding may be sub-optimal and when encoding errors occur, they can influence the perception of the target quantity (Cohen, 2009, 2010; Cohen & Quinlan, 2016). Nonetheless, in simplifying the modeling, many assume that the perceptual representations of the integer are veridical. Following perceptual encoding, all number symbol and line inputs are assumed to activate a psychological representation of quantity, Ψ_q. The nature of this varies according to which model of number representation is assumed (i.e., Linear, Logarithmic, or Scalar Variance).

Turning to mensuration skills, we have argued that whereas the bounded number-line task implicates the ability to subtract accurately (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014), in the unbounded number-line task, the requisite mensuration skills include the ability to add accurately (Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014). According to Cohen and Sarnecka (2014) young children master the mensuration skills necessary to complete the unbounded number-line task prior to those necessary to complete the bounded number-line task. As mensuration skills are primarily mathematical, we include a mathematical skills component in our model of performance on number-line tasks (See Figure 1). It is often assumed that the inputs to this mathematical component are direct from the encoding processes (see e.g., Dehaene's Triple Code Model, Dehaene, 1992; Dehaene, & Akhavein, 1995; Dehaene, & Cohen, 1991). However, some have hypothesized that quantity representation influences this mathematical component as well (e.g., Halberda, Mazzocco, & Feigenson, 2008).

2.1 Mathematical Instantiation

We take it that both the bounded and unbounded number-line tasks are cross-domain matching tasks. In such tasks, the participant must transform variation in one stimulus domain (e.g., integer quantities in the numerical domain) into variation in another (e.g., line lengths in the length domain). We use the following symbols for clarity: φ symbolizes the physical stimulus; Ψ symbolizes the psychological representation of the stimulus; and Θ symbolizes the participants physical response to the task. These symbols take a subscript such that the subscript denotes the type of stimulus; an *I* denotes an integer, *q* denotes a quantity, and *L* denotes a line.

The number-line task requires the participant to transform an integer, *I,* into a psychological representation of quantity, Ψ_q. Whenever a physical stimulus is transformed into a psychological representation, there is a *stimulus transformation* function, *S*, such that:

$$\Psi_q = S_I(\varphi_I) + e_I, \tag{1}$$

In Equation 1, S_I is the stimulus transformation function for converting an integer into a psychological representation of quantity, and e_I is trial-by-trial variation or error. We may describe this function as reflecting a perceptual bias (for a review, see Cohen, Ferrell, & Johnson, 2002; Shepard, 1981). The log and linear transformations are particular instances of a stimulus transformation function, S_I, but other functions are possible.

In addition, whenever a psychological representation is transformed into a physical response, there is a *response transformation* function, *R*. In the number-line task, this transformation converts the psychological quantity (Ψ_q) into a line of a particular length. A key assumption is that the response transformation function of a specific stimulus (e.g., a line) is the inverse of the corresponding stimulus transformation function. For the number-line task, the response transformation function is as follows,

$$\Theta_L = R_L(\Psi_q)^{-1} + e_L, \quad (2)$$

where Θ_L is the line length response. Replacing Ψ_q with Equation 1, the following formula relates an integer to the line-length produced by a participant,

$$\Theta_L = R_L(S_I(\varphi_I) + e_I)^{-1} + e_L. \quad (2a)$$

Equation 2a illustrates that the participant's response (Θ_L) is influenced by his or her perception of both quantity (S_I) and line length (R_L) and both are subject to bias. S_I is the quantity bias associated with integers and R_L is the length bias associated with lines.

The model can be expanded to address issues with estimation (position-to-number, after Siegler & Opfer, 2003) versions of the number-line task. For example, in an estimation task, a participant is presented with a number-line with a tick mark placed on it and is instructed to judge the quantity that is denoted by the position of the tick mark. The participant's response is an integer, Θ_I, and is given by the inverse of Equation 2a.

$$\Theta_I = R_I(S_L(L) + e_L)^{-1} + e_I \quad (2b)$$

In setting out Equations 2a and b in this manner the estimation variant of the number-line task (position-to-number) and the production variant (number-to-position, after Siegler & Opfer, 2003) are shown to be identical, yet, inverse tasks (Brooke & MacRae, 1977). As a consequence, the estimation and production tasks are subject to inverse biases and should produce inverse patterns of data. Indeed, Equation 2b explains why the bounded number-line task produces inverse results when presented as a line-to-number task, rather than a number-to-line task (Siegler & Opfer, 2003).

2.1.1 The Relevance of Stevens' Power Law

In cross-domain matching tasks, the biases we have labeled as S_I and R_L follow Stevens' Power Law (Stevens, 1956; for review, see Gescheider, 1988). Stevens' Power Law states that the numbers assigned to a perceptual event (Θ_I) takes the form of a power function,

$$\Theta_I = k\varphi^{\beta} \qquad (3)$$

where k is a function of the units of measurement, φ represents the physical stimulus intensity (i.e., in the current context, line length), and β is the characteristic exponent that describes the perceptual bias (Stevens, 1956). More specifically, β describes the mean placement of the perceptual distributions along the perceptual continuum, it does not describe the relative variances of the distributions[1]. Within the context of a number-line task, we take β to reflect the overall contribution of the quantity and line biases and we therefore label this the *line/quantity bias*. β describes whether the interaction between the mean placement of the quantity distributions and the mean placement of the line distributions (i.e., the line/quantity bias) is linear, negatively accelerating, or positively accelerating. Estimates of this overall line/quantity bias can be recovered from the behavioral data in number-line tasks. As with any such characteristic exponent, when $\beta > 1$, the line/quantity bias is described by a positively accelerating function (i.e., exponential). When $\beta < 1$, the line/quantity bias is described by a negatively accelerating function (i.e., logarithmic) (see Figure 2). When $\beta = 1$ this indicates unbiased performance, and is consistent with a linear model.

2.1.2 The bounded number-line task

Spence (1990) applied Steven's Power Law to the estimation of proportions. Specifically, Spence assumed that participants judged the magnitude of a presented proportion (φ_p) by estimating the magnitude of both φ_p, and $(1 - \varphi_p)$. If a perceptual bias associated with proportion estimation exists and it follows Steven's Power Law, then the estimated proportion, Θ_p, would be described by the following formula:

$$\Theta_p = \varphi_p{}^{\beta} / [(\varphi_p{}^{\beta}) + (1-\varphi_p)^{\beta}], \tag{4}$$

Spence (1990) termed Equation 4 the *Power Model*. In this case, β is the same characteristic exponent present in Steven's Power Law. In fact, Spence's Power Model is simply an instantiation of Steven's Power Law that captures the division strategy when the target is between 0 and 1 and the two end points (0 and 1) are used as the only reference points. Estimates of proportions, that are based on Spence's Power Model, describe an ogive when $\beta > 1$, or inverse ogive, when $\beta < 1$. The ogive always inflects at 0.5 because the sum is constrained to 1 and, at this inflection point, both φ_p and $(1-\varphi_p)$ will be perceptually equivalent.

In addressing the estimation of proportions, Spence's (1990) Power Model has two fixed reference points, namely 0 and 1. Hollands and Dyre (2000) generalized Spence's (1990) Power Model to include multiple reference points and applied it to an integer scale. Hollands and Dyre termed their model the *Cyclic Power Model* (CPM) because their formula (see Equation 5), creates multiple ogive cycles depending on the number of reference points when modified for the production (i.e., number-to-position) number-line task (see Figure 2).

$$\Theta_L = \frac{(\varphi_I - R_{i-1})^{\beta}}{(\varphi_I - R_{i-1})^{\beta} - (R_i - \varphi_I)^{\beta}} * (R_i - R_{i-1}) + R_{i-1}, \text{ if } R_{i-1} \leq \varphi_I \leq R_i \tag{5}$$

where Θ_L is the line response, φ_I is the presented integer, R_i is nearest larger reference point and R_{i-1} is the nearest smaller reference point.

