Journal of Experimental Psychology: General

Mapping Numerical Magnitudes Along the Right Lines: Differentiating Between Scale and Bias

Vyacheslav Karolis, Teresa Iuculano, and Brian Butterworth

Online First Publication, July 18, 2011. doi: 10.1037/a0024255

Journal of Experimental Psychology: General

0096-3445/11/$12.00 DOI: 10.1037/a0024255

Mapping Numerical Magnitudes Along the Right Lines: Differentiating Between Scale and Bias

Vyacheslav Karolis, Teresa Iuculano, and Brian Butterworth
University College London

Previous investigations on the subjective scale of numerical representations assumed that the scale type can be inferred directly from stimulus–response mapping. This is not a valid assumption, as mapping from the subjective scale into behavior may be nonlinear and/or distorted by response bias. Here we present a method for differentiating between logarithmic and linear hypotheses robust to the effect of distorting processes. The method exploits the idea that a scale is defined by transformational rules and that combinatorial operations with stimulus magnitudes should be closed under admissible transformations on the subjective scale. The method was implemented with novel variants of the number line task. In the line-marking task, participants marked the position of an Arabic numeral within an interval defined by various starting numbers and lengths. In the line construction task, participants constructed an interval given its part. Two alternative approaches to the data analysis, numerical and analytical, were used to evaluate the linear and log components. Our results are consistent with the linear hypothesis about the subjective scale with responses affected by a bias to overestimate small magnitudes and underestimate large magnitudes. We also observed that in the line-marking task, participants tended to overestimate as the interval start increased, and in the line construction task, they tended to overconstruct as the interval length increased. This finding suggests that magnitudes were encoded differently in the 2 tasks: in terms of their absolute magnitudes in the line-marking task and in terms of numerical differences in the line construction task.

Keywords: subjective scale, response bias, admissible transformations, central tendency effect, logarithmic compression

It has long been noted that number processing reveals a striking similarity to processing physical magnitudes (Moyer & Landauer, 1967; Restle, 1970). Both are subject to the *distance effect*; that is, longer response times are required to distinguish between close magnitudes than distant ones (Moyer & Landauer, 1967). Both are also affected by the *size effect*; that is, it takes more time to distinguish between large magnitudes relative to small magnitudes when they are separated with the same numerical distance (Buckley & Gillman, 1974; Parkman, 1971; Parkman & Groen, 1971). Furthermore, converging evidence suggests a strong association between small numbers and the left side of space and large numbers and the right side of space (Dehaene, Bossini, & Giraux, 1993; Zorzi, Priftis, & Umiltà, 2002), which was interpreted to mean that subjective number representations have an implicit and unique spatial architecture (Hubbard, Piazza, Pinel, & Dehaene, 2005; Izard & Dehaene, 2008).

These findings justify the use of psychophysical methods in the study of numerical cognition (e.g., Izard & Dehaene, 2008; Siegler & Opfer, 2003) and raise the question about the internal (or subjective) scale of numerical representations. A unifying framework, accounting for the above phenomena, treats numerical magnitude as a univariate random Gaussian variable (Izard & Dehaene, 2008; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004) that maps in an increasing order from left to right onto a subjective continuum: the "mental number line" (Dehaene, 1992). The distance effect, then, is the result of an increasing overlap between distributions for numbers with close magnitudes as compared with an overlap between distributions for far magnitudes. Two hypotheses were proposed to account for the size effect. According to one, the scale of mental number line is logarithmically compressed (Dehaene, 2003). As a result of such compression, spacing between two neighboring numbers should decrease with an increase of their magnitudes, leading to a greater overlap of distributions. An alternative hypothesis proposes that the subjective scale is linear but the noisiness of mapping increases with number magnitude, also leading to a greater overlap (Gallistel & Gelman, 1992; Whalen, Gallistel, & Gelman, 1999).

Because both linear and logarithmic hypotheses can predict similar outcomes (Dehaene, 2003), distinguishing between them is a nontrivial empirical problem. An attempt to solve this problem has been made by studies exploiting the number-to-location mapping paradigm. For example, Siegler and Opfer (2003) presented participants selected from four age groups (second, fourth, and sixth graders and adults) with a line labeled with "0" at one end

Vyacheslav Karolis, Teresa Iuculano, and Brian Butterworth, Institute of Cognitive Neuroscience, University College London, London, England.

We are grateful to Marinella Cappelletti, Dmitry Laptev, and Jorn Diedrichsen for their advice and valuable comments during the preparation of this article.

Correspondence concerning this article should be addressed to Vyacheslav Karolis, Institute of Cognitive Neuroscience, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, England. E-mail: ucjtvka@ucl.ac.uk or karoleezz@yahoo.co.uk

and "100" or "1,000" at the other and asked them to mark on the line the magnitudes of numbers contained within those intervals. The study showed that the younger groups (second and fourth graders) exhibited responses that were best modeled by a logarithmic function, whereas older children and adults used linear mapping. The authors suggested that children initially possess a logarithmic subjective scale associated with a more primitive sense of numbers (Dehaene, 1997), and that the scale becomes linear at later stages of development through the use of counting series and language.

There are a number of reasons why these findings may not be conclusive. First, the study uses numerical intervals that are standards for the decimal counting and metric systems. These intervals may be highly overtrained through education and life experience. The near 100% of variance explained by the linear fit in adult data suggests that the task was very easy, giving rise to a ceiling effect. Second, it is possible that the access to the magnitude of a number is not required for an adult to correctly perform in this task. In fact, the problem can be solved algorithmically on the basis of precise ordering provided by the counting system. For example, the interval can be roughly partitioned into parts hallmarked by the multiples of 10, and the required number between them can be found by interpolation.

Other studies suggest that a logarithmic component persists in the subjective scale of adults if approximate estimation is required. Thus, Dehaene, Izard, Spelke, and Pica (2008) asked participants to rate the magnitude of a nonsymbolic numerosity (dots or tones) on the line bracketed with either 1 and 10 or 10 and 100 dots. They showed that small nonsymbolic numerosities up to 10 items, which could be easily counted, were rated linearly. By contrast, rating large nonsymbolic numerosities (10–100) exhibited a significant logarithmic component, suggesting that the scale for approximate estimation is not completely linearized.

The evidence for the logarithmic compression of nonsymbolic magnitudes can also be found in dot enumeration. It has been shown (Izard & Dehaene, 2008; Krueger, 1982) that mapping from a number of dots to digits obeys Stevens power law (Stevens, 1957); that is, the relation between stimulus numerosity D and participants' response N is captured by a power function $N = \alpha D^{\beta}$, with exponent $\beta < 1$. This form of mapping is consistent with the idea that both dependent and independent variables are the logarithmic interval scales (in other words, there is a linear relation between their logarithms, $\log N = \log \alpha + \beta \log D$; see Luce, 1959, Theorem 9). Given that two logarithmic scales are required for the power law to hold (if the magnitude encoded on the log subjective scale was mapped directly into behavior, the response function would be of the form $N = \beta \log D + \alpha$), Izard and Dehaene (2008) proposed the following mechanism. At the first stage, perceived numerosities are encoded on the log-scaled mental number line. At the second stage, the analogue representations on the mental number line are transformed into a response by means of a response grid. The latter is also log scaled, but it can be "calibrated" with respect to the mental number line with affine transformations (stretch or shrink and shift), allowing for the adjustment of response criteria as a result of a feedback, comparison to a standard, etc.

Longo and Lourenco (2007) also demonstrated the presence of a logarithmic component in the subjective scale in the estimation of symbolic magnitudes. The novelty of their approach was to vary the start of the interval and its length, making the task much more challenging for participants when they needed to estimate the interval midpoint. Assuming homomorphism between physical and numerical magnitudes, the authors proposed that the bisection of numerical intervals should be affected by "pseudoneglect." This phenomenon is characterized by the tendency, found in healthy adults, to bisect physical lines to the left of the objective center (for review, see Jewell & McCourt, 2000). Provided that pseudoneglect represents an attentional bias of a constant strength, the authors argued that the error in the interval bisection task (i.e., the underestimation of an interval mean) should depend on the magnitude of this mean. Specifically, the authors predicted a greater underestimation for the interval mean of a larger magnitude, as the distance between large numbers on the logarithmic scale is smaller than between small numbers, meaning that an attentional bias of a constant strength should span a greater numerical distance. The results confirmed the predictions, showing that the underestimation of the interval mean increased with its magnitude. In another study, Lourenco and Longo (2009) administered a similar task, this time also asking participants to retain in memory small or large numbers presented in the beginning of each trial. When participants retained a small number, the modulation of the bias by number magnitude persisted; when participants memorized a large number, no modulation was found. Following Banks and Coleman (1981), Lourenco and Longo (2009) proposed that the use of either logarithmic or linear scale may depend on the specifics of a numerical problem at stake.

Methodological Issues

Although the previous research provided evidence in favor of both linear and logarithmic hypotheses, there are reasons to believe that the methodology used in the study of subjective number scaling is problematic. First of all, the presence or absence of a log-like nonlinearity of the trend does not generally guarantee the presence or absence of a logarithmic component in the subjective scale. In the number line tasks (Barth & Paladino, 2011; Moeller, Pixner, Kaufmann, & Nuerk, 2009; Siegler & Opfer, 2003), this problem is reflected by an ongoing discussion about the function that should be used to fit the data of young children. For example, Moeller et al. (2009) suggested that the nonlinearity is better modeled by a segmented linear regression line. More recently, Barth and Paladino (2011) showed that the performance in the task can also be accounted for by Spence's power model of proportional judgments (Spence, 1990). The model predicts that the proportion P of some unit magnitude (e.g., length of a line or the length of numerical interval) will be overestimated if $P < .5$ and underestimated if $P > .5$. The predicted response function for the model is not linear, though, as the over- and underestimation starts converging monotonically to zero for extreme values of P (i.e., $P \to 0$ or $P \to 1$).

