

Knowledge on the line: Manipulating beliefs about the magnitudes of symbolic numbers affects the linearity of line estimation tasks

Dana L. Chesney · Percival G. Matthews

Published online: 17 May 2013
© Psychonomic Society, Inc. 2013

Abstract It has been suggested that differences in performance on number-line estimation tasks are indicative of fundamental differences in people's underlying representations of numerical magnitude. However, we were able to induce logarithmic-looking performance in adults for magnitude ranges over which they can typically perform linearly by manipulating their familiarity with the symbolic number formats that we used for the stimuli. This serves as an existence proof that individuals' performances on number-line estimation tasks do not necessarily reflect the functional form of their underlying numerical magnitude representations. Rather, performance differences may result from symbolic difficulties (i.e., number-to-symbol mappings), independently of the underlying functional form. We demonstrated that number-line estimates that are well fit by logarithmic functions need not be *produced* by logarithmic functions. These findings led us to question the validity of considering logarithmic-looking performance on number-line estimation tasks as being indicative that magnitudes are being represented logarithmically, particularly when symbolic understanding is in question.

Keywords Numerical cognition · Symbol comprehension · Estimation · Number sense

It has become increasingly popular for researchers to interpret the spacing and precision of responses on number-line estimation tasks, in which people place symbolic numbers (e.g., 74) on lines anchored by other symbolic numbers (e.g., 0–100), as being directly indicative of the spacing and precision of the mental representations of numerical magnitudes referenced by symbolic numbers. Particularly, changes in task performance are interpreted as reflecting developmental changes in children's underlying representations of numerical magnitudes (e.g., Siegler & Booth, 2005; Siegler, Thompson, & Opfer, 2009). According to the “log-to-linear shift” hypothesis, children initially rely on logarithmic representations of numerical magnitude to complete these tasks—representations tied to the system that allows us to perceive numerosities of dot arrays without counting. The hypothesis holds that with experience and education, children develop linear representations of numerical magnitudes that they can apply to this task (Siegler & Opfer, 2003). Some have claimed that this apparent shift is evidence that the functional form of number-line estimation performance reflects that of underlying magnitude representations (e.g., Dehaene, Izard, Spelke, & Pica, 2008).

However, successful number-line estimation requires more than simply assigning perceived magnitudes to the appropriate spatial line positions. One must translate between symbolic numbers and their referenced magnitudes before translating the magnitudes to spatial line lengths. Missteps in translating between symbolic numbers and numerical magnitudes may alter the shapes of estimation patterns, independently of the underlying functional form of those magnitudes' representations. This is particularly likely to occur in cases in which the mappings between symbolic numbers and the magnitudes they are meant to represent are not strongly established (Barth & Paladino, 2011; Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008).

Electronic supplementary material The online version of this article (doi:10.3758/s13423-013-0446-8) contains supplementary material, which is available to authorized users.

D. L. Chesney (✉) · P. G. Matthews
Department of Psychology, University of Notre Dame,
118 Haggard Hall,
Notre Dame, IN 46556, USA
e-mail: dlchesney@gmail.com

P. G. Matthews
e-mail: pmatthews@wisc.edu

P. G. Matthews
Department of Educational Psychology, University of Wisconsin,
1025 W. Johnson Street, #884,
Madison, WI 53706-1796, USA

We must bear in mind that such mistranslations can affect performance, or we risk conflating effects of basal magnitude representation with those due to symbol/referent mapping. This issue is especially important when evaluating the log-to-linear shift hypothesis. A child must recognize a numeral before placing it on a number line with any consistency. Consider a young child who cannot correctly name the numeral 74. There is no *a priori* reason why this child would link the symbol “74” to a definite magnitude. Research has shown that many kindergartners cannot properly label written numerals above 10 (e.g., Clarke & Shinn, 2004; Wright, 1991). This poses a problem for interpreting such children’s performance on number-line estimation tasks ranging from 0 to 100. It is reasonable to expect disorganized behavior when children encounter number symbols without clear referents. This could lead to flat performance functions in unfamiliar ranges.