The CPM is a mathematical instantiation of the subtraction/division strategy with *r* reference points, where $r \geq 2$. Figure 2 presents instantiations of the CPM when $\beta < 1$ and $\beta > 1$, and when $r = 2$ (also Steven's Power Model), $r = 3$, and a mixed model whereby there are an equal mixture of trials when $r = 2$ and $r = 3$. The CPM (Hollands & Dyre, 2000) has been shown to model successfully the strategies deployed by both adults and old children (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Rouder & Geary, 2014; Slusser, et al., 2013).

2.1.3 The unbounded number-line task

In addressing performance on the unbounded number-line task, Cohen and Blanc-Goldhammer (2011) developed the Scallop Power Model (SPM) to model participants' strategy. Recall that in the unbounded number-line task the participant need merely estimate the distance of the line based on the quantity denoted by the integer. Cohen and Blanc-Goldhammer (2011) expressed this dead reckoning strategy in terms of three variations of the SPM based on Steven's Power Law. The differences between these variations reflect how the additions are carried out when the participant estimates the length of the line. In all formulas set out here, β estimates the line/quantity bias. Figure 3 presents the three variations of the SPM (Equations 6-8) when $\beta < 1$ and $\beta > 1$.

The *single scallop* variation is a strategy whereby the participant judges the length of the line in a single global estimate. This strategy is modeled by a simple power function given in Equation 6: -

$$\Theta_L = \varphi_I^{\beta} \tag{6}$$

where Θ_L is the estimated line length, φ_I is the presented integer, and β is the characteristic exponent. Equation 6 is equivalent to Stevens' Power Law when $k = 1$. We assume that this reflects the simplest form of dead reckoning: more complex strategies are described respectively by the dual scallop and multi-scallop data patterns.

The *dual scallop* variation describes a strategy in which the observer estimates the length of the line to the currently adopted working window, and then adds the remaining quantity to get to the target value. This creates a visible scallop in the data. The dual scallop power model is a variation of Equation 6 that accommodates a single working window and is defined as follows:

$$\text{if } (\varphi_I < d) \text{ then } (\Theta_L = \varphi_I{}^{\beta}) \text{ else } (\Theta_L = d^{\beta} + (\varphi_I - d)^{\beta}), \quad (7)$$

where d is the size of the working window of numbers. In this variation, the target minus d ($\varphi_I - d$) can be greater than d.

The final variation, the *multi-scallop power model*, specifies a strategy in which the observer successively estimates the length of the line to d until the remainder is less than d, and then adds the remaining quantity to get to the target value. The multi-scallop variation is a variation of Equation 6 that accommodates multiple working windows and is defined as follows:

$$\Theta_L = \text{truncate}(\varphi_I/d, 0) * d^{\beta} + (\varphi_I \text{ modulo } d)^{\beta}, \quad (8)$$

Collectively, these instantiations of the scallop power model have been successful in describing adult's and children's performance in the unbounded number-line task (Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014). We may therefore conclude that, in the unbounded number-line task, (i) the line/quantity bias is well described by Steven's Law, and, (ii) the dead reckoning strategy is well described by Cohen and Blanc-Goldhammer's (2011) application of Steven's Power Law.

2.2 Predictions

If (i) our model is correct, (ii) one models the data with the appropriate Equation (i.e., 5-8), and (iii) no other influences are at play, then the estimated beta parameters should be the same in the bounded and unbounded number-line tasks. Critically, though, it is expected that the estimates of beta may diverge under sub-optimal conditions. For example, Cohen and Sarnecka (2014) have shown that beta estimates in the bounded and unbounded number-line tasks diverged for young children, who did not have the mensuration skills necessary to implement the strategies required to successfully complete the bounded number-line task. It is also important to note that without a proper consideration of the operation of cognitive strategies, estimates of bias may carry little useful information about the underlying psychological representation of quantity (see e.g., Cohen & Quinlan, 2018).

In Experiment 1, we test the model predictions in the production (i.e., number-to-position) versions of the bounded and unbounded number-line tasks. In Experiment 2, we test the model predictions in the estimation (i.e., position-to-number) versions of the bounded and unbounded number-line tasks. In Experiment 2, we also varied the presentation time of the number line itself in a bid to examine performance under time pressure. A more detailed rationale of the time manipulation is presented in Experiment 2.

3 Experiment 1

In Experiment 1, we presented participants with production versions of the bounded and unbounded number-line task. In the current cases, responses were not time-limited, and as a consequence, there was ample opportunity for strategies to influence performance. The expectation was, therefore, that successful modelling of the data would involve successful modelling of such strategies.

3.1 Method

3.1.1 Participants

One hundred and sixty-three undergraduate volunteers from an introductory level psychology class participated for class credit. Sample size was determined by estimating a minimum number of participants to achieve .8 power to detect a medium effect size (126 participants given the estimated bias from Cohen and Blanc-Goldhammer, 2011). We then estimated the time necessary to collect that number of participants, posted all available experimental slots for the time estimated, and ran all participants who signed up. This procedure resulted in the collection of more than the minimum number of participants because of a higher than expected sign-up rate.

3.1.1 Apparatus and Stimuli

All stimuli were presented on 24-in. LED color monitors with 72-Hz refresh rates controlled by a Mac mini. The resolution of the monitor was 1,920 by 1,200 pixels.

Versions of the bounded and unbounded number-line tasks were generated (after Cohen & Blanc-Goldhammer, 2011). In both cases, the number-line was constructed from 1-pixel-thick red lines, with a 10-pixel-high vertical line (a tick mark) marking the start of the number-line. For the bounded number-lines, this left boundary was labeled with the number "0." A similar tick mark indicated the right end of the number-line and was labeled "22." The two tick marks were connected at the bottom by a red horizontal line (i.e., the number-line). For the unbounded number-lines, the left and right boundaries were unlabeled tick marks.[2] The tick marks were connected at the bottom by a horizontal line, the length of which represented the distance of one unit. Centered directly below the unit was the label "1."

In the production task, a target number was placed half an inch below the "0" at the left boundary, for the bounded number-lines, and half an inch below the "1", for the unbounded number-lines. The target numbers ranged from 2 to 21 and were chosen randomly from a uniform distribution from trial-to-trial in both practice and experimental trials.[3] Although such a range may seem unnatural, we explain our rationale for using targets 21 and under in the following.

With the unbounded number-line, the displayed line represented a single unit and for this to be to be visible, it had to be at least 2 pixels in length. Importantly, responses to the largest probe values can be truncated if the edge of the computer screen acts as a hard boundary. To prevent the edge of the computer screen acting as a hard boundary, one must provide sufficient space in the left side of the screen for the largest probe value plus the largest expected error associated with that probe. The size of this space was calculated accordingly; ((largestTarget)^bias + error)*LargestUnitSize. For our experiment the largestTarget = 21; the largestExpectedBias = 1.4 (see Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014); and the largestUnitSize = 32 (pixels). As error is scalar (see Cohen & Blanc-Goldhammer, 2011), responses will increasingly deviate from the probe value as this increases. If we exclude error, we needed about (21^1.4)*32 = 2271 pixels. As such, with a large bias and a target of 21, we could accommodate most single unit sizes, but for larger targets*larger unit sizes, responses would be hitting the edge of the screen. Thus, the maximum probe value we used was set at 21.