The same problem applies to the dot enumeration studies. There is no objective reason (see Luce, 1959, Theorem 1) for assuming the hypothesis of logarithmic interval scales (as in Izard & Dehaene, 2008) to interpret the data, given a basic principle underpinning the power law. According to the law, subjects use ratio scaling, where equal stimulus ratios tend to produce equal sensation ratios (Stevens, 1957). Given that the exponent in dot enumeration tasks is less than 1, it means that each time the number of dots is doubled, an estimate increases less than twice. Moreover, there is a discrepancy between the results in the dot enumeration and line-mapping tasks, given that in the former the response function is a power one, whereas in the latter it is a linear combination of the linear and logarithmic components (Dehaene et al., 2008). This implies a different estimation mechanism in the line-mapping task, as the response here is no longer calibrated with the log-scaled response grid (a second log scale is necessary for the power law to hold).

Another problem is that the previous studies assumed that any trend observed in the responses follows solely from the idiosyncrasies of the subjective scale. However, this is not a valid assumption, as some systematic tendencies may result from response biases. One of these biases, the *central tendency effect*, was branded by Stevens (1971) as "one of the most obstinate" and "perhaps most important" (p. 428). It was first described by Hollingworth (1910), who found that judgments of physical magnitudes reveal a tendency to "gravitate toward a mean magnitude" (p. 461) of a series of presented stimuli, termed by him the *indifference point*. In other words, the stimuli of small magnitudes tended to be overestimated, and the stimuli of large magnitudes tended to be underestimated. The indifference point is not necessarily equal to a linear mean of the series. Helson (1947) argued that central tendency represents the pooled effect of all stimuli. Consequently, if the magnitudes of stimuli are represented on the compressive scale, the indifference point will be close to a geometric mean of the series; that is, its proportional magnitude to the range of series will be greater than that of an arithmetical mean.

The relevance of the central tendency issue for the study of number magnitude scaling is implied by two facts. First, the central tendency forces the response function to be less steeply increasing. For a cross-modal matching paradigm (e.g., dot enumeration), it means a smaller value for the exponent of power function. Because the "true" exponent is not available, the inference based solely on the analysis of exponent values (as is the case in Izard & Dehaene, 2008) may be inaccurate. Second, the central tendency may provide an alternative interpretation for compressive signatures found in the studies by Longo and Lourenco (2007; Lourenco & Longo, 2009). It is not clear whether the change in the size of the bias with number magnitude, reported in those studies, resulted from logarithmic spacing between magnitudes on the underlying mental continuum or from a tendency to overestimate small numbers and underestimate large numbers. In other words, a weaker pseudoneglect, found for the small magnitude of an interval mean, could also occur if the underestimation due to pseudoneglect was counterbalanced by the response bias to overestimate small numbers; and conversely, a greater pseudoneglect for the large magnitude of the mean could be a sum of pseudoneglect and the response bias to underestimate large numbers.

Given that behavior may be affected by various distorting processes, such as the central tendency or any other form of response bias, the type of subjective scale cannot be determined by simply asking people to estimate the stimulus magnitude (Gallistel & Gelman, 2005; Stevens, 1971). To address this issue, it is necessary to consider which criteria are used to determine the type of scale. The theory of measurement proposes that the type of scale is defined by the transformational rules according to which a number gets its assignment (Luce, 1959; Stevens, 1951, 1968). That is, a magnitude N on a particular scale can be constructed by applying those transformational rules to an arbitrary set of other magnitudes. For example, the characteristic feature of the logarithmic scale is that $\log A + \log B = \log AB$, whereas for the linear scale that is not an admissible transformation (Luce, 1959), because $A + B = AB$ does not hold, unless $A = 0$ and $B = 0$.

Taking admissible transformations as a criterion defining the scale leads to the consistency principle (Luce, 1959): If the manipulations on stimulus magnitude are closed under a specific transformation, then the behavioral outcomes should also be closed under a specific transformation, though not necessarily the same one. That is, to determine whether the subjective scale is log or linear, one needs to determine whether behavioral outcomes in response to combinatorial operations with stimulus magnitudes are closed under the transformations admissible for the logarithmic scale. The critical point is that without combinatorial operations, it is not possible to tell whether the subjective scale is log or linear on the basis of the observed behavior in any mapping task. A log subjective scale could result in a linear mapping to a physical continuum by a log-to-linear transformation in the response generation process; similarly, a linear subjective scale could equally result in a log external mapping by a linear-to-log transformational process.[1]

Combinatorial Method

The main aim of the study was to present a method for differentiating the hypotheses of linear and logarithmic mapping for numerical magnitudes while controlling for response bias. We apply this method to the data obtained using modified versions of the number-to-position paradigm. Participants were required either to indicate the relative position of a number within a numerical interval by marking a physical line (the line-marking task, Experiment 1) or, given the segment of an interval, to extend the physical line to fit the length of the whole interval (the line construction task, Experiment 2). The numerical start and length of an interval were varied. Thus, in the line-marking task (see Figure 3), the problem participants could face would be marking the location of the number 23 within an interval bracketed by 12 on one side and 45 on the other. A correct location would be, then, a third of the line from the end bracketed by 12. In the line construction task, the problem was somewhat different. Participants would be presented with a physical line bracketed by 12 and 23, and they would be required to extend the physical line such that it would correspond to the length of the interval from 12 to 45. We presented the intervals in two orientations: left to right (L-R) and right to left (R-L). The hypothesis of an obligatory L-R mapping on the mental number line suggests that the performance for the R-L orientation should be less accurate as a result of costs associated with remapping of a R-L interval on the L-R mental continuum. The change in the parameters of a response function is also a possibility.

In our study we consider three hypotheses: (a) strong linear (i.e., the subjective scale is linear), (b) strong logarithmic (i.e., the subjective scale is logarithmic), and (c) weak logarithmic (i.e., the subjective scale is partly linearized but contains a significant

[1] We are grateful to one of the reviewers for insisting on this point, which seems to have been ignored by many researchers.

logarithmic component). To make explicit the combinatorial operations underlying the method, let us define two numbers bracketing a numerical interval as Start and End and a number falling within this interval as Target. (The constraint that Target should lie between Start and End in its numerical value applies to the line-marking task only, but the predictions for the line construction task, where the Target magnitude falls outside that interval, are identical, with the only difference that the labels Target and End are swapped.) Next, we express each stimulus magnitude as the arithmetical sum of two numbers. Taking S as a distance between Start and 0, and L and T are some arbitrary scalar magnitudes, such that $0 < T < L$, we define

$$\begin{aligned} \text{Start} &= 0 + S = S \\ \text{End} &= L + S \\ \text{Target} &= T + S. \end{aligned} \tag{1}$$

The question we want to address now is, What is the form of the admissible transformations on the subjective scale that would account for the position of Target within the interval bracketed by Start and End, given the combinatorial operations with stimulus magnitudes, listed in Equation 1? First of all, the position of a Target magnitude within an interval is given as a relative distance between Target and Start to the length of the whole interval, that is,

$$\text{Target}_{[\text{Start-End}]} = \frac{f(\text{Target}) - f(\text{Start})}{f(\text{End}) - f(\text{Start})}, \tag{2}$$

for some scaling function f. We will use the convention of adding the subscript to a variable to denote its relative position within an interval bounded by variables in the subscript brackets as opposed to the absolute value of that variable.

For the linear mapping function, $\text{Target}_{[\text{Start-End}]}$ becomes

$$\text{lin Target}_{[\text{Start-End}]} = \frac{T + S - S}{L + S - S} = \frac{T}{L}. \tag{3}$$

That is, lin $\text{Target}_{[\text{Start-End}]}$ does not depend on the start S of the interval but only on the relative magnitude of T to the interval length L. In addition, lin $\text{Target}_{[\text{Start-End}]}$ is not affected by the length of the interval as long as the proportion between T and L is preserved.

For the strong logarithmic hypothesis, $\text{Target}_{[\text{Start-End}]}$ becomes

$$\text{log Target}_{[\text{Start-End}]} = \frac{\log(T + S) - \log(S)}{\log(L + S) - \log(S)} = \frac{\log\left(\frac{T + S}{S}\right)}{\log\left(\frac{L + S}{S}\right)} = \frac{\log\left(\frac{T}{S} + 1\right)}{\log\left(\frac{L}{S} + 1\right)}. \tag{4}$$

From the above expression, it can be seen that S does not cancel out; therefore, log $\text{Target}_{[\text{Start-End}]}$ depends on where the interval starts. In addition, the premultiplication of T and L by a common factor n does not imply that log $\text{Target}_{[\text{Start-End}]}$ remains the same. That is, log $\text{Target}_{[\text{Start-End}]}$ will depend on how wide the interval is, even though the linear proportionality between T and L is preserved.

Finally, the weak logarithmic hypothesis suggests that mapping is partially linearized but preserves a log component. Consequently, the relative distance is obtained by summing nominators and denominators of linear and log $\text{Targets}_{[\text{Start-End}]}$, that is,

$$\text{lin log Target}_{[\text{Start-End}]} = \frac{w_1 T + w_2 \log\left(\frac{T}{S} + 1\right)}{w_1 L + w_2 \log\left(\frac{L}{S} + 1\right)}, \tag{5}$$

where w_1 and w_2 are the weighting parameters for the linear and logarithmic components, respectively. In general, the term *linearization* of number representations implies that the size of the logarithmic component decreases as the size of the linear component increases. Consequently, all three hypotheses can be expressed by means of a single expression,

$$\text{Target}_{[\text{Start-End}]} = \frac{wT + (1 - w)\log\left(\frac{T}{S} + 1\right)}{wL + (1 - w)\log\left(\frac{L}{S} + 1\right)}, \ 0 \leq w \leq 1, \tag{6}$$

where linear and logarithmic components form a convex combination.[2] The weighting parameter w determines the identity of the scale, such that the strong linear hypothesis corresponds to $w = 1$; the strong logarithmic hypothesis corresponds to $w = 0$; and the weak log hypothesis, because $n \ll \log n$, corresponds to $w \ll 1$. The critical point is that Equation 6 represents a general case for admissible transformations on the subjective scale, under which the arithmetical operations with stimulus magnitude, listed in Equation 1, are closed. That is a direct implementation of Luce's consistency principle (Luce, 1959).