Some researchers (e.g., Ebersbach et al., 2008; Moeller, Pixner, Kaufmann, & Nuerk, 2009) have suggested that such situations could yield bilinear performance patterns resembling logarithmic functions. For example, children’s performance may *appear* logarithmic if they perform linearly on the range that they understand, but heap unfamiliar symbols into a general class of “big” numbers, placing them without concern for specific value on the “big” end of the line. This predicts excessively steep slopes in the known range and overly flat slopes in the unknown range. Such bilinear performance yields functions that are reasonably well fit by logarithmic curves. Under this hypothesis, logarithmic-looking performance is due to symbol-mapping difficulties, not to reliance on logarithmic magnitude representations. Stated alternatively, behavior that is well fit by logarithmic functions is not necessarily produced by logarithmic functions.

Several studies have shown that young children perform linearly within lower number ranges and curvilinearly on tasks including higher number ranges (e.g., Siegler & Booth, 2004, 2005), which young children are likely to have poorly mapped (see Clarke & Shinn, 2004; Wright, 1991). Notably, Barth and Paladino (2011) suggested that kindergartners’ curvature on 0–100 number lines results from symbolic uncertainty inducing underestimation of high anchor values—not from underlying magnitude spacing. However, no work has experimentally manipulated the interpretations of symbolic numbers used in line estimation tasks, leaving any possible effects of age-related symbolic uncertainty to speculation. Indeed, we know of no studies investigating how symbol-to-magnitude mismappings affect number-line estimation.

We believe that it is important to address this gap in the literature, particularly given prior findings that providing incorrect information about specific array cardinalities can systematically alter participants’ estimates of other nonsymbolic numerosities (Izard & Dehaene, 2008). We sought to experimentally test the hypothesis that symbol-to-magnitude

mismatching can yield logarithmic-looking performance on line estimation tasks, even when participants clearly can represent numbers linearly in the tested range. We used different symbolic number formats to manipulate adult participants’ mappings between to-be-placed number symbols and numerical magnitudes. This let us test specific predictions about the resulting functional forms of their performance on number-line estimation tasks. We used formally educated adult participants to ensure that any observed effects resulted from symbolic factors, as opposed to hypothesized differences in basal magnitude representation that might confound children’s performance.

Experiment 1

Adults completed several number-line estimation tasks, some using unfamiliar notation to induce participants to mismap numeric symbols to magnitudes. We hypothesized that their performances would be linear with slopes reflecting the ratio between the presumed and actual symbol magnitudes, except when participants thought that the stimuli exceeded the lines’ upper anchors. For these unexpectedly high stimuli, we predicted that participants would place responses together near the upper anchor, yielding a relatively flat slope approaching zero. We predicted that when analyzed together, estimates for stimuli above and below the assumed high anchor value would resemble a logarithmic curve, despite not being produced by a logarithmic function. Performance was predicted to become more accurate when assumptions about the upper anchor values were corrected with a calibration slide.

Method

Participants

A group of 68 University of Notre Dame undergraduates (29 female, 39 male; ages 18–22 years) participated for course credit. Two participants were dropped (for noncompliance or experimenter error), leaving a final analytic sample of 66.

Materials

The participants completed the experiment individually. All stimuli were presented on iMac 5.1 computers using Superlab 4 software (Cedrus Corporation, 2007).

Stimuli

Each task used 11 different numerical stimuli, presented twice each. These were adapted from Siegler and Opfer’s (2003) 0–1,000 number-line stimuli by multiplying values by a scaling factor of $10^{1.5}$. The stimuli were presented in

either decimal (64, 128, 192, 576, 800, 2,272, 4,960, 7,360, 12,480, 18,390, 25,920) or exponential ($.002 \times 10^{4.5}$, $.004 \times 10^{4.5}$, $.006 \times 10^{4.5}$, $.018 \times 10^{4.5}$, $.025 \times 10^{4.5}$, $.072 \times 10^{4.5}$, $.157 \times 10^{4.5}$, $.233 \times 10^{4.5}$, $.395 \times 10^{4.5}$, $.582 \times 10^{4.5}$, $.820 \times 10^{4.5}$) notation. Each trial involved a 147-mm horizontal line with anchor values printed below the line's left endpoint (always 0) and right endpoint (in exponential notation, $.999 \times 10^{4.5}$, or decimal notation, 31,623). The to-be-estimated numbers appeared ~15 mm above the center of the line.