Without due consideration of these constraints, participants' responses to large target values will be correspondingly truncated. A consequence of this is that the response function begins to approximate the log-linear pattern (see Cohen & Quinlan, 2018). It appears that this physical task constraint is probably responsible for the recent results reported by Kim and Opfer

(2016, 2017) which show a log-linear response function in the unbounded number-line task. Specifically, they presented participants with displays in which the screen edge acted as a boundary identical to that of the bounded number-line. As such, “unbounded” number-line displays were, for all intents and purposes, identical to the bounded number-line displays.

To prevent participants from using reference points external to the number-line (e.g., the position of the left edge of the monitor or the center of the monitor), we varied the location and physical length of the number-line. The number-line was centered along the *y*-axis of the screen, and randomly placed between 100 and 200 pixels from the left side of the screen. The length of the single unit unbounded number-line was identical to the length of a single unit in the bounded number-line condition. If a participant extended the unbounded number-line without bias to the value of 22 (the right hand bound of the bounded number-line) the two types of number-lines would be of identical length. The length of the bound number-line varied between 42 and 704 pixels and the length of the single unit unbound number-line varied between 2 and 32 pixels. In addition, a strip of black tape was used to conceal the illuminated Apple symbol on the bottom of the monitor’s facia.

3.1.3 Procedure

Participants were tested individually in dedicated darkened testing rooms. Participants were instructed to identify the appropriate position of the quantity indicated by the target number on the number-line by using the click-and-drag functions of the mouse. In the bounded number-line task, the participant moved the cursor over the left tick mark, at which point, a grey line appeared covering the tick mark. The participant then pressed the left mouse button and dragged the grey line to the estimated target location. As the grey line was dragged, the left tick mark remained in place. In the unbounded number-line task, participants used the mouse to click-and-drag the terminal boundary line to the estimated target location. Just as with the bounded number-line, the click-and-drag method produced a grey tick mark. A red horizontal line connected the original unit to the grey tick mark. Participants extended the line out to the estimated target location. The line could not be dragged any closer than 100 pixels to the right edge of the screen. No response feedback was provided for either practice or experimental trials.

In both the bounded and unbounded number-line conditions, participants could freely move the grey line, dragging both right and left and releasing without submitting a response. When the participant determined the placement of the target line was accurate, he or she pressed the space bar to submit the response. The next trial appeared one second after the submission of the response. Participants' accuracy to the nearest pixel and response times (henceforth, RTs) were recorded.

Each participant was tested in one of two conditions: either the Production/Bounded version of the task, or Production/Unbounded version of the task. There were 12 practice trials and 360 experimental trials (on average, 18 trials per target). We required the large number of trials to permit fitting our model to individual participants. The participants had a self-timed break every 120 trials. The experiment typically lasted under one hour.

3.2 Results and Discussion

We identified the model that best fit each individual participant's data. Because the individual participant dataset contain relatively few points, and the models have relatively similar structures, it is important to reduce the probability of a spurious fit to a particular model. To accomplish this, for each participant's dataset, we conducted a robust analysis in two stages. First, we removed potential outlier points that would have a disproportionate influence on the data. Second, we fit the models using a bootstrap procedure that randomly sampled the participant's data, thus reducing the influence of spurious data points. We identified potential outliers in two ways. First, we removed all estimates in which the individual participant's RTs were greater than 45 s or less than 400 ms.[4] Second, we rounded targets to the nearest 5 (e.g., grouping targets 1-5 into a single bin, 6-10 into a single bin, etc.) and, within each bin, we removed estimates that were over three standard deviations from the mean error for that bin. These constraints eliminated less than 1.5% of the data.

To identify the best model fit for each participant by condition[5], we conducted a bootstrap procedure. For each participant by condition, we randomly sampled, with replacement, n trials, where n equals the total number of trials the participant ran in that condition (in Experiment 1, $n = 360$, because there was only 1 condition). From these trials, we calculated the mean estimate of each target value. For the bounded number-line task, we tested the fit of four models using generalized nonlinear least squares (gnls) methods. The models used were as follows: (i) a power function, (ii) a CPM with two reference points (i.e., where the reference points corresponded to the bounds of the number-line), (iii) a CPM with three reference points (i.e., the bounds and the midpoint of the number-line), and, (iv) a 50% mixture of 2 and 3 reference points CPM (see Hollands & Dyre, 2000). The 50% mixture model captures the situation whereby the participant sometimes uses 2 reference points and sometimes uses 3 reference points when completing the CPM strategy. We included this model because Hollands and Dyre (2000) showed it was a strategy participants adopt when completing similar tasks. For each of these models, β was the only free parameter and was unconstrained when fit to each model[6].

For the unbounded number-line task, the models used were as follows: (i) a power SPM, (ii) a dual SPM, and, (iii) a multi-SPM (see Cohen & Blanc-Goldhammer, 2011). For the power

SPM, β was the only free parameter. For the remaining SPM models, β and *d* (the working window) were the only free parameters. In all cases, the free parameters were unconstrained when fit to each model.

To reiterate, when β = 1 an observer's response is unbiased, and the corresponding response function is linear. Because the log-to-linear theory assumes adults represent increasing quantities in a linear fashion, it predicts that β will equal 1. In contrast, the power model will capture a logarithmic fit when β < 1 (though adults in college are presumed to have outgrown that fit) for both number-line tasks. The key evidence is with respect to estimates of β from the data of the various versions of the number-line task. The log-linear account predicts estimates of β to be 1: departures from a value of 1 are inconsistent with this and hence demand a different explanation. For each model fit, we calculated the Bayesian Information Criterion (BIC), which adjusts its value based on the number of free parameters and the r^2, which is a standard measure of effect size. If a model could not be fit on one of the runs or the fit resulted in an $r^2 <= 0$, all the parameters and fits statistics were assigned an NA (i.e., a null value indicating that the model could not be fit) for that replication[7]. Less than 1% of the participant by condition datasets resulted in NAs.

For each participant, we repeated the procedure above 100 times. We then calculated the mean of the BIC, r^2, and each parameter fit. The best-fit model for each participant by condition was determined by identifying the model with the lowest mean BIC, and, less than 5% of the runs resulting in an NA. The mean parameter fits and r^2 of the 100 runs for the best fit model was used as our dependent variables below.

The average r^2 of the best fit model was 0.95 (SD = 0.06). Given the superior fits of the models, we removed poor fitting models ($r^2 < 0.76$; those 3 SDs below the mean fit) from the dataset prior to the following analyses. This criterion removed less than 3% of the models. Table 1 presents the number (and percent) of participants' best fit by each model. Although all models were fit in the bounded variation of the task, only the power SPM model resulted in the best fit in the unbounded cases. This is likely because of the robust nature of the power model compared with the dualSPM and multiSPM and the relatively few data points in the individual participant's datasets.

Throughout we primarily report frequentist statistics (null hypothesis statistical testing), but we also include Bayes factors when they are informative. Specifically, we do not report Bayes factors when both the frequentist test statistic and the Bayes factor show strong evidence in the same direction (i.e., Bayes factor > 10 or < 0.1).