We assume that an estimate of a $\text{Target}_{[\text{Start-End}]}$ is subject to random Gaussian noise and is mapped into behavior via some response function with coefficients $\mathbf{B} = \{\beta_i\}$. For the purposes of the current study, we assume that the response function is linear, that is,

$$\text{Response} = \beta_1 \text{Target}_{[\text{Start-End}]} + \beta_0. \tag{7}$$

In what follows, we will address the model given by Equations 6 and 7 as the full model and the model given by Equations 3 and 7 as the linear model.

It is easy to see why the method is well posed for differentiating between the subjective scale and the response bias. For a particular value w, we can construct some arbitrary $\text{Target}_{[\text{Start-End}]}$ in a multiple ways using two or more sets of values for S, T, and L.

[2] An alternative formulation for the weak hypothesis as $\text{Target}_{[\text{Start-End}]} = w$ lin $\text{Target}_{[\text{Start-End}]} + w$ log $\text{Target}_{[\text{Start-End}]}$ appears to be conceptually inappropriate. It would imply that the estimation of a Target position within an interval is performed twice, on the linear and log scales separately, and the result is then determined by mixing the results of two estimations proportionally to the weight w. In other words, this formulation would imply that two independent scales are used concurrently.

Every such set will generate a different $\text{Target}_{[\text{Start-End}]}$ for a different value of w (i.e., on a different scale). Owing to multiple assignments, the method does not confound magnitude information contained in $\text{Target}_{[\text{Start-End}]}$ with **B**. The latter provides the estimation for the size of response bias that, by definition, should be indifferent to the combinations of S, T, and L as long as they produce the same magnitude of $\text{Target}_{[\text{Start-End}]}$. Obviously, manipulations with any two of the triplet S, T, and L, while keeping a third one fixed, would suffice to generate an infinite number of a particular $\text{Target}_{[\text{Start-End}]}$ replications. This implies that orthogonal manipulations with any two variables (or, alternatively, with their sums, products, etc.) are both necessary and sufficient for discriminating between the linear and log hypotheses experimentally.

By contrast, the studies that used the standardized intervals (e.g., Dehaene et al., 2008; Siegler & Opfer, 2003) could not decouple the contributions of response bias and weights to the values of regression coefficients, as there was only one way to assign the magnitude of $\text{Target}_{[\text{Start-End}]}$. Longo and Lourenco (2007) apparently came closer than others to a realization of the combinatorial method when they manipulated the beginning and the length of the interval. However, they did not make use of the method, taking into consideration only one variable: the magnitude of the interval mean.

The above formulation allows for two approaches to the data analysis. First, the contribution of the logarithmic component can be estimated directly by optimizing the model given by Equations 6 and 7. The second approach is analytical and provides a broader picture about the factors affecting the performance. The geometric interpretation of this idea is given in Figures 1A and 1B. We will drop the symbolic notation of Equation 6 and, instead, use the labels that define features of a numerical interval. The way S, L, and T were defined implies that S stands for the magnitude of an interval Start, L stands for a linear Length of the interval, and T/L stands for a linear $\text{Target}_{[\text{Start-End}]}$. The primary concern here is how the difference between log and linear $\text{Target}_{[\text{Start-End}]}$ will change for different choices of Start, Length, and linear $\text{Target}_{[\text{Start-End}]}$. The examples in Figure 1C indicate that the difference between log and linear $\text{Target}_{[\text{Start-End}]}$ is greater (a) for the values of linear $\text{Target}_{[\text{Start-End}]}$ between .1 and .6, (b) for the intervals with a smaller Start, and (c) for the intervals with a greater Length. Furthermore, we can marginalize the effect of each variable by averaging across the other two. Figure 2A shows the marginalized difference between log and linear $\text{Target}_{[\text{Start-End}]}$ for different values of linear $\text{Target}_{[\text{Start-End}]}$, and Figure 2B shows that for different interval Starts and Lengths. All functions are nonlinear, but their linear approximations have distinctive slopes. The predominantly decreasing trend for linear $\text{Target}_{[\text{Start-End}]}$ and the ever-decreasing trend for Start can be approximated by a line with a negative slope, whereas the increasing trend for Length can be approximated by a line with a positive slope. Importantly, for mappings that are partially linearized, the sign of slopes for all

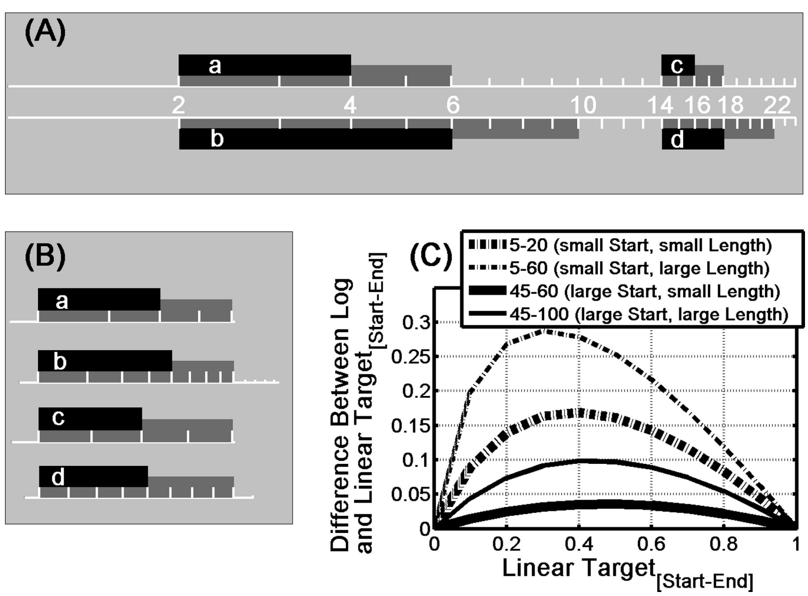

Figure 1. (A) Placement on the logarithmic scale at different Starts and different interval Lengths. Bars represent numerical distances: from Start to Target (in black) and from Start to End (in gray). The length of black bars relative to the length of gray bars on the linear scale would always be .5. However, it is not the case for the logarithmic scale, as can be seen in Figure 1B. Interval c (greater Start and smaller Length) matches the linear proportion most closely. (C) The change in the difference between log and linear $\text{Target}_{[\text{Start-End}]}$ with the change in linear $\text{Target}_{[\text{Start-End}]}$, Start. and Length. The scale of axes is normalized (both linear and log $\text{Target}_{[\text{Start-End}]}$ are proportional magnitudes). In the legend, the first digit stands for Start (i.e., the beginning of an interval), and the second digit stands for End (i.e., the end of an interval). Individual curves provide a 2×2 example of four intervals, with two choices for Start (small: 5; large: 45) and two choices for Length (small: $20 - 5 = 15$; $60 - 45 = 15$; large: $60 - 15 = 55$; $100 - 45 = 55$). On average, the difference between log and linear $\text{Target}_{[\text{Start-End}]}$ is larger for linear $\text{Target}_{[\text{Start-End}]}$, between .1 and .6, for small Start (5) and for large Length (55).

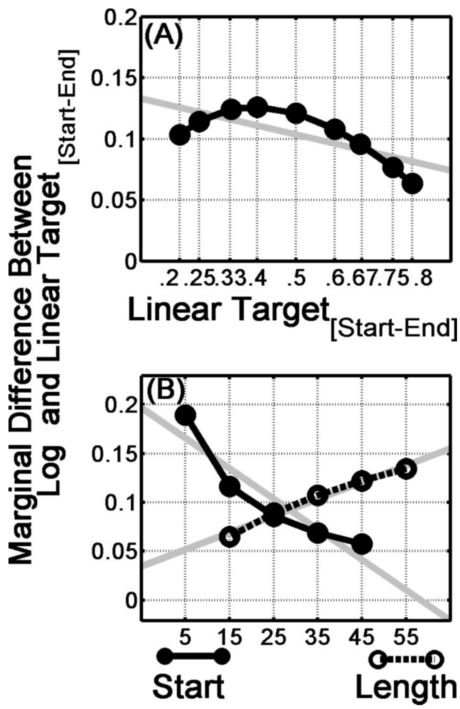

Figure 2. The predictions for logarithmic mapping. (A) The marginal difference between log and linear Target$_{[End\text{-}Start]}$ for nine choices of linear Target$_{[End\text{-}Start]}$ (⅕, ¼, ⅓, ⅖, ½, ⅗, ⅔, ¾, ⅘) averaged across five Starts (5, 15, 25, 35, 45) and five Lengths (15, 25, 35, 45, 55). The range of values approximates that used in the study (see Method sections). The *y*-axis is a normalized scale. The predicted trend is predominantly decreasing. The gray line shows the linear approximation to the trend. (B) Marginal difference between log and linear Target$_{[End\text{-}Start]}$ as a function of Start and Length. The labels for the *x*-axis show numerical magnitudes for Start and Length; the *y*-axis is a normalized scale. The choices for Start, Length, and linear Target$_{[End\text{-}Start]}$ are as in Figure 2A. For Start, the marginal difference is calculated by averaging across Length and linear Target$_{[End\text{-}Start]}$. For Length, the marginal difference is calculated by averaging across Start and linear Target$_{[End\text{-}Start]}$. Logarithmic mapping predicts a decreasing trend for Start and an increasing trend for Length. The gray lines show their linear approximations. For mapping, which is partially linearized or affected by the central tendency, the steepness of the trends for Start and Length will be smaller, but the directions remain unchanged.

variables remains unchanged, though the steepness of the trends will depend on the relative contribution of a logarithmic component. Moreover, the sign of slopes for Start and Length would remain unaffected by response bias (i.e., when $\beta_0 \neq 0$, $\beta_1 \neq 1$).