Design and procedure

Participants were assigned to groups (Group 1, $n = 34$; Group 2, $n = 32$) that completed different experimental tasks, but the same control tasks (described below). In all tasks, participants indicated estimates via mouse click. The trials timed out after 15 s.

Group 1: Decimal stimuli with exponential anchors (Dec-Exp) in experimental tasks We designed this task to create a situation in which naive adults incorrectly underestimate upper anchor values. The upper anchor was symbolized as $.999 \times 10^{4.5}$. Pilot participants never correctly assessed this value as 31,623. Instead, most rounded to $.999 \times 10^4$ (~10,000), underestimating the anchor by a factor of 3.16. The to-be-placed stimuli, in contrast, were presented in standard decimal format, which should be well understood. Participants were expected to initially overestimate the positions of stimuli relative to their correct positions by the ratio of the true to the assumed anchor magnitudes (31,623:10,000), performing linearly with a slope approaching 3.16 for stimuli below the assumed upper anchor value (typically 10,000). We further predicted that when confronted with stimuli greater than this presumed value, participants would compress estimates into a small space at the right end of the line.

Group 2: Exponential stimuli with decimal anchors (Exp-Dec) in experimental tasks We designed this task to create a situation in which naive participants would underestimate the ratio of the stimuli to the upper anchor. This reversed the setup of the Dec-Exp manipulation, such that we expected the stimuli, rather than the upper anchor, to be underestimated by a factor of 1/3.16 (about 0.32). We predicted that this would yield a linear performance pattern with a slope approaching 0.32 over the entire stimulus range, as no values would violate expectations by exceeding the anchor value.

Both groups: Control tasks Control tasks displayed both stimuli and anchors in the same format, either decimal (Dec-Dec) or exponential (Exp-Exp).

Stimulus order For the Dec-Exp tasks, stimuli were presented in three blocks. First, in the *low* block, stimuli of values under 10,000 were presented in random order. Next, in the *high* block, stimuli over 10,000 were presented in random order. Finally, in the *mixed* block, the entire range of stimuli were presented in random order. We used this low-high-mixed sequence out of concern that, when confronted with stimuli that exceeded their evaluation of the upper anchor, participants might recalibrate that evaluation upward to be consistent with their assumptions that all stimuli should be less than the upper anchor. This sequence was also followed for the control tasks, to confirm that the ordering itself did not yield bilinear performance (it did not; see below). For the Exp-Dec tasks, we had no expectation that the stimuli would be perceived as greater than the anchor. Thus, the stimuli were simply presented twice in random order, with each stimulus being presented at least once before any were repeated.

Task order Each participant completed all tasks in an hour-long session that also included other estimation tasks not discussed in this report. The participants completed their assigned experimental tasks (Group 1, Dec-Exp; Group 2, Exp-Dec) twice, once before (*precalibration*) and once after (*postcalibration*) viewing a calibration slide indicating which stimulus corresponded to the midpoint of the line (Fig. 1). Thus, the task order acted as an experimental condition. We predicted that participants would use the calibration slide to form an accurate mapping of the exponential notation, leading to linear performance with slopes approaching 1.0 in the postcalibration experimental tasks. Group 1 completed the Dec-Dec control task before the experimental tasks and the Exp-Exp control after, whereas Group 2 reversed this order.

Results and discussion

Analyses

We used Booth and Siegler's (2006) method for all evaluations of differences in fit among linear, logarithmic, and bilinear models: We compared the absolute values of the differences between the group's median estimates and the values predicted by the best-fitting functions via paired-samples *t* tests. Because three tests were needed to compare these model fits, we used a Bonferroni correction, requiring $p < .017$ for statistical significance. When possible, median estimates from the first and second presentations of the stimuli for a given task were included as separate data points. The results are summarized in Table 1 and Fig. 1.