The models under consideration are all variations of Steven's Power Law. As such, the form of the line/quantity bias is described by the characteristic exponent, β. The overall average estimated beta of the production number-line tasks (M = 1.12, SD = 0.15) was significantly greater than 1, $t(158) = 10.66$, $p < 0.001$. This replicates the similar finding reported by Cohen and Blanc-Goldhammer (2011). A mixed-model ANOVA with participant as a random variable determined that the estimated beta of the bounded number-line task (M = 1.10, SD = 0.17) was significantly smaller than that of the unbounded number-line task (M = 1.15, SD = 0.11), $F(1, 157) = 4.07$, $p = .045$. We have very low confidence in this finding because the associated JZS Bayes factor (Rouder, Speckman, Sun & Morey, 2009) value is 1.074 in favor of the prediction that the bias in the unbounded task is greater than that in the bounded version. This is extremely weak evidence and, on these grounds, it is inadvisable to attribute any theoretical significance to it. Importantly, the estimated beta from the bounded number-line task (the variation that yielded the mean estimated beta closest to 1) was also significantly greater than 1, $t(81) = 5.30$, $p < .001$.

In sum, the estimated betas for both the bounded and unbounded production number-line tasks were significantly greater than 1. This basic finding fits comfortably with previous research with adults on the number-line task that has shown similar estimates of bias (Cohen & Blanc-Goldhammer, 2011; Slusser, et al., 2013). Critically, this result shows, yet again, that adults do not produce linear responses to the production version of the number-line task.

In addition, although the data also revealed that the estimated betas for the bounded and unbounded production number-line task were significantly different, the associated Bayes factor suggests the evidence is extremely weak and therefore little confidence should be assigned to it. We will return to this point in the General Discussion. Next, we assess the validity of our model by examining performance in estimation rather than production versions of the number-line task.

4 Experiment 2

Two modifications to the procedure used in Experiment 1 were implemented in Experiment 2. First, we used an estimation (position-to-number) task, and, second, we examined performance in this task under varying degrees of time pressure. In the estimation version of the task, participants are presented the number-line as well as an indicator of the position of the target (e.g., a "tick mark"). In this task, the participant responds by providing a numerical estimate that corresponds to the position of the tick mark on the line. Because the estimation task does not require the participants to physically interact with the line, it provides an opportunity to explore the influence of time limitations on participants' responses. As such, we systematically varied the presentation time across three conditions, that is, presentation time was either 500 ms, 1 s, or 3 s.

Here, we test a strong prediction of the model that the estimation task should manifest equivalent biases as the production task when we take the inverse of the estimated bias

($1/\beta_{estimation}$). A weaker prediction – one that assumes there are as yet unidentified factors in the model – predicts that the inverse of the estimated biases in both task should remain greater than 1. That is, the direction of the bias should remain constant (i.e., negatively accelerating). The model makes no specific predictions concerning the time limitations. Nevertheless, if one assumes that the fundamental perceptual bias underlying the number-line task (the line/quantity bias) is unaffected by time limitations, then the model predicts that the estimated beta should be constant across time presentation conditions – even if the time limitations influence how the participants implement the strategies (e.g., by finding a fewer or inconsistent number of reference points in the CPM, etc.). If one assumes that time limitations influences the line/quantity bias, then the model should be able to accurately estimate that influence, even if time limitations also influence how participants implement the strategies.

4.1 Method

4.1.1 Participants

One hundred and sixty-one undergraduate volunteers from an introductory level psychology class participated for class credit. The sample size was determined by the same procedure described in Experiment 1.

4.1.2 Apparatus and Stimuli

The apparatus and stimuli were identical to Experiment 1, with the following exceptions. In the estimation task, no target number was presented. Instead, a small vertical grey target line was located within the end bounds of the number-line for the bounded number-line, and beyond the terminal right bound of the unbounded number-line. The value indicated by the target line was a whole number ranging from 2 to 21 and was chosen randomly from a uniform distribution from trial-to-trial.

4.1.3 Procedure

The procedure was identical to Experiment 1, with the following exceptions. In the estimation task, participants were instructed to estimate the value on the number-line portrayed by the target line by inputting the numerical value into a dialogue box. Each trial began with the left vertical boundary presented for 500 ms, then the rest of the number-line became visible for a pre-determined amount of time. For the experimental trials, the duration was randomly selected to be 500, 1000, or 3000 ms (we did not use the 500 ms duration in the practice trials because we thought that it may be too quick of a presentation for the participants to learn the task). After the presentation time elapsed, a mask consisting of random lines of the same color as the number-line was presented for 1000 ms. This was followed by the dialogue box. The dialogue box allowed participants to input values between 00.1 and 99.9. That is, the response dialogue box allowed any number that contained up to two integers and one decimal place. The dialogue box remained on the screen until the participant pressed the "OK" button on the dialogue box and then a new trial began. Each participant was tested in one of two conditions, that is, in either the bounded version or the unbounded version of the number-line task.

4.2 Results and Discussion

We scored and analyzed the data using the same procedures as Experiment 1. The outlier constraints eliminated less than 1.5% of the data.

For each participant by condition by presentation time, we bootstrapped the models to the data as we did in Experiment 1. Because the expected bias from the estimation task is the inverse of that from the production task, we transformed the estimated biases from the estimation task by taking their inverse (i.e., $1/\beta_{estimation}$). In this way, the estimated betas from the Experiment 2 are directly comparable to that of Experiment 1.

The average r^2 of the best fit model was 0.87 (*SD* = 0.14). Similar to Experiment 1, poor fitting models ($r^2 < 0.46$) were removed from the dataset prior to the following analyses. This criterion removed 3% of the models and another 3% of the models resulted in NAs from the bootstrap procedure. Table 1 presents the number of participants best fit by each model.

To assess the influence of presentation time and number-line task on estimated beta we first computed the estimated beta from the best fitting model for each participant for each conditions of interest. These data were then entered into a 3 (presentation time) x 2 (number-line task: bounded vs. unbounded) mixed model ANOVA, with participant as a random variable. There was a significant effect of number-line task, $F(1, 150) = 27.26$, $p < .001$, with the bounded number-line task (*M* = 1.27, *SD* = 0.23) producing a larger estimated beta than the unbounded (*M* = 1.14, *SD* = 0.11) number-line task. There was no significant effect of presentation time, $F(2, 296) = 0.31$, *ns*. Nor was there a significant interaction between presentation time and number-line task, $F(2, 296) = 0.13$, *ns*. Importantly, the smallest average beta (i.e., that for the unbounded condition) was significantly greater than 1, $t(74) = 11.59$, $p < .001$.

Cross-experiment comparisons allow us to examine the size of the betas across the production and estimation versions of the task (see Figure 4). The average value of the estimated betas for the bounded number-line task was greater in the estimation task than the production task, $t(189) = 7.01$, $p < .001$. In contrast the average value of the estimated betas for the unbounded number-line task was not significantly different than those from than the production task, $t(132) = 0.51$, *ns* (Bayes factor = 0.11 – moderate to strong evidence in favor of the null).

5 General Discussion

Here, we have presented a general model of number-line completion that is based on Steven's Power Law. In our model, the estimated beta quantifies a conflation of two psychological biases associated with the perception of quantity and lines, respectively. We refer to the estimated betas as reflecting a line/quantity bias. The two biases are conflated in the current model because neither our model, nor any other mathematical model of number-line task completion, disambiguates them.