Consequently, a simple tool for testing both the weak and strong logarithmic hypotheses can be the following. Provided that linear Target$_{[Start\text{-}End]}$, Start, and Length are uncorrelated or orthogonal, the deviations of the response from a correct value can be fitted with linear multiple regression. If numbers are represented on a (partially linearized) logarithmic scale, the regression coefficients for linear Target$_{[Start\text{-}End]}$ and Start are expected to be negative, whereas for Length they are expected to be positive.

Experiment 1: Line-Marking Task

In the first experiment, we implemented the combinatorial method in a task that required mapping numbers to a location on the line. We systematically manipulated parameters *S*, *L*, and *T*/*L* (experimental variables Start, Length, and linear Target$_{[Start\text{-}End]}$) and used both numerical and analytical approaches to the data analysis. In addition, using a Bayesian statistical approach, we compared the full model of Equations 6 and 7 with the simpler linear model of Equations 3 and 7.

Method

Participants. Twenty healthy adults (10 men, 10 women) 19–40 years of age (M = 24.1, SD = 5.33) participated in the study. They all gave informed consent, had normal or corrected-to-normal vision, and declared themselves to be right-handed.

Stimuli and apparatus. The line-marking task was administered by means of a custom-made MATLAB program and displayed with a 19-in (48.26-cm) LCD monitor (1440 × 900 pixels; pixel size = .265 mm). All stimuli in the experiment were designed in terms of a pixel size. Participants saw a gray 15-pixel-wide line presented against a black background in the middle of the screen along the vertical axis (see Figure 3). Along the horizontal axis, the location of the line center varied randomly within 50 pixels off the monitor center in either direction. The length of the line varied randomly between 480 and 520 pixels, subject to constraints discussed below. A thin red vertical strip (1 pixel thick, 31 pixels long), functioning as a cursor, was presented simultaneously with the line. The cursor split the line into two parts, and on presentation it could occupy any randomly selected location between the ends of the line. The cursor displacement, constrained to the horizontal dimension, was manipulated by a computer mouse. The trial was terminated by clicking the left button of the mouse. The location of the cursor at the time of the click was registered and used to calculate accuracy of the response. The resolution of the response was equal to the pixel size (.265 mm).

In each trial, participants saw three numbers (font size = 20). Two of them (in white) were presented at the opposite ends of the line. A smaller number, Start, signified the beginning of an interval, and a larger number, End, signified the end of the interval. A number to be marked, Target (in red), lay between Start and End in its numerical value. The orientation of the line could be either left to right (L-R) or right to left (R-L). In the L-R condition, a

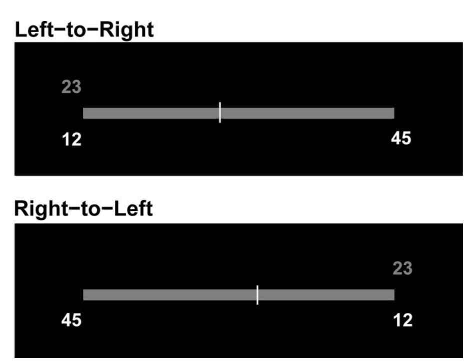

Figure 3. Stimuli in the line-marking task. Participants are required to mark the location of the gray number (red in actual experimental settings) within the interval defined by the two white numbers by sliding the cursor (vertical strip) along the line. At the beginning of each trial, the cursor was presented at a random location on the line.

Start was presented at the left end of the line and an End at the right end. The layout was reversed for the R-L condition.

Design. Three numerical factors were manipulated in the task: (a) linear $\text{Target}_{[\text{Start-End}]}$ (i.e., the relative distance between Start and Target to the length of the interval), (b) Start (i.e., the origin of the interval), and (c) Length (i.e., the length of the interval). These variables corresponded directly to the values of T/L, S, and L, respectively, used in the description of the combinatorial method. The choice of a Target magnitude was such that it divided the interval proportionally into one of nine linear $\text{Target}_{[\text{Start-End}]}$ values: ⅕, ¼, ⅓, ⅖, ½, ⅗, ⅔, ¾, ⅘.

The experiment had a blocked design. In the Start-controlled block, the values for Start were drawn at random from one of the "bins": 1–9, 11–19, 21–29, 31–39, and 41–49. A Start from each bin was presented once with each linear $\text{Target}_{[\text{Start-End}]}$. The assignment of Length was random in this block but was subject to two constraints. First, the value for Length was within the range between 10 and 60. Second, a numerical interval initially generated by the computer program was corrected to the nearest value divisible without remainder by the denominator of a linear $\text{Target}_{[\text{Start-End}]}$. The reason for using the latter constraint was to ensure that Target was always an integer.

In the Length-controlled block, a Length from each bin was presented once with each linear $\text{Target}_{[\text{Start-End}]}$. The bins for Length were 11–20, 21–30, 31–40, 41–50, and 51–60. Again, the adjustments of Length magnitude were required to ensure that the value of Target was an integer. The assignment of Start was random in this block, but its magnitude was contained in the range between 1 and 49, excluding the multiples of 10.

The final constraint in stimulus generation relates to the length of the presented line. Although the length for the line was drawn in the first instance from the uniform distribution to be between 480 and 520 pixels, the length of an actually presented line was adjusted to a nearest value, such that the line would contain a number of pixels divisible by the denominator of a given linear $\text{Target}_{[\text{Start-End}]}$ without remainder. This allowed for a precise marking of the line with respect to a required linear $\text{Target}_{[\text{Start-End}]}$.

Given the 2×2 design (L-R or R-L orientation by Start-controlled or Length-controlled block), the task consisted of four blocks of 45 trials (nine levels of $\text{Target}_{[\text{Start-End}]}$ times five levels of Start/Length) each. Both within- and between-block orders of presentation were randomized.

Procedure. Participants were shown the stimulus material, explained the task, and instructed how to respond. They were asked to provide an approximate and unspeeded estimate of the position of the Target number on the line without performing exact arithmetical calculations. In order to respond, participants were required to move the cursor along the line to an estimated location and mark the line by clicking on the left button of the mouse. Participants were asked not to hurry or spend too much time on a trial. As guidance, the time interval of 5–10 s per trial was suggested. However, it was made clear that this time window was not obligatory. Participants also underwent a training session to become familiar with the tasks. The training session involved a different set of linear $\text{Target}_{[\text{Start-End}]}$ values—namely, 1⁄7, 2⁄7, 5⁄7, and 6⁄7—and consisted of two (L-R and R-L) blocks, where each linear $\text{Target}_{[\text{Start-End}]}$ was presented twice within each block, giving eight trials in the session. Both Start and Length were drawn randomly. Each block in the experimental session was preceded by a message on the screen specifying the orientation of the line.

Data analysis. The responses were normalized by calculating them as proportions of the line segment between the beginning of the line and the marked point divided by the length of the whole line. This transformation placed responses onto identical scale with linear $\text{Target}_{[\text{Start-End}]}$ and allowed for a straightforward calculation of the error as a difference between response and linear $\text{Target}_{[\text{Start-End}]}$. Three main issues were addressed in the analysis: (a) the selection of a model for the data, (b) the response bias, and (c) the marginal effects of linear $\text{Target}_{[\text{Start-End}]}$, Start, and Length. Within each subsection, the effect of orientation was also investigated.

Model selection. The parameters for the full model given in Equations 6 and 7 were calculated for each subject and for each orientation separately. The magnitudes of Start, Length, and the difference between Start and Target were plugged into Equation 6 in place of S, L, and T, respectively. Values for β_0, β_1, and w were calculated according to the least squares criterion, with an optimization algorithm (function *fmincon* in MATLAB). The initial values for β_0, β_1, and w were set to 0, 1, and 0, respectively, corresponding to a null hypothesis that the subjects responded in accordance with the strong logarithmic hypothesis and a zero response bias. It should be stressed that a traceable logarithmic component would require a small value for the weight w (roughly, smaller than .1), given that $\log(n) \ll n$.

To evaluate the performance of the full model, we compared it with the linear model, given by Equations 3 and 7, which is just a linear regression model with the linear $\text{Target}_{[\text{Start-End}]}$ as a predictor. To approximate the posterior distribution of the parameters β_0, β_1, and w, we drew 10,000 Markov chain Monte Carlo parameter samples for each model, subject, and orientation, dropping the first 500. The proposal distributions were assumed to be Gaussian. To correct for a small proportion of the interval between 0 and 1, for which parameter w implies a traceable contribution of the logarithmic component, we used an inverse arcsine transformation of the form $w = [\sin(w') - 1]^2$, and sampled w' instead of w. The value of the parameter w' was bound to be between 0 and $\pi/2$; the values that were sampled outside that interval were reflected back into the interval. Owing to the transformation, the proportion of the interval between 0 and $\pi/2$ that was compatible with the log hypothesis was roughly .5. We calculated the log of the marginal likelihood, $L(\text{model})$, for each model by transforming logarithmically the average likelihood over all Markov chain Monte Carlo samples. The differences in the logs of average likelihoods for two models, $L(\text{linear}) - L(\text{full})$ (i.e., the logs of the individual Bayes factors between the models), was then tested against zero with nonparametric Wilcoxon signed-rank test. The values greater than 0 would support the hypothesis of the linear scaling, and the values smaller than 0 would support the logarithmic hypothesis. The cross-subject log of the Bayes factor was calculated by summing the individual logs. Similarly, the effect of the line orientation was studied by looking at the Bayes factor between L-R and R-L conditions.