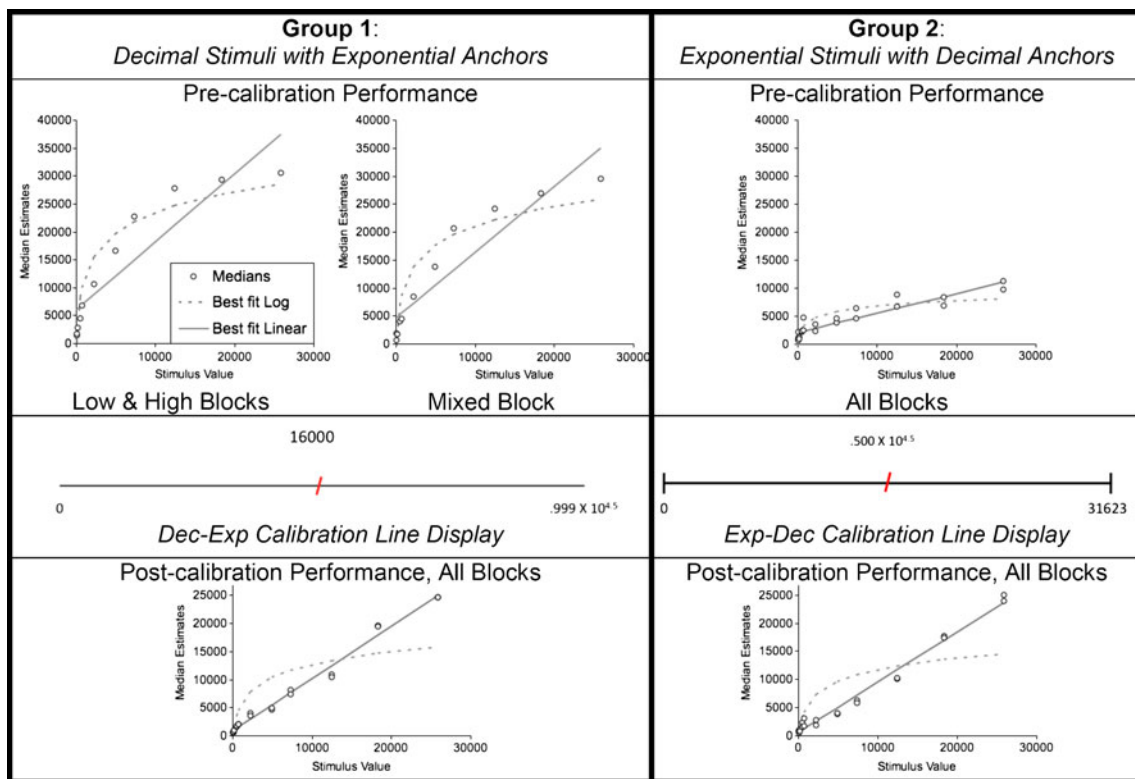


Fig. 1 Experiment 1: Median estimates and best-fitting linear and logarithmic functions for pre- and postcalibration Dec–Exp and Exp–Dec tasks, presented with the tasks’ calibration lines

Control tasks

Linear functions with slopes near 1.0 accounted for at least 99 % of the variance in each control condition. Responses for both the Dec–Dec and Exp–Exp control tasks were better fit by linear than by logarithmic functions, even though blocks used the low–high–mixed task order.

Dec–Exp tasks

We separately analyzed the patterns from first and second presentations of stimuli in the precalibration Dec–Exp blocks (the low and high blocks vs. the mixed block). As predicted, there was a stark contrast in performance for stimuli below 10,000 and above 10,000, not only in the low and high blocks, but also in the mixed block. For stimuli below 10,000, the aggregate performance was linear with a slope well above 1.0, approaching the predicted slope of 3.16 (low and high, 2.82; mixed, 2.56). In contrast, performance on values over 10,000 yielded very shallow slopes (low and high, 0.20; mixed, 0.39).

When considering the full range of values, the resulting function took on the predicted curvilinear appearance. Although there was no statistically significant difference between the fits of the linear and logarithmic functions, the absolute value of the variance explained was greater for the logarithmic than for the linear model. As predicted, both

models were outperformed by a bilinear model separately fitting values above and below 10,000 ($R^2_{bi} > .99$). Still, despite the clear bilinear nature of the participants’ performance, the best-fitting logarithmic function had an R^2 of over .9. Similar logarithmic fits in children’s performance functions have been considered evidence that children use logarithmic basal magnitude representations (Siegler & Booth, 2004). In contrast, the performance on the Dec–Exp task after the presentation of the calibration slide was clearly linear and not logarithmic. With the presentation of a single slide, performance changed dramatically. Patterns in the individual data mirrored this aggregate tendency (see the [online supplement](#) for individual analyses).