The model, if valid, will produce similar estimates of the underlying line/quantity bias that drives participants' responses in all variations of the number-line task. We tested the predictions of this model in two experiments that compared participants' production and estimation responses in the bounded and unbounded versions of the number-line task. The model's estimates of line/quantity bias across all variations of the task were similar (e.g., about 1.12). The only exception occurred with the bias for the estimation variation of the bounded number-line task. More specifically, the estimated line/quantity bias from the unbounded number-line task followed the predictions of the model flawlessly. The estimated line/quantity bias from the bounded number-line task were slightly underestimated in the production task and overestimated in the estimation task. We discuss these findings below.

Our cross-domain model of number-line estimation predicts that the data patterns produced by different variations of the number-line task are a function of strategy, rather than people's underlying psychological representation of quantity. In cross-domain tasks, participants are presented with a standard that provides the link between the two domains. In the number-line task, the number-line itself is the standard whereby the overall length of the line is assigned to the value of the right hand bound. First, because the estimation task and the production task simply reverse the stimulus and response in the cross-domain task, they should produce inverse estimates of line/quantity bias. Second, the bounded and unbounded number-line tasks are simply variations of the cross-domain matching task whose differences influence participant strategy. In the bounded number-line task, the right hand bound sets a hard limit on the greatest target value. In contrast, in the unbounded number-line task the right hand bound is smaller than all target values. This difference dictates the most appropriate strategy in each task. When the target value is a fraction of the standard value (numerical or line length), the participant must use either subtraction or division to equate line length to number value. This is the case for the bounded number-line. In contrast, when the target value is a multiple of the standard value (numerical or line length), the participant must use either addition or multiplication to equate line length to number value. This is the case for the unbounded number-line. Data from the bounded number-line task is often best fit by the CPM, which models a subtraction/division strategy. The data from the unbounded number-line task is often best fit by the SPM, which models an addition strategy. The data from the current experiments accords well with these conclusions (see Table 1). Although both tasks produce very different patterns of data, the line/quantity bias underlying performance should be the same. Therefore, if the formulas used to estimate beta from the data are accurate, the estimated betas in both versions of the task should be equivalent.

In Experiments 1 and 2, the estimates of the line/quantity biases from the production and estimation versions of the unbounded number-line task were virtually identical. Furthermore, the estimated biases were not influenced by presentation time in the estimation variation. These findings support our proposed model and we claim therefore that that it captures the underlying processes driving completion of the unbounded number-line task.

In Experiments 1 and 2, the estimates of the line/quantity biases from the production and estimation versions of the bounded number-line task revealed partial support for the model. Specifically, the production version of the bounded number-line task produced estimates of line/quantity biases similar to those of the unbounded number-line task. This further supports our proposed model. However, the estimation version of the bounded number-line task produced over-estimates of line/quantity biases. This divergence in our estimates suggests that the estimates of bias in the bounded number-line task may be unstable. This, in turn, implies that we do not fully understand the cognitive mechanisms underlying performance in the bounded number-line task.

One possible source for the deviations of the betas from the model prediction in the bounded number-line task is that there may be a systematic bias in the CPM model fit. Cohen and Quinlan (2018) ran a simulation of the production variation of the bounded number-line task whereby they simulated specific quantity representations and completion strategies, and fit the CPM. They found that the estimated betas from the CPM fits systematically underestimated the biases present in the simulated quantity representations. Because beta in the production variation is equal to the inverse (1/beta) in the estimation variation, the CPM's bias to underestimate the beta in the production task results in the overestimation of beta in the estimation task. This underestimation in the production task and overestimation in the estimation task is exactly what we observed in our bounded number-line data. As a consequence, it is likely that the CPM model produces somewhat biased estimates of beta.

Another possible source for the deviations of the betas from the model prediction in the bounded number-line task is the potential of a *direction* bias. That is, in Experiment 1 the target tick mark was "grabbed" at the left-hand boundary and dragged right. Of course, the participant could move the tick mark freely from there (e.g., they could move it freely both right and left within the bounds of the number-line before they submitted their response). The consistent starting position was implemented in the bounded number-line task so that it matched the starting drag position of the unbounded number-line task. Nevertheless, some might be concerned that this start position may bias the participants' responses in the bounded number-line task. In our model, the influence of dragging to the right vs, left would be a response bias (see Equation 2) and would manifest in the line/quantity bias. There are two reasons to suspect that a *direction* bias does not influence the data. First, Karolis, Iuculano, & Butterworth, (2011) directly tested for a direction bias in the bounded number-line task and found no evidence for such a bias. Second, in the estimation version of the number-line task, a direction bias cannot influence the line/quantity bias because the target tick mark is presented as part of the number-line. Therefore, the difference between the estimated betas of the production (Experiment 1) and estimation (Experiment 2) versions of the task can give some indication of the presence of a direction bias. We found no difference between the estimated betas in the production and estimation versions of the unbounded number-line task. Because there is no a priori reason to suspect a direction bias only in the bounded number line task, this suggests that a direction bias did not influence our data. Nevertheless, researchers may want to explore this possibility in future.

Data from the number-line task have been used to make claims about the underlying psychological representation of quantity (Dehaene et al., 2008; Opfer et al., 2016; Siegler & Booth, 2004; Siegler & Opfer, 2003). In contrast via our model we can make explicit the following counter points: (i) no account of performance on the number-line tasks can be complete without taking account of the strategy that is deployed to estimate perceptual bias. In our model this is addressed via incorporating β, and, (ii) the bias so estimated confounds the quantity bias and line bias. Therefore, to make claims regarding the quantity bias alone, one must assume the line bias equals 1 (β_{line}=1). Whether this is a sensible assumption and how best to test this are empirical questions for future research.

Here we emphasize that in all conditions, our data revealed estimated betas that are greater than 1. Because our analysis explicitly assessed whether participants produced a linear response function (i.e., a function whose $\beta = 1$), we were able to assess the validity of the log-to-linear hypothesis if we assume $\beta_{line}=1$. Assuming $\beta_{line}=1$ is a common, unstated assumption in the number-line literature, and an assumption that researchers' supporting the log-to-linear hypothesis implicitly make. We do not claim this is a valid assumption. Rather, we simply adopted the assumption as being true so that we could assess the log-to-linear hypothesis on its own terms. The log-to-linear hypothesis predicts that adults should produce a linear pattern of results in the number-line task. Furthermore, if a bias is present, the theory predicts that the bias should result in a negatively accelerating bias (i.e., $\beta < 1$) for those who were unable to switch from a logarithmic representation to a linear representation. Our results show that adults produce a positively accelerating bias and this stands in stark contrast to the predictions of the log-to-linear hypothesis.

The positively accelerating bias identified in Experiments 1 and 2 accords well with the findings of Cohen and Blanc-Goldhammer (2011), but is inconsistent with the earlier literature (Berteletti, Piazza, Dehaene & Zorzi, 2010; Booth & Siegler, 2006; Geary, Hoard, Nugent & Byrd-Craven, 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003; Thompson & Opfer, 2008). Prior to Cohen and Blanc-Goldhammer (2011), researchers assessed children's general performance in the bounded number-line task but did not attempt to fit their data with specific models such as the CPM. Instead they estimated biases as either logarithmic (negatively accelerating) or linear (Berteletti, et al., 2010; Booth & Siegler, 2006; Geary et al., 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003; Thompson & Opfer, 2008). In the adult data reported by Cohen and Blanc-Goldhammer (2011), the estimated biases in the production versions of the tasks using the CPM and SPM were positively accelerating and averaged about 1.13 (very similar to the average production bias obtained in the current production experiment – Experiment 1). We note that positively accelerating bias identified in Experiments 1 and 2 does not fit well with the traditionally hypothesized psychological representations of quantity discussed in the introduction. Nevertheless, Cohen and Blanc-Goldhammer (2011) present an example psychological representation of quantity based on their data ($\beta_{quantity}$ = 1.13 assuming that β_{line} = 1.0). The presented example is very similar to the scalar variance model, with the only deviation being that the mean placement of the distributions has a slight positive acceleration rather than being linear. As such, this proposed quantity representation may be considered a variation of the scalar variance theory.