Analysis of bias. Two parameters were of interest in the analysis of response bias. The first was the slope β_1 of the full model (or of the linear model, if it performed better than the full one), which can be treated as a spread–compression index. For example, a value smaller than 1 would imply that the spread of the mean responses was smaller than it was required by the variance in

$\text{Target}_{[\text{Start-End}]}$ of Equation 6, and therefore some values should be either overestimated or underestimated, or both. The test of the slopes against 1 was complemented by the test suggested in Matthews and Stewart (2009). This required testing the standard deviation of responses against the standard deviation of independent variable. In some respect, this test is more robust, as it takes into account the within-subject variability of responses.

The second parameter of interest was the value of the regression models at $\text{Target}_{[\text{Start-End}]} = .5$ (i.e., the regression mean). This parameter provided information about the symmetry of the compression–stretch and can be interpreted as a marker of global under- or overestimation. For regression slopes that were smaller than 1 (compressed responses), the regression mean below .5 would indicate that there was a tendency to underestimate in general, and vice versa if the mean value was above .5.

The effect of linear $\text{Target}_{[\text{Start-End}]}$, Start, and Length. We used the multiple regression analysis with Start, Length, and linear $\text{Target}_{[\text{Start-End}]}$ as predictors to obtain the estimation of the marginal effect of each variable on the performance. The dependent variable of the analysis was the error, calculated as a difference between response and linear $\text{Target}_{[\text{Start-End}]}$. The betas for randomly generated variables (i.e., Start and Length in the Length-controlled and Start-controlled blocks, respectively) were disregarded. Consequently, we analyzed two samples of beta values for Start (L-R and R-L in Start-controlled condition) and Length (L-R and R-L in Length-controlled condition) and four samples of betas for linear $\text{Target}_{[\text{Start-End}]}$ (L-R and R-L in both Start- and Length-controlled conditions). The significance of a trend was established by testing betas for each variable against zero with t-test statistics. In addition, betas for the L-R and R-L conditions were tested against each other, in order to see whether manipulations with the line orientation had any effect on the data.

Results

Fifty-two trials (1.4%) were excluded from analysis, either because reaction time was less than 200 ms (16 trials) or because the deviation from a correct response was more than .4 (36 trials).

Model selection. The estimated median weight w for the full model was equal to 1 (all $w > .27$). The linear model provided a better account for the data than the full model, as the median log of the Bayes factor was significantly greater than zero (L-R: $z = 2.69$, $p < .01$; R-L: $z = 2.17$, $p = .03$). The results were supported by the analysis on the basis of Akaike information criterion, calculated for the numerically optimal models (L-R orientation: $z = 3.92$, $p < .001$; R-L orientation: $z = 3.88$, $p < .001$). The cross-subject log of the Bayes factor was equal to 4.95 and 3.78 for L-R and R-L orientations, respectively, implying very strong evidence in favor of the linear hypothesis. The effect of orientation was not significant ($z < 1$ for both the full and linear models, confirmed by Akaike information criterion). The linear model accounted for 75% of variance for L-R orientation and 76% of variance for R-L orientation.

Analysis of bias. Because the linear model predicted the data better than the full model, we used this model for the analysis of response bias. The average response function was $\text{Response}_{(\text{L-R})} = .816 \times \text{Target}_{[\text{Start-End}]} + .106$ and $\text{Response}_{(\text{R-L})} = .821 \times \text{Target}_{[\text{Start-End}]} + .108$ for L-R and R-L orientations, respectively (see Figure 4). The slopes of the regression models fitted to each

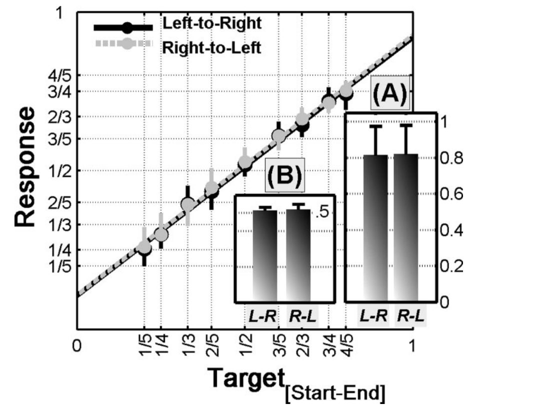

Figure 4. Line-marking task results. The group means and their standard deviations of responses with the linear $\text{Target}_{[\text{Start-End}]}$ as a predictor. Figure 4A shows that the slopes of linear regression models are significantly smaller than 1, indicating the presence of linear compression in the data, that is, a central tendency. Figure 4B shows the bars for the regression means. The latter were slightly greater than the middle of the interval for both line orientations, suggesting a small overall overestimation. L = left; R = right.

subject data were significantly smaller than 1, $t_{\text{L-R}}(19) = 5.15$, $p < .001$, and $t_{\text{R-L}}(19) = 5.02$, $p < .001$, indicating the presence of the central tendency bias in the data. The slopes for L-R and R-L orientations did not differ from each other ($t < 1$). The alternative test for the central tendency (Matthews & Stewart, 2009) showed that participants' standard deviations of responses were significantly smaller than the standard deviations of linear $\text{Target}_{[\text{Start-End}]}$ values, $t_{\text{L-R}}(19) = 2.40$, $p = .027$, and $t_{\text{R-L}}(19) = 2.41$, $p = .026$.

The regression means were slightly greater than .5 (L-R: .51; R-L: .52). Because of the small between-subjects variability, the difference from .5 was statistically significant, $t_{\text{L-R}}(19) = 3.36$, $p < .005$, and $t_{\text{R-L}}(19) = 2.86$, $p = .01$, which indicate some tendency to globally overestimate. There was no difference in regression means for L-R and R-L orientations.

The effects of linear $\text{Target}_{[\text{Start-End}]}$, Start, and Length. The marginal error in responses, calculated for each experimental variable by averaging across the others, is shown in Figure 5. The values for both Start and Length are arranged into five bins to enable averaging across participants. The multiple regression analysis, meanwhile, was run on the actual numerical magnitudes for these variables.

As would be expected from the fact that the slopes of linear models were considerably smaller than 1, there was a significant negative trend in errors as a function of linear $\text{Target}_{[\text{Start-End}]}$: Start-controlled, L-R: $t(19) = 4.81$, $p < .001$, $R^2 = .20$;[3] Start-controlled, R-L: $t(19) = 5.40$, $p < .001$, $R^2 = .18$; Length-controlled, L-R: $t(19) = 5.07$, $p < .001$, $R^2 = .18$; Length-controlled, R-L: $t(19) = 4.26$, $p < .001$, $R^2 = .16$. The average sample slopes were $\beta_{\text{L-R/Start}} = -.190$, $\beta_{\text{L-R/Length}} = -.202$, $\beta_{\text{R-L/Start}} = -.187$, and $\beta_{\text{R-L/Length}} = -.160$. A repeated measures

[3] We calculated R^2 using linear models with each variable separately as a predictor to fit the data.

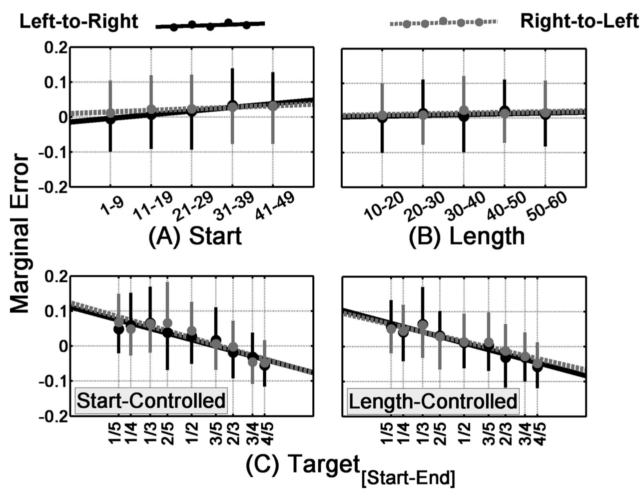

Figure 5. Line-marking task. The marginal effects of interval Start, interval Length, and linear Target$_{[\text{Start-Target}]}$ on the errors in responses. The error bars show the group mean standard deviations. The effect of Start was positive, whereas the effect for Length was completely flat, contrary to the predictions of the log scale hypothesis. Neither experimental variable showed a significant difference between left-to-right and right-to-left orientation.

2×2 analysis of variance on betas (block: Start/Length; orientation: L-R/R-L) showed no significant main effect or interaction (all Fs < 1.6). There was a remarkable consistency in the beta values at a within-subject level, with correlational coefficients between betas for different blocks ranging from .62 to .81. In addition, there was a strong negative correlation between betas and R^2 estimated for the linear model (see the previous subsection; all $\tau > .52$, $p < .001$ [nonparametric Kendall's test]).