Exp–Dec tasks

Performance on the precalibration Exp–Dec task was well fit by a linear function, although the linear fit was not significantly better than the logarithmic fit ($R^2_{lin} = .87$ vs. $R^2_{log} = .80$). Furthermore, participants typically placed all stimuli on the leftward third of the line, yielding a slope of 0.35 ($SE = 0.03$). This was close to the slope of 0.32 that we hypothesized would result if participants thought that $10^{4.5}$ equaled 10,000. Performance changed dramatically following the presentation of the calibration slide. After calibration, performance was clearly better fit by a linear function. Furthermore, the slope

Table 1 Linear, logarithmic, and bilinear fit information for the estimation tasks in Experiment 1

Task	Linear R^2	Log R^2	Bilinear R^2	Best-Fitting Function t Tests (Cohen's d)		
	Slope (SE)	Beta (SE)	Slope, <10,000 (SE); Slope, >10,000 (SE)	Log vs. Linear	Linear vs. Bilinear	Log vs. Bilinear
Dec–Dec	1.00 0.91 (0.01)	.70 3,171 (459)	1.00 0.89 (0.04); 0.95 (0.05)	Linear 7.1 (1.5)*	n.s. 0.6 (0.12)	Bilinear 7.2 (1.5)*
Exp–Exp	.99 0.95 (0.02)	.63 3,133 (538)	1.00 0.76 (0.04); 1.19 (0.02)	Linear 6.0 (1.3)*	Bilinear 3.7 (0.8)*	Bilinear 6.4 (1.4)*
Precalibration Dec–Exp, low & high blocks combined	.83 1.22 (0.19)	.92 5,322 (515)	.99 2.82 (0.18); 0.20 (0.03)	n.s. −1.19 [^]	Bilinear 3.75* [^]	Bilinear 4.90* [^]
Precalibration Dec–Exp, mixed block	.87 1.18 (0.15)	.90 4,974 (550)	1.00 2.56 (0.12); 0.39 (0.03)	n.s. −0.26 [^]	Bilinear 3.81* [^]	Bilinear 6.20* [^]
Postcalibration Dec–Exp	.99 0.93 (0.02)	.69 3,223 (482)	.99 0.91 (0.05); 1.03 (0.12)	Linear 6.8 (1.5)*	n.s. 0.0 (0.0)	Bilinear 7.5 (1.6)*
Precalibration Exp–Dec	.87 0.35 (0.03)	.80 1,384 (151)	.91 0.53 (0.09); 0.21 (0.10)	n.s. 1.2 (0.3)	n.s. 0.4 (0.1)	Bilinear 1.9 (0.4) ⁺
Postcalibration Exp–Dec	.99 0.89 (0.02)	.63 2,961 (505)	.99 0.65 (0.06); 1.06 (0.06)	Linear 5.5 (1.2)*	Bilinear 2.4(.5) ⁺	Bilinear 6.0 (1.3)*

[^] $dfs = 10$, otherwise $dfs = 21$; ⁺ $p < .075$ (marginally significant); * $p < .017$ (significance threshold with Bonferroni correction)

increased from 0.35 ($SE = 0.03$) to 0.89 ($SE = 0.02$), approaching the 1.0 slope of unbiased performance. This slope increase was statistically significant ($p < .001$).

Discussion

Performance patterns accorded with our predictions of how mistaken beliefs about magnitudes of numbers presented in exponential notation would influence response patterns. We concluded that the logarithmic-looking performance on the precalibration Dec–Exp task was an artifact of symbol-to-magnitude mismapping, not an indicator of underlying numerical magnitude representations. The fact that performance became linear with a slope approaching 1.0 when confusion about the exponential value was resolved further supports this conclusion.

Experiment 2

In Experiment 1, we assumed that participants would typically estimate that $.999 \times 10^{4.5}$ was about $.999 \times 10^4$, or 10,000, but we did not gather self-reports confirming that this was the case. To gain more direct evidence of participants' misinterpretations, in Experiment 2 we replicated the Dec–Exp task and gathered self-

reports of participants' interpretations of the high anchor value. To decrease the likelihood of participants rounding upward, we used 4.4 instead of 4.5 for the exponential notation format (e.g., $.999 \times 10^{4.4}$). Furthermore, we increased the number of high (>10,000) stimulus values presented, enhancing our ability to detect differences in estimation patterns in this range.