6 Conclusions

In sum, the present paper contains a detailed and unified mathematical model of number-line task completion that assumes that the line/quantity bias is well described by Steven's Power law. We posit that data from number-line tasks reflect a person's underlying representation of quantity, together with the cognitive strategies and skills required to be successful in equating line length and quantity. We have provided further evidence that the bounded and unbounded number-line tasks, respectively, invoke qualitatively different strategies. We demonstrate that when the participant's cognitive strategy is adequately modeled, the two variations of the number-line task manifest equivalent estimates of perceptual bias. Performance on the unbounded number-line task fits perfectly with the model as described. Systematic deviations from the model arise in the data for performance on the bounded number-line task. Such deviations are understood when a careful analysis of the corresponding task constraints associated with the bounded number-line task is undertaken.

Critically we have cautioned about using data from the bounded number-line as a means to examine theories of numerical cognition. Performance on this task may well reflect a complex interaction between, on the one hand, of number processing and, on the other, task constraints that have nothing to do with number processing. Strikingly different estimates of beta arise across the production and estimation versions of the bounded number-line task. More transparent answers to questions about number processing are forthcoming when unbounded versions of the task are used (Cohen & Blanc-Goldhammer, 2011; Cohen & Quinlan, 2018). Performance on variants of the unbounded number-line task is captured well by a unified model of number processing that accommodates both production and estimation capabilities that assumes a common psychological representation of quantity.

References

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14*, 545-551. doi: http://dx.doi.org/10.1111/j.1467-7687.2010.00962.x

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology, 46*, 545-551. doi: http://dx.doi.org/10.1037/a0017887

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure number estimation. *Developmental Psychology, 41*, 189-201. doi: http://dx.doi.org/10.1037/0012-1649.41.6.189

Brooke, J. B., & MacRae, A. W. (1977). Error patterns in the judgment and production of numerical proportions. *Perception & Psychophysics, 21*, 336-340. doi: http://dx.doi.org/10.3758/BF03199483

Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on "Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures". *Science, 323*. doi: http://dx.doi.org/10.1126/science.1164773

Chesney, D. L., & Matthews, P. G. (2013). Knowledge on the line: Manipulating beliefs about the magnitudes of symbolic numbers affects the linearity of line estimation tasks. *Psychonomic Bulletin & Review, 20*, 1146-1153. doi: http://dx.doi.org/10.3758/s13423-013-0446-8

Church, R. M., Meck, W. H., & Gibbon, J. (1983). Application of scalar timing theory to individual trials. *Journal of Experimental Psychology: Animal Behavior Processes, 9*, 135-155. doi: http://dx.doi.org/10.1037/0097-7403.20.2.135

Cohen, D. J. (2009). Integers do not automatically activate their quantity representation. *Psychonomic Bulletin & Review, 16*, 332-336. doi: http://dx.doi.org/10.3758/PBR.16.2.332

Cohen, D. J. (2010). Evidence for Direct Retrieval of Relative Quantity Information in a Quantity Judgment Task: Decimals, Integers, and the Role of Physical Similarity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36*, 1389-1398. doi: http://dx.doi.org/10.1037/a0020212

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

Cohen, D. J., Farrell, J., & Johnson, N. (2002). What very small numbers mean. *Journal of Experimental Psychology: General, 131*, 424-442. doi: http://dx.doi.org/10.1037//0096-3445.131.3.424

Cohen, D. J., & Quinlan, P. T. (2016). How numbers mean: Comparing random walk models of numerical cognition varying both encoding processes and underlying quantity representations. *Cognitive Psychology, 91*, 63-81. doi: http://dx.doi.org/10.1016/j.cogpsych.2016.10.002

Cohen, D. J., & Quinlan, P. T. (2018). The log-linear response function of the bounded number-line task is unrelated to the psychological representation of quantity. *Psychonomic Bulletin & Review, 25*, 447-454. doi: https://doi.org/10.3758/s13423-017-1290-z

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

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1-42. doi: http://dx.doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S. (2003). The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends in Cognitive Science, 7*, 145-147. doi: http://dx.doi.org/10.1016/S1364-6613(03)00055-X

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 314-326. doi: http://dx.doi.org/10.1037/0278-7393.21.2.314

Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: A case study of severe acalculia with preserved approximation. *Neuropsychologia, 29*, 1045-1074. doi: http://dx.doi.org/10.1016/0028-3932(91)90076-K

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 626-641. doi: http://dx.doi.org/10.1037/0096-1523.16.3.626

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

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5-to 9-year old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology, 99*, 1-17. doi: http://dx.doi.org/10.1016/j.jecp.2007.08.006

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 43-74. doi: http://dx.doi.org/10.1016/0010-0277(92)90050-R

Geary, D., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neurobiology, 33*, 277-299. doi: http://dx.doi.org/10.1080/87565640801982361

Gescheider, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology, 39*, 169-200.

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665-668. doi: http://dx.doi.org/10.1038/nature07246

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

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

Kim, D., & Opfer, J. E. (2016). *A Unified Framework for Bounded and Unbounded Numerical Estimation*. Paper presented at the Proceedings of the 38th Annual Conference of the Cognitive Science Society, Austin, TX.

Link, T., Nuerk, H. C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *The Quarterly Journal of Experimental Psychology, 67*, 1597-1613. doi: http://dx.doi.org/10.1080/17470218.2014.892517

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9*, 320-334. doi: http://dx.doi.org/10.1037/0097-7403.9.3.320

Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 11*, 591-597. doi: http://dx.doi.org/10.1037/0097-7403.11.4.591

Nuerk, H. C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition, 82*, B25-B33. doi: http://dx.doi.org/10.1016/S0010-0277(01)00142-1

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

Peeters, D., Sekeris, E., Verschaffel, L., & Luwel, K. (2017). Evaluating the effect of labeled benchmarks on children's number line estimation performance and strategy use. *Frontiers in Psychology, 8*. doi: https://doi.org/10.3389/fpsyg.2017.01082

Peeters, D., Verschaffel, L., & Luwel, K. (2017). Benchmark-based strategies in whole number line estimation. *British Journal of Psychology, 108*, 668-686. doi: http://dx.doi.org/10.1111/bjop.12233

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

Rouder, J. N., Speckman, P. L., Sun, D., & Morey, R. D. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225-237.

Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology, 24*, 21-57. doi: http://dx.doi.org/10.1016/0022-2496(81)90034-1

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

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

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The Logarithmic-To-Linear Shift: One Learning Sequence, Many Tasks, Many Time Scales. *Mind, Brain and Education, 3*, 143-150. doi: http://dx.doi.org/10.1111/j.1751-228X.2009.01064.x

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

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

Stevens, S. S. (1956). The direct estimation of sensory magnitudes-loudness. *The American Journal of Psychology, 69*, 1-25. doi: http://www.jstor.org/stable/1418112

Table 1

The number of participants who were best fit by each model in each number-line presentation condition.