Testing the betas for Length against zero revealed no significant trend ($t < 1$ for both L-R and R-L), whereas the trend for Start was significant for both orientations, $t(19) = 3.38$, $p < .01$, $R^2 = .06$, and $t(19) = 2.64$, $p = .016$, $R^2 = .02$, for L-R and R-L, respectively. As Start increased, the error grew positively with the mean rate $\beta_{\text{L-R}} = .012$ and $\beta_{\text{R-L}} = .006$ per 10 number units. The positive value was found in 29 out of 40 cases (20 participants by two line orientations). The magnitude of the effect was somewhat greater for L-R than for R-L, with a marginally significant difference between two orientations, $t(19) = 1.97$, $p = .06$. Despite this difference, there was a significant correlation between individual betas for L-R and R-L ($r = .48$, $p = .032$), suggesting that the effect (unlike the effect of Length, $p = .19$) was consistent at a within-subject level.

Discussion

We applied the combinatorial method to differentiate between subjective scale and response bias in a task where participants marked the location of a numerical magnitude within numerical intervals in which Start and Length were varied systematically. The results unambiguously show that approximate estimation in this particular task is performed on a strictly linear scale. The linear regression model predicted the data better than the model that included the weight for the logarithmic component. This finding was supported by the analysis of the marginal effects of linear Target$_{[\text{Start-End}]}$, Start, and Length on the error. For the logarithmic mapping, the regression slope for linear Target$_{[\text{Start-End}]}$ is expected to be negative and complemented with the negative slope for Start and the positive slope for Length. However, the results show an opposite trend for Start with no significant effect of Length.

The results showed that performance was affected by linear compression due to a response bias, known as the central tendency effect. In other words, the small values within a numerical interval were systematically overestimated, and large numbers were systematically underestimated. The strength of the central tendency generally reflected participants' ability to solve the task, such that the smaller central tendency was associated with higher proportion of the variance, explained by the regression models. The regression mean was close to, but statistically greater than, the middle of the numerical interval. This slight shift has a simple explanation in another factor that biased performance: the magnitude of Start. It can be noted that the intercepts of the least squares lines for Start in Figure 5A are approximately equal to zero. Consequently, each level of Start contributed to the magnitude of a responded ratio, causing, on average, a slight increase in the regression mean. It can also be confirmed by the fact that the regression means significantly correlated with the betas for Start at a within-subject level for either line orientation (L-R: $r = .61$, $p < .005$; R-L: $r = .49$, $p = .03$).

Ascribing the biasing effect specifically to the magnitude of Start may be inappropriate, as the display of the task constitutes a rather complex composition of different numbers, where the magnitude of Start can strongly correlate with other magnitudes and their sums (but not with the differences between numbers). Consequently, the performance can be better accounted for by saying that participants tended to provide a greater estimate for the magnitude of Target when they faced numerical problems of a greater numerical size.

The question remains whether this bias, linear mapping, and the central tendency effect generalize to a task involving a different set of constraints and response requirements. The following study aimed at extending the understanding of the processes that affect the mapping of the internal magnitude scale into behavior in a novel number-to-position task.

Experiment 2: The Line Construction Task

To test the generality of our findings, we designed a new task in which participants had to construct an interval. As before, they were presented with a line signifying the length of the numerical interval between Start and End. However, this numerical interval was deemed to be just a part of a whole interval. Given the length of the line and the numerical length of the part, participants were asked to extend the line up to a magnitude of Target, which was always greater than End. For example, participants could be presented with a line bracketed by 12 and 23. Given 45 as a Target number, participants had to add an extension to the line, such that the length of the extension would correspond to the numerical distance between 23 and 45. In this particular example, the length of a correctly constructed extension would be twice as long the initially presented line segment.

The differences in the task do not prevent us from using the same analytic apparatus for testing the linear and logarithmic hypotheses. To account for the fact that Target is larger than End, we simply redefine Length as the distance between Start and Target and exchange Target and End in Equation 2 to get

$$\text{End}_{[\text{Start-Target}]} = \frac{f(\text{End}) - f(\text{Start})}{f(\text{Target}) - f(\text{Start})}. \quad (8)$$

The result that follows is identical to Equation 6, except that $\text{End}_{[\text{Start-Target}]}$ substitutes for $\text{Target}_{[\text{Start-Target}]}$, T represents the distance between Start and End, and L represents the distance between Start and Target. In this formulation, the predictions for the logarithmic hypothesis remain identical to the line-marking line; that is, the slopes are expected to be negative for $\text{End}_{[\text{Start-Target}]}$ and Start and positive for Length.

Apart from the differences in response requirements, it is also worth considering the differences in the constraints between two tasks. In the previous task, no cues, apart from numerical values, were available on where the line should be marked. On the other hand, the response was constrained to lie within a closed spatial interval, represented by the physical line. Given that the center of the line is easily identified, the tendency to overestimate or underestimate the magnitudes around the middle of a numerical interval could be artificially induced by the spatial format of the task.

In the context of the line construction task, the central tendency involves different processes and would manifest itself as the tendency to underconstruct a long addend to the part and overconstruct a short addend. In the current design, the length of a presented part provided a cue as to how long an added line should be. If an initially presented line was short, participants could figure out fairly quickly that they need to construct a rather long addend to the line, and vice versa if an initially presented line was long. On the other hand, because the standard for the whole line was never shown to participants, the presented segment did not clearly indicate how long the line should be and where the middle of an interval should lie. In this respect, the line construction task can be more sensitive for the study of the number magnitude scale than the line-marking task, as participants were free to construct the size of representational space. If a logarithmic component was indeed present in the estimation, then participants would systematically underconstruct the line (i.e., causing the shift of the regression mean toward a greater value).

Method

Participants. Twenty participants (eight men, 12 women) 20–52 years of age (M = 25.1, SD = 7.62) took paid participation in the study. They all gave informed consent, had normal or corrected-to-normal vision, and declared themselves to be right-handed.

Stimuli and apparatus. In this experiment, participants saw a gray horizontal line that was deemed to be just a part of a longer whole line. The line width was identical to that used in Experiment 1. In the L-R condition, a Start and an End were presented in white below the line at its left and right ends, respectively. At the right end and above the line, a Target was presented in red. The location of the line's right end varied randomly between 60 and 140 pixels to the left of the monitor center. Moving the mouse to the right enabled participants to extend the line by adding a white strip (extension) to the initially presented gray part. The extension continuously prolonged with the movement of the mouse, and it could also be reduced by moving the mouse backward. The spatial layout for R-L condition was reversed. The manipulations with the mouse had no effect on the length of the initially presented gray part. Any displacement of the mouse along the vertical axis was ignored, and the speed of the extension growth or shrink was identical to the speed of the cursor in Experiment 1.

Design and procedure. The experimental design and procedure of the line construction task, with appropriate adjustments, mirrored those of the line-marking task in Experiment 1. By contrast to the line-marking task, the Target magnitude in the current task was always greater than End. Consequently, to make two tasks comparable, we made two changes in experimental variables. First, linear $\text{End}_{[\text{Start-Target}]}$ substituted for linear $\text{Target}_{[\text{Start-End}]}$; that is, we manipulated a relative distance between Start and End to the distance between Start and Target instead of a relative distance between Start and Target to the distance between Start and End. Second, Length was defined as the distance between Start and Target (between Start and End in the line-marking task). The values for linear $\text{End}_{[\text{Start-Target}]}$, Start, and Length were generated in the same way as described for the line-marking task.

The only difference in the experimental procedure was in the response requirements: Instead of marking a presented line, participants had to construct the line as far as it was implied by the magnitude of Target, given a numerical distance between Start and End and the length of the gray line, representing the physical analogy of that numerical distance. The length of the presented line was such that a correct estimation would require a whole line to be between 460 and 540 pixels long. The value for the correct line length was drawn from a uniform distribution but subject to the same constraints as described for the line-marking task. We chose not to vary the length of the line to a greater extent, as it would make a comparison between the current and line-marking task problematic due to the differences in spatial parameters of the tasks.

In line with the change in the definition of experimental variables, the response was calculated as a relative magnitude of a gray part to the sum of the gray part and a constructed white segment.

Results

Eleven trials ($<$1%) were excluded from the analysis. For all of them the deviation from a correct response was more than .4.

Model selection. The median weight w for the full model given in Equations 6 and 7 was equal to 1. Only for one subject the magnitude of the weight (w = .065) was sufficiently small to suggest the presence of the logarithmic component in the responses. The model comparison showed that the linear model predicted data better than the full model (L-R: z = 3.25, p < .005; R-L: z = 2.24, p = .025). The result was confirmed by the statistics on the basis of Akaike information criterion (z = 3.67, p < .001) for both orientations. The cross-subject log of Bayes factor was equal to 5.45 for L-R orientation and 4.11 for R-L orientation (very strong evidence in favor of the linear model). The linear model accounted for 87% of the variance for the L-R condition and 85% for the R-L condition. The effect of the line orientation was not significant (z < 1 for both models, confirmed by Akaike information criterion).

Analysis of bias. The mean equations of the linear regression model with the linear $\text{End}_{[\text{Start-Target}]}$ as a predictor were

$\text{Response}_{(L\text{-}R)} = .811 \times \text{End}_{[\text{Start-Target}]} + .081$ and $\text{Response}_{(R\text{-}L)} = .808 \times \text{End}_{[\text{Start-Target}]} + .092$ (see Figure 6). The slope of the linear regression line was significantly smaller than 1, $t(19) = 7.61$, $p < .001$, and $t(19) = 6.95$, $p < .001$, for L-R and R-L, respectively, and there was no difference between L-R and R-L orientations ($t < .1$). The alternative test for the central tendency showed that the standard deviation of responses was significantly smaller than the standard deviation of the linear $\text{End}_{[\text{Start-Target}]}$ values, $t_{L\text{-}R}(19) = 5.95$, $p < .001$, and $t_{R\text{-}L}(19) = 5.55$, $p < .001$. The regression means for L-R and R-L were very close to and not statistically different from .5 (L-R: .49, $t < 1.5$; R-L: .5, $t < 1$) and each other ($t = 1.63$, $p = .11$).