Method

Participants

A group of 24 University of Notre Dame undergraduates participated for course credit (2 male, 22 female; ages 18–20).

Stimuli, materials, design, and procedure

The stimuli and materials used in this experiment were nearly identical to those of Experiment 1, with the following exceptions. First, the line length was increased to 285 mm, to reduce effects of noise due to fine motor control. Second, the Exp–Exp and Exp–Dec tasks were not included in any sessions. Third, in the notation we used the exponent 4.4 instead of 4.5. Pilot participants never correctly determined that $.999 \times 10^{4.4}$ equaled 25,094. Instead, most assumed that

it was $\sim 10,000$. Fourth, we used different stimuli, to fit the new number range and to expand coverage of the high values: 75, 176, 477, 1,306, 2,587, 3,969, 6,029, 7,460, 8,691, 8,741, 9,570, 9,821, 9,872, 12,610, 15,297, 17,659, 19,090, 21,150, 24,943. Finally, at the end of the session, participants were asked what value they had used as an initial estimate for $.999 \times 10^{4.4}$. We predicted that a plot of median estimates versus the stimuli would yield a slope approaching 2.51 (corresponding to $.999 \times 10^{4.4}/10,000$) on uncalibrated Dec–Exp trials for values less than the assumed anchor value of 10,000, and a slope near 0 for values in excess of 10,000. The task and block order matched that of Group 1 in Experiment 1, except that Exp–Exp trials were not included. All tasks were completed during an hour-long session that also included other estimation tasks not discussed in this report.

Results and discussion

The analyses matched those in Experiment 1. Table 2 and Fig. 2 summarize the findings.

Beliefs about high anchors

Most participants reported guessing that $.999 \times 10^{4.4}$ was roughly equal to 10,000, suggesting that they rounded $10^{4.4}$ down to 10^4 . Twelve of the 24 participants reported guessing values between 9,990 and 10,000. One reported guessing 2,500. Five reported guessing between 12,000 and 15,000. Three reported guessing between 99,900 and 100,000, suggesting that they rounded the exponent up to 10^5 . Only three of 24 reported guessing somewhat accurately (22,000–25,000).

Control task

The baseline performance was again linear across the number range tested. A linear function with a slope near 1.0 accounted for at least 99 % of the variance in the Dec–Dec control task. Performance was better fit by a linear than by a logarithmic function.

Dec–Exp tasks

Precalibration The results from the low and high blocks replicated findings from the precalibration trials in Experiment 1. For the low block ($<10,000$), the aggregate performance was best fit by a linear function with a slope of 2.25, closely approaching our predicted value of 2.51. In contrast, the high-block performance ($>10,000$) yielded a very shallow slope of 0.17. Indeed, responses in this range did not seem to follow a well-behaved pattern: A regression including the data from all participants explained a statistically insignificant amount of the variance ($R^2 = .02$). When combined, the aggregate performance on the low and high blocks was better fit by a logarithmic function than by a linear function ($p < .05$). Once again, both models were outperformed by a bilinear model in which values above and below 10,000 were fit separately ($R^2_{bi} = .99$). The individual data mirrored the aggregate response patterns (see the [online supplement](#) for individual analyses).

In contrast, performance in the mixed block was linear (low and high, $R^2_{lin} = .75$, $R^2_{log} = .85$; mixed, $R^2_{lin} = .93$, $R^2_{log} = .78$). This suggests that participants recalibrated their estimates of the exponential anchor $.999 \times 10^{4.4}$ when presented with unexpectedly high decimal stimuli. A similar trend was seen in Experiment 1, though there, it did not change the

Table 2 Linear, logarithmic, and bilinear fit information for the estimation tasks in Experiment 2