	Unbounded							
			Estimation (presentation time in ms)					
	Production		500		1000		3000	
	n	%	n	%	n	%	n	%
Power SPM	77	100	72	100	74	100	75	100
	Bounded							
			Estimation (presentation time in ms)					
	Production		500		1000		3000	
	n	%	n	%	n	%	n	%
Power	34	41	16	21	10	13	14	18
One-Cycle CPM	21	26	13	17	21	27	19	25
Mixed CPM	11	13	36	47	29	38	31	40
Two-Cycle CPM	16	20	12	16	17	22	13	17

Figure Captions

Figure 1. A schematic of the different biases associated with the number-line tasks. Here, all digits and lines are subject to an encoding bias and theoretically transformed into a psychological quantity representation. The number-line tasks require, however, mathematical operations to be successfully completed, and these are represented in the diagram as well. The mathematical operations may be influenced by the accuracy of the quantity representation as well as the participant's mathematical knowledge.

Figure 2. Examples of the pattern of data produced by the Cyclic Power Model for $\beta > 1$ (solid lines) and $\beta < 1$ (dashed lines). The shades of gray indicate the number of reference points.

Figure 3. Examples of the pattern of data produced by the Scalloped Power Model for $\beta > 1$ (solid lines) and $\beta < 1$ (dashed lines). The shades of gray indicate the number of scallops.

Figure 4. The estimated beta from the bounded and unbounded estimation number-line task for the Production and Estimation tasks. The error bars represent 1 SE above and below the mean.

Footnotes

[1] For clarity, we discuss the perceptual distributions as if they are discrete. This simplification is practical because (i) a specific stimulus (e.g., a line of a certain length) is presented to a participant, even though line length is on a continuum, and (ii) the simplification does not change the interpretation of the beta parameter.

[2] We excluded the "0" label because it interfered with the "1" label when the physical distance of the unit was small. The participants understood the task well enough without the physical reminder of the value of the left boundary.

[3]In fact, in Cohen and Blanc-Goldhammer (2011) the right boundary was labeled "26" but the authors found that participants produced estimates with the unbound number-line that hit the edge of the screen for target values 23-25.

[4] The 45 second value gave rise to the removal of only the most extreme outliers (see Cohen and Blanc-Goldhammer, 2011). The 400 ms is too fast to perceive and complete the task.

[5] In Experiment 1, each participant was tested once in a single condition. In Experiment 2, however, each participant was tested across three conditions (presentation times of 500, 1000, and 3000). The analysis was run separately for each condition.

[6] The reference points for all the CPMs were fixed at the lower bound, upper bound, and midpoint (for the three-reference point CPM). For the mixture model, the percent mixture was fixed at 0.5. The highly constrained nature of these formulas are a strength of the model. These constraints reduce the likelihood of fitting the data simply by chance. If one were to convert these fixed parameters to free parameters, then the formulas would be more flexible. However, that flexibility would be penalized in the model fit statistics (i.e., as reflected in the BIC).

---

[7] Negative r^2s result when the model accounts for less variance than the mean. Because our non-linear models are highly constrained (they often only have a single free parameter), they are unlikely to fit the data well simply by chance. Rather, our constraints may occasionally result in a fit that accounts for less variance than the mean.

Figure 1

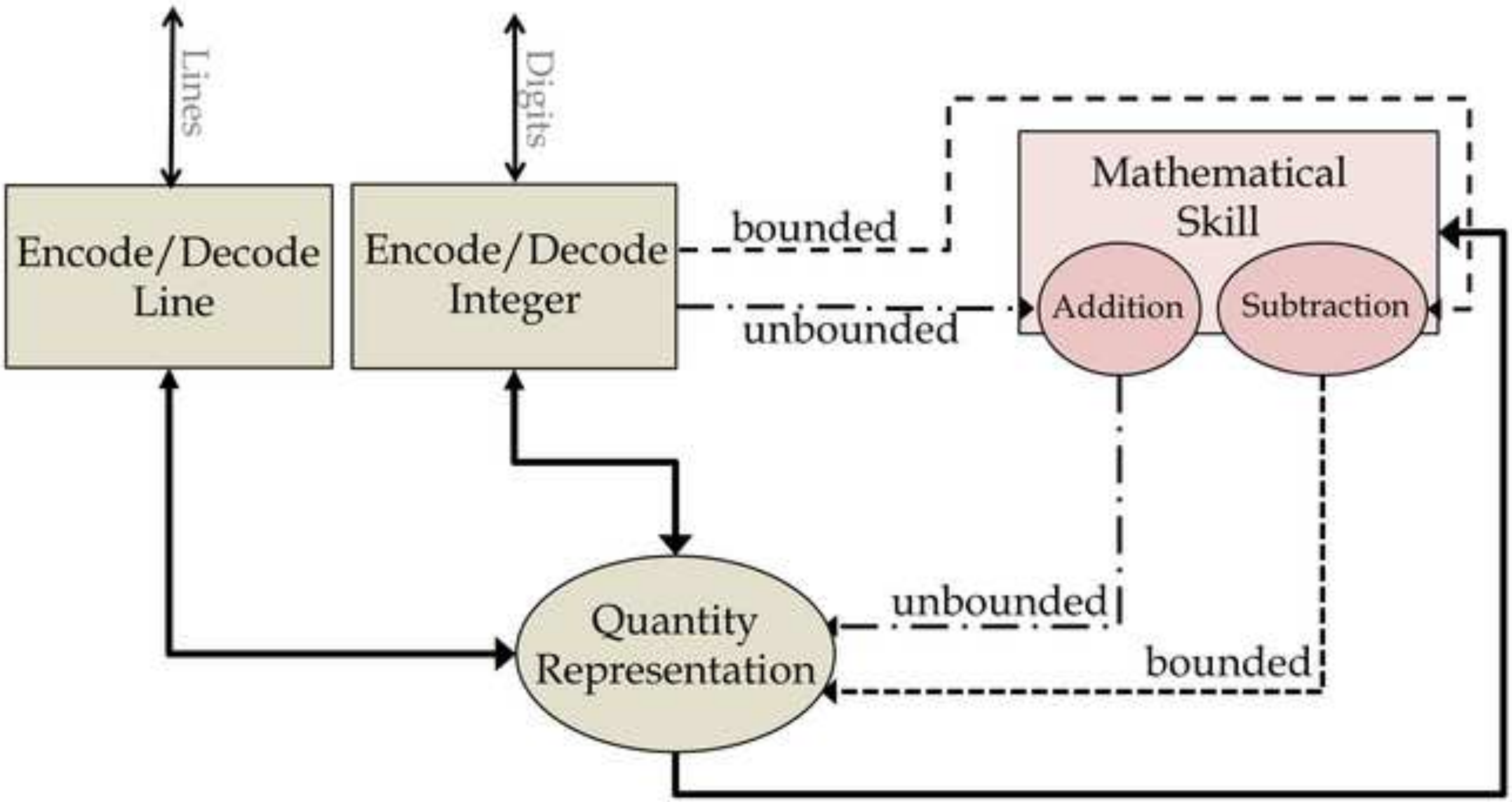
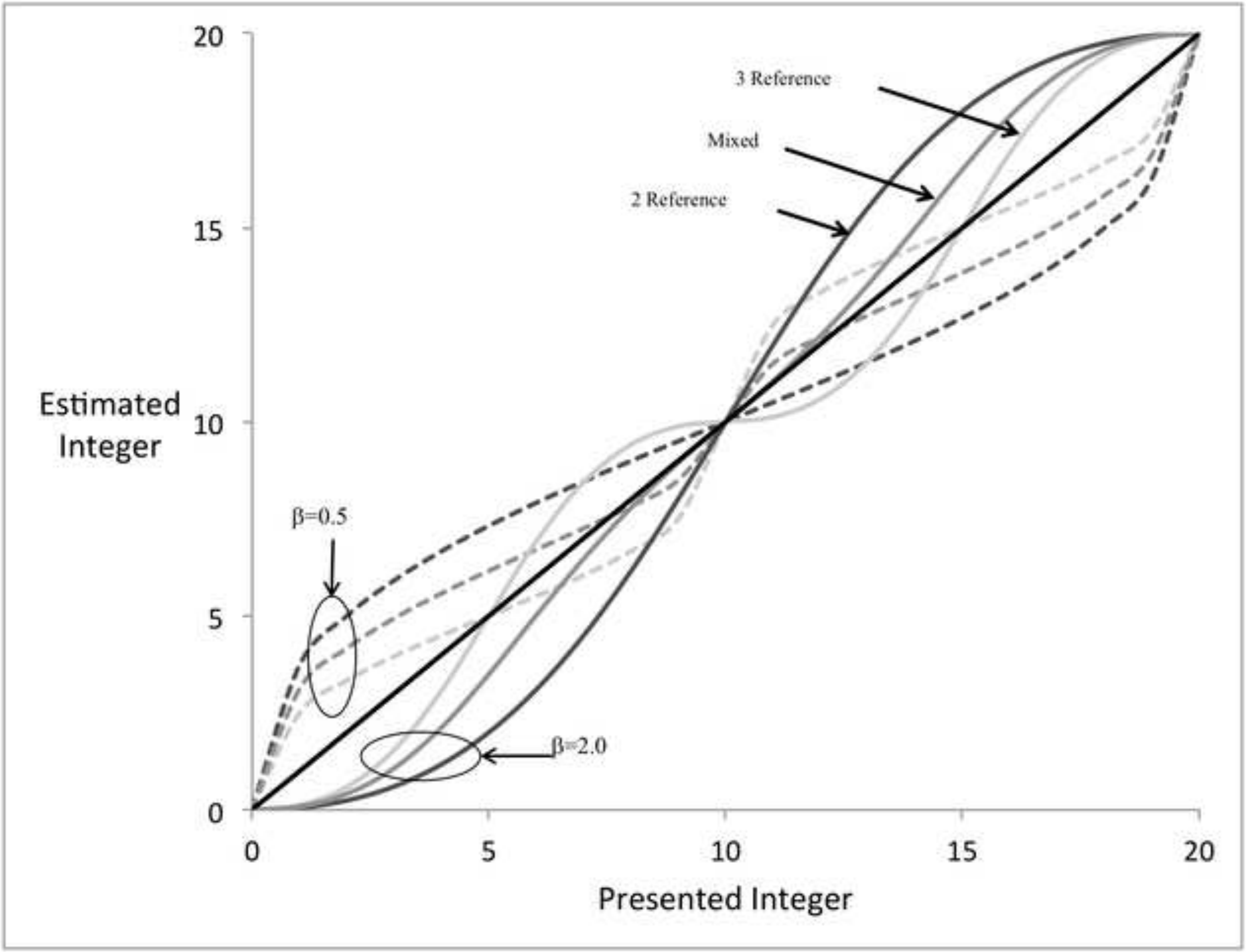