The effect of linear $\text{End}_{[\text{Start-Target}]}$, Start, and Length on error. The results for linear $\text{End}_{[\text{Start-Target}]}$, Start, and Length are shown in Figure 7. There was a significant negative trend as a function of linear $\text{End}_{[\text{Start-Target}]}$: Start-controlled, L-R: $t(19) = 7.50$, $p < .001$, $R^2 = .34$; Start-controlled, R-L: $t(19) = 5.31$, $p < .001$, $R^2 = .29$; Length-controlled, L-R: $t(19) = 6.52$, $p < .001$, $R^2 = .27$; Length-controlled, R-L: $t(19) = 7.52$, $p < .001$, $R^2 = .34$. A repeated measures 2 × 2 analysis of variance on slopes (block: Start/Length; orientation: L-R/R-L) showed that neither main effects nor their interaction was significant (all Fs < 1.2). The mean slopes were $\beta_{L\text{-}R/Start} = -.191$, $\beta_{L\text{-}R/Length} = -.181$, $\beta_{R\text{-}L/Start} = -.183$, and $\beta_{R\text{-}L/Length} = -.20$. The beta values for linear $\text{End}_{[\text{Start-Target}]}$ were very consistent at a within-subject level, with the correlation between them for different experimental blocks ranging from .71 to .80 (all ps < .001). In addition, there was a strong negative correlation between betas and R^2 of the linear models (all $\tau > .43$, $p < .01$).

A t test on the regression slopes for Start showed that they did not statistically differed from zero and there was no difference between L-R and R-L (all ts < 1.46, $p > .16$). The distribution for betas of Length visibly deviated from normality, approximating the form of a nonsymmetrical one-tailed Gaussian. Therefore, we used Wilcoxon signed-rank test instead of t test. The beta values for two line orientations were significantly smaller than zero (L-R: $z = 3.81$, $p < .001$, $R^2 = .07$; R-L: $z = 3.88$, $p < .001$, $R^2 = .06$) and were not different from each other ($z < 1$). The betas for Length were negative in 37 cases out of 40 (median $\beta = -.007$ per 10 number units for both L-R and R-L). Nonparametric correlation analysis showed that the correlation between the betas of Length for L-R and R-L was very close to significance (Kendall's $\tau = .32$, $p = .055$), suggesting that the effect was moderately consistent at a within-subject level.

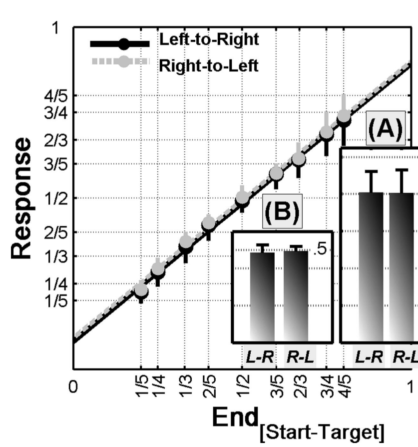

Figure 6. Line construction task. The group means and their standard deviations of responses with the linear $\text{End}_{[\text{Start-Target}]}$ as a predictor. Figure 6A shows that the slopes of linear regression models are significantly smaller than 1, indicating the presence of linear compression in the data, that is, a central tendency. Figure 6B shows the bars for the regression means. The latter were not statistically different from .5, implying that the crossover from overestimation to underestimation took place at the middle of the interval for both line orientations. L = left; R = right.

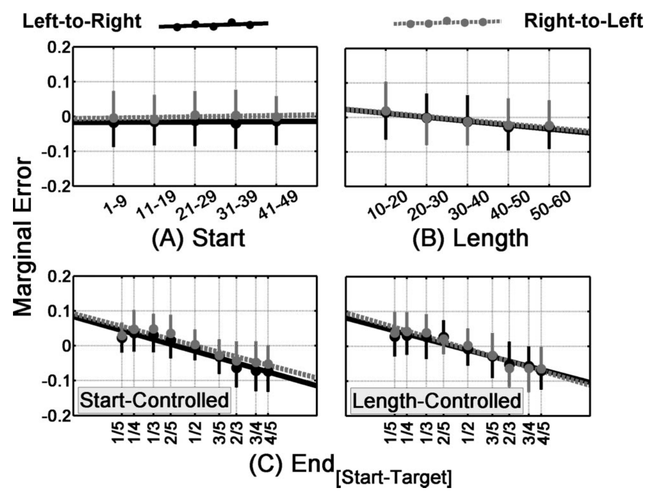

Figure 7. Line construction task results. The marginal effects of interval Start, interval Length, and linear $\text{End}_{[\text{Start-Target}]}$ on the errors in responses. The error bars show the mean standard deviations. The effect of Start was no longer positive, whereas the effect for Length was consistently negative, contrary to the predictions for the log-scale hypothesis. Neither experimental variable showed a significant difference between left-to-right and right-to-left orientation.

One of the possibilities why no significant effect of Start was found is that there was a nonzero correlation between pairs of independent variables. In the Start-controlled blocks, the values for Length were generated randomly but were not orthogonal to Start by design. Although the group mean correlation between Length and Start in the Start-controlled blocks was close to zero, it ranged from −.24 to .32 for individual participants. Consequently, we asked whether the positive trend for Start did not show up because it was counterbalanced by a stronger and more consistent effect of Length in this task. There is an indirect way of inquiring into this issue. It can be expected that the counterbalancing would reveal itself as a negative correlation between individual beta values for Start and the strength of the correlation between Length and Start for each subject. Testing this hypothesis, however, did not support that the counterbalancing took place, as the strength of the correlation was found to be negligible ($p > .67$ for either line orientation).

Discussion

The results of Experiment 2 demonstrated that the approximate estimation of symbolic numerical magnitudes was performed on the linear scale. We showed that the linear regression model predicted the data better than the full model with a weight for a logarithmic component. The analytical method, decoupling the effects of linear $End_{[Start\text{-}Target]}$, Start, and Length, provided a further support for the strong linear hypothesis. The logarithmic hypothesis predicts a negative trend for linear $End_{[Start\text{-}Target]}$, complemented by the negative trend for Start and the positive trend for Length. However, we found that the effect of Start was nonsignificant, whereas the negative trend for Length was very consistent at a between-subjects level and moderately consistent at a within-subject level. The results suggest that the negative trend for linear $End_{[Start\text{-}Target]}$ was due to the central tendency bias. The regression means were statistically indistinguishable from .5, implying that the switch from overestimation to underestimation was in the middle of a numerical interval. Once again, the strength of the central tendency was a marker of randomness in performance, such that the stronger effect was accompanied with lower variances explained by the regression models.

The effect of Length essentially implies that participants tended to construct longer extensions to the line as the numerical difference between Target and Start increased, resulting in an increasing underestimation of $End_{[Start\text{-}Target]}$. The findings of this effect with the null effect for Start are in a striking contrast to the results of the line-marking task, where participants were biased by the magnitude of Start, not Length. Taking into account that the effect cannot be ascribed exclusively to Length (i.e., the difference between Start and Target), as the latter should correlate with the difference between other stimulus magnitudes, one can interpret the results as showing that participants' decisions in the line construction task were biased by the magnitude of numerical differences between numbers rather than the individual absolute magnitudes of the latter.

General Discussion

The investigation of the subjective scale for magnitude representations cannot take for granted the notion that the type of the scale can be inferred directly from the stimulus–response mapping. The aim of our study was to describe and exploit a method that addresses the theoretically motivated problem of differentiating between the linear and logarithmic scaling hypotheses for numerical magnitudes, while controlling for the bias in the decision making. The method exploits the idea that a scale is defined not by its appearance but by the transformational rules according to which magnitudes get assigned (Luce, 1959; Stevens, 1968). The method was implemented in a modified version of the number-to-position paradigm, where participants were required either to mark the position of an Arabic numeral within an interval of varying length and start or to complete such interval by constructing the line of an appropriate length. The modification also allowed us to avoid the shortcomings of the previous studies, where a digit number was positioned within standardized intervals. This sort of interval is easy to deal with for adult participants, and hence, the null result does not provide convincing evidence for the linearization of the numerical scale.

The results show that responses were derived from a linear subjective scale irrespective of whether participants were required to mark or construct the line. We used two complementary approaches to the data analysis, those of model fitting and analytical decomposition, and none of them revealed signatures of a logarithmic trend in responses. We found the presence of the central tendency, that is, a form of linear compression where small numbers in an interval are overestimated and large numbers are underestimated. This effect has been observed in diverse experimental settings (e.g., Huttenlocher, Hedges, & Duncan, 1991; Matthews & Stewart, 2009; Nakamura, 1987; Preston & Baratta, 1948; Sheth & Shimojo, 2001) and is likely to represent a general response bias under uncertainty. This view is supported by the findings in our study, showing that the responses were more randomly distributed for the participants with a stronger central tendency.

One of the possible reasons why the performance in our tasks revealed a perfectly linear mapping is that the magnitudes presented as Arabic numerals are more susceptible to algorithmic computations than those presented nonsymbolically. For example, the judgments could be partially based on the analyses of the decade differences. That might impose a roughly linear structure on the estimation, even though the latter remained approximate. Consequently, a generalization of our findings to the other formats for numerical magnitudes (i.e., dots or the number of tones) should be treated with caution. However, the combinatorial method, when applied to numerosities presented nonsymbolically, provides an opportunity to resolve the issue.