Lines
Digits
Encode/Decode Line
Encode/Decode Integer
bounded
unbounded
Mathematical Skill
Addition
Subtraction
Quantity Representation
unbounded
bounded

Figure 2



3 Reference
Mixed
2 Reference
β=0.5
β=2.0
Estimated Integer
Presented Integer
0
5
10
15
20
0
5
10
15
20

Figure 3

Click here to access/download;Figure;Figure3.tiff

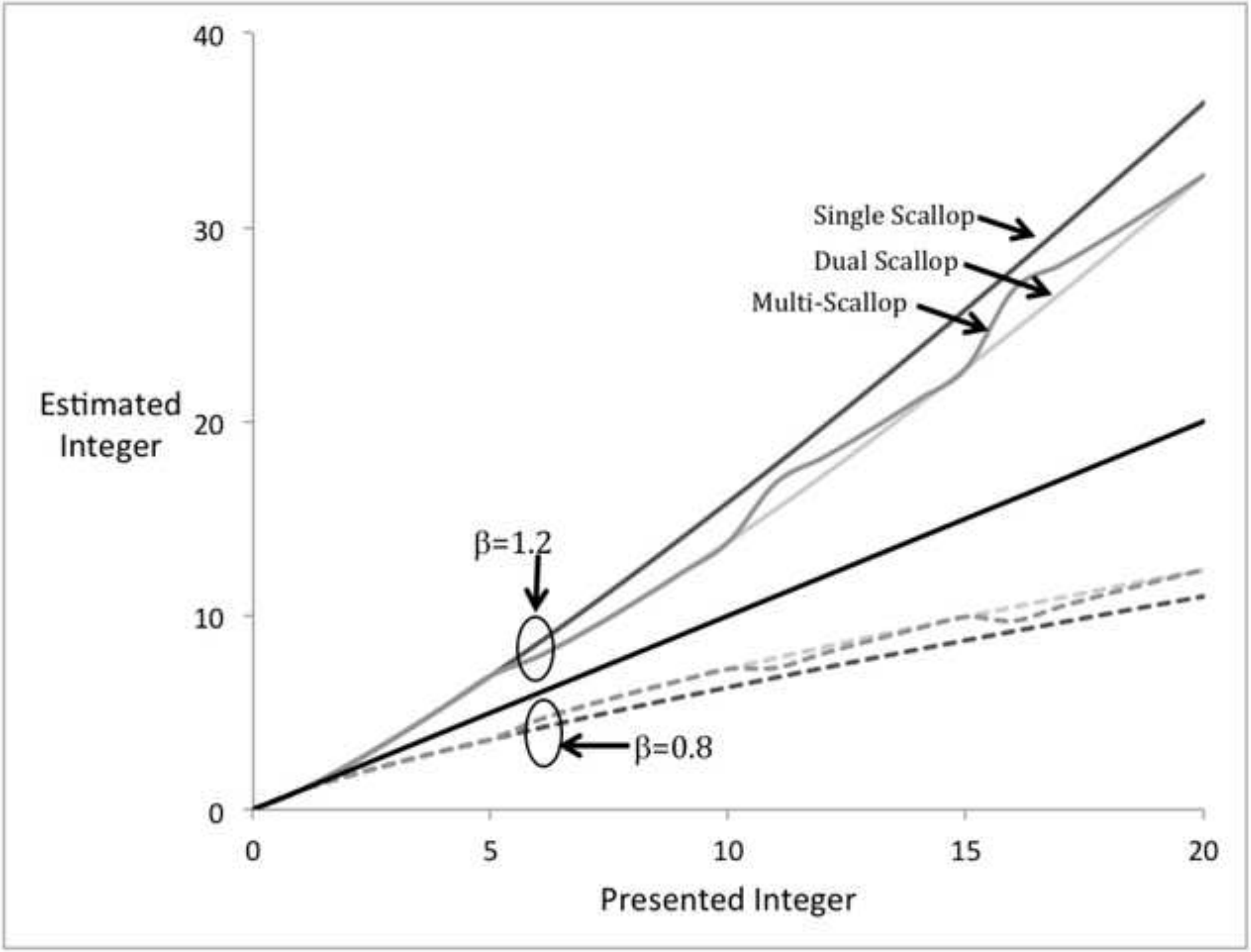


Figure 4