The mental number line hypothesis plays an important role in our understanding of the processes underlying the representations of number. The hypothesis can be characterized by two statements. First, the mental number line is held to represent magnitudes in one orientation only, left to right in our alphabetic cultures (Shaki & Fischer, 2008). Second, the mental number line is held to be the representation of numerical magnitudes automatically and obligatorily activated in all numerical tasks. This implies that the performance in the right-to-left condition would require some sort of mental rotation, which could have the effect of producing more internal noise, and hence more variability in responses. Despite the fact that we explicitly used a number line analogy in the design of the study, our results did not show any accuracy differences between L-R and R-L conditions in either task. However, as most evidence for an oriented representational continuum is derived from reaction time data, it is possible that the accuracy measures in the absence of a limit on reaction times may be insufficiently sensitive to detect the costs.

Our results also demonstrate that the performance was affected by task-specific effects. In the line-marking task, participants tended to overestimate target magnitude when the start of an interval increased, whereas in the line construction task they overestimated when the length of the interval increased. These particular trends are not compatible with the logarithmic mapping that predicts a greater overestimation for smaller starts in the line-marking task and for smaller lengths in the line construction task. Meanwhile, the finding that participants were biased in different ways clearly indicates that marking magnitude and constructing magnitude emphasized different numerical relations and that, other factors being equal, the way participants manipulate and combine the quantities can have a specific effect on an estimation outcome. Although it is not possible to answer what factors were critical, two alternatives may be considered. One possibility is that the process of estimation was tuned to the actual mode of behavior. More concretely, if the task required an assignment of a discrete magnitude to a location on the physical line, then the relations between numbers were represented in terms of their absolute magnitudes. If the task was to complete an interval, that is, something that extends from *A* to *B*, then those relations were represented in terms of differences between numbers. Another possibility is that the task-specific effects might result from the between-task difference in the position of a target with respect to an interval. If the target lay outside the interval, as it was in the construction task, that might induce participants to estimate in terms of how far it is positioned from the other numbers (i.e., in terms of numerical distances), instead of how big its magnitude is with respect to others.

Alternatively, two possible mechanisms predicting the overestimation of a target number can be envisaged in terms of the mental number line hypothesis. The first possibility is that the bias can be a result of an attentional shift, evoked by the canonical L-R orientation of the line. As proposed by Lourenco and Longo (2009), the amount of compression in the mental number line may depend on whether some part of the number line is in a focus of attention. The segment of mental number line becomes decompressed when it is in the focus; otherwise it returns to a default compressed state. For example, in our line-marking task, participants may tend to fixate on the interval between Start and Target more than on the interval between Target and End, as the position of Target should be marked at some distance from the interval Start. As a result, the unattended part may become represented compressively, resulting in the overestimation. This idea seems to account for the findings that the overestimation was somewhat smaller for the noncanonical orientation. Here the magnitude of interval End was presented in the location of Start for a canonically oriented interval and therefore could have a greater saliency than in the noncanonical condition.

The second possibility is that the target overestimation may be closely related to so-called operational momentum bias, reported for the operations of addition and subtraction (Knops, Viarouge, & Dehaene, 2009; McCrink, Dehaene, & Dehaene-Lambertz, 2007). The phenomenon is characterized by participants' tendency to increasingly overestimate for addition and underestimate for subtraction as the true sum or the true difference increase. It is thought that the effect arises from dynamic representations of symbolic operations on the mental number line and can be described with a physical analogy: Before a moving body stops under the effect of counteracting forces, it travels some distance, which is greater for heavier bodies. In this analogy the body is a number, the mass is its magnitude, and the path along which the body moves is the mental number line. Given that, from a mental number line perspective, finding a location of a number within an interval could require moving along the mental continuum from left to right, the process of mapping numbers and performing addition appear to be operationally similar to each other and can cause similar behavioral outcomes.

The main reason why both possibilities provide at best a partial interpretation for our results is that an obligatory mapping that is automatic and beyond cognitive control, as required by the mental number line hypothesis, presumes a unique mode for representing the relation between magnitudes. The contrast of the numerical factors biasing performance in our tasks clearly shows that it was not the case. In keeping with the physical analogy, the performance in the line construction task would require a different sort of dynamics, as compared with the line-marking task: Here the overestimation was caused not by the mass of the body (i.e., number absolute magnitude) but by the differences between two masses (i.e., numerical difference). If the mental number line allows for such flexibility, then it represents an adaptive strategy used to operate with abstract quantities: convenient and conventional but not obligatory.

The question remains what these task-specific and number-related effects tell us about magnitude representation and processing. First, the presence of a consistent bias per se suggests that there is a capacity limit that constrains representing the relations between two pairs of numbers simultaneously. If the difference between Start and Target versus the difference between Start and End could be optimally contrasted, then participants would not consistently weight the difference between one pair of numbers more than the other. Second, our findings suggest that the choice of the format for representing numerical relations on the numerical scale depends on the particular task requirements. In one task, the response bias was triggered by the absolute magnitudes of the numbers, with no effect of differences between numbers, and vice versa for the other task. As this contrast suggests, the relations between magnitudes would be encoded as either the difference between two magnitudes or the magnitude of their difference. From the point of view of formal arithmetical rules, the distinction is meaningless because the statements are numerically equivalent. However, from the point of view of the mental operations with magnitudes, each way of encoding may be better suited than the other for a numerical problem at stake.

In summary, our results imply that the subjective scale of numerical magnitudes in adults is linear. Mapping from the subjective scale into behavior is affected by response biases and can be deployed flexibly according to task demands.

References

Banks, W. P., & Coleman, M. J. (1981). Two subjective scales of number. *Perception & Psychophysics, 29,* 95–105. doi:10.3758/BF03207272

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

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology, 103,* 1131–1136. doi:10.1037/h0037361

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

Dehaene, S. (1997). *The number sense.* New York, NY: Oxford University Press.

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

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and mental magnitude. *Journal of Experimental Psychology: General, 122,* 371–396. doi:10.1037/0096-3445.122.3.371

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:10.1126/science.1156540

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44,* 43–74. doi:10.1016/0010-0277(92)90050-R

Gallistel, C. R., & Gelman, R. (2005). Mathematical cognition. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 559–588). New York, NY: Cambridge University Press.

Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. *American Journal of Psychology, 60,* 1–29. doi:10.2307/1417326

Hollingworth, H. L. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology, and Scientific Methods, 7,* 461–469. doi:10.2307/2012819

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience, 6,* 435–448. doi:10.1038/nrn1684

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98,* 352–376. doi:10.1037/0033-295X.98.3.352

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106,* 1221–1247. doi:10.1016/j.cognition.2007.06.004

Jewell, G., & McCourt, M. E. (2000). Pseudoneglect: A review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia, 38,* 93–110. doi:10.1016/S0028-3932(99)00045-7

Knops, A., Viarouge, A., & Dehaene, S. (2009). Dynamic representations underlying symbolic and nonsymbolic calculation: Evidence from operational momentum effect. *Attention, Perception, & Psychophysics, 71,* 803–821. doi:10.3758/APP.71.4.803

Krueger, L. E. (1982). Single judgments of numerosity. *Perception & Psychophysics, 31,* 175–182. doi:10.3758/BF03206218

Longo, M. R., & Lourenco, S. F. (2007). Spatial attention and the mental number line: Evidence for characteristic biases and compression. *Neuropsychologia, 45,* 1400–1407. doi:10.1016/j.neuropsychologia.2006.11.002

Lourenco, S. F., & Longo, M. R. (2009). Multiple spatial representations of number: Evidence for co-existing compressive and linear scales. *Experimental Brain Research, 193,* 151–156. doi:10.1007/s00221-008-1698-9

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review, 66,* 81–95. doi:10.1037/h0043178

Matthews, W. J., & Stewart, N. (2009). Psychophysics and the judgment of price: Judging complex objects on a non-physical dimension elicits sequential effects like those in perceptual tasks. *Judgment and Decision Making, 4,* 64–81.

McCrink, K., Dehaene, S., & Dehaene-Lambertz, G. (2007). Moving along the number line: Operational momentum in nonsymbolic arithmetic. *Perception & Psychophysics, 69,* 1324–1333. doi:10.3758/BF03192949

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology, 103,* 503–515. doi:10.1016/j.jecp.2009.02.006

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215,* 1519–1520. doi:10.1038/2151519a0

Nakamura, M. (1987). Central tendency of decision making under uncertainty. *Behaviormetrika, 22,* 29–48. doi:10.2333/bhmk.14.22_29

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology, 91,* 191–205. doi:10.1037/h0031854

Parkman, J. M., & Groen, G. J. (1971). Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology, 89,* 335–342. doi:10.1037/h0031198

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44,* 547–555. doi:10.1016/j.neuron.2004.10.014

Preston, M. G., & Baratta, P. (1948). An experimental study of the auction-value of an uncertain outcome. *American Journal of Psychology, 61,* 183–193. doi:10.2307/1416964

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology, 83,* 274–278. doi:10.1037/h0028573

Shaki, S., & Fischer, M. H. (2008). Reading space into numbers—A cross-linguistic comparison of the SNARC effect. *Cognition, 108,* 590–599. doi:10.1016/j.cognition.2008.04.001

Sheth, B. R., & Shimojo, S. (2001). Compression of space in visual memory. *Vision Research, 41,* 329–341. doi:10.1016/S0042-6989(00)00230-3

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:10.1111/1467-9280.02438

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 683–692. doi:10.1037/0096-1523.16.4.683

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York, NY: Wiley.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64,* 153–181. doi:10.1037/h0046162

Stevens, S. S. (1968). Measurement, statistics, and the schemapiric view. *Science, 161,* 849–856. doi:10.1126/science.161.3844.849

Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review, 78,* 426–450. doi:10.1037/h0031324

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10,* 130–137. doi:10.1111/1467-9280.00120

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Neglect disrupts the mental number line. *Nature, 417,* 138–139. doi:10.1038/417138a

Received August 6, 2010
Revision received May 3, 2011
Accepted May 4, 2011 ■