Task	Linear R^2	Log R^2	Bilinear R^2	Best-Fitting Function t Tests (Cohen's d)		
	Slope (SE)	Beta (SE)	Slope, $<10,000$ (SE); Slope, $>10,000$ (SE)	Log vs. Linear	Linear vs. Bilinear	Log vs. Bilinear
Dec–Dec	.99 1.03 (.01)	.63 3,706 (472)	1.00 0.91 (0.01); 0.98 (0.04)	Linear 10.2 (1.7)*	Bilinear 4.6 (.8)*	Bilinear 10.9 (1.8)*
Precalibration Dec–Exp, low & high blocks combined	.75 1.07 (.15)	.85 5,125 (524)	1.00 2.25 (0.06); 0.17 (0.04)	Log –2.2 (.5) ⁺	Bilinear 8.0 (1.8) ⁺	Bilinear 4.3 (1.0) ⁺
Precalibration Dec–Exp, mixed block	.93 1.09 (.07)	.78 4,484 (582)	.99 1.50 (0.08); 0.47 (0.06)	Linear 2.9 (.7) ⁺	Bilinear 4.0 (.9) ⁺	Bilinear 5.5 (1.3) ⁺
Postcalibration Dec–Exp	.99 1.03 (.01)	.63 3,715 (473)	1.00 0.914 (0.01); 0.98 (0.05)	Linear 10.3 (1.7)*	Bilinear 6.4 (1.0)*	Bilinear 11.3 (1.8)*

⁺ $dfs = 18$, otherwise $dfs = 37$; ⁺ $p < .05$ (marginally significant); * $p < .017$ (significance threshold with Bonferroni correction)

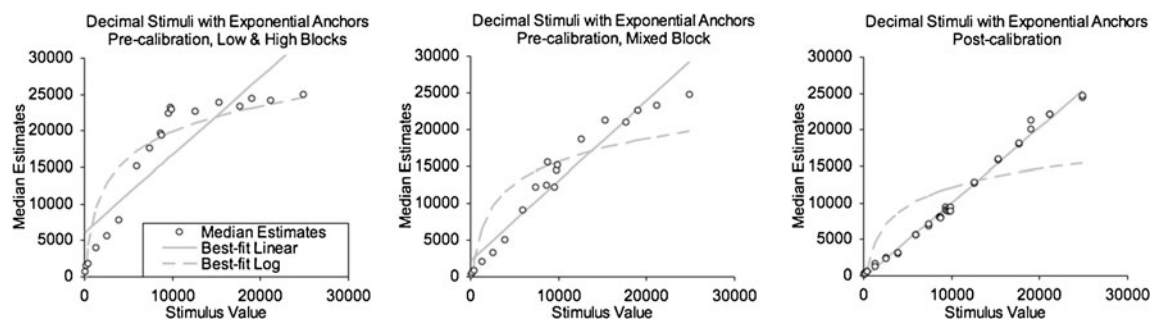


Fig. 2 Experiment 2: Median estimates and best-fitting linear and logarithmic functions for the precalibration Dec-Exp task in the low and high blocks (left) and in the mixed block (center), and for the postcalibration Dec-Exp task (right)

outcomes of the log versus linear comparison (low and high, $R^2_{\text{lin}} = .83$, $R^2_{\text{log}} = .92$; mixed, $R^2_{\text{lin}} = .87$, $R^2_{\text{log}} = .90$).

Postcalibration Performance on the Dec-Exp task following the calibration slide was again better fit by a linear than by a logarithmic function. The linear function accounted for 99 % of the variance. This pattern of performance again confirmed that the “logarithmic” performance seen in the precalibration trials was an artifact of symbol-to-magnitude mismapping.

Discussion

The performance patterns again matched our predictions of how mistaken symbolic beliefs would influence response patterns. The logarithmic-looking performance on the precalibration Dec-Exp task was an artifact of symbol-to-magnitude mismapping, not an indicator of underlying numerical magnitude representations. However, the two experiments’ results did differ, in that Experiment 2’s mixed block showed more evidence of recalibration. This difference was likely attributable to participants in Experiment 2 having been presented with more high values, though sample differences (e.g., gender distribution or time-of-semester effects) may also have played a role.

General discussion

Our findings demonstrated that symbolic misunderstandings can and do result in number-line estimation patterns that can be well fit by logarithmic functions, independently of participants’ underlying basal representations of numerical magnitude. The performances followed predictable trajectories based on participants’ beliefs about relative values of stimuli and anchors. Simply using a confusing notation induced logarithmic-looking performance. Importantly, although performance on these lines was *well fit* by logarithmic functions, there is little reason to believe that it was *produced* by logarithmic functions. Thus, it would not be

logically sound, here, to assume that the performance curves fit by logarithms were produced from underlying logarithmic representations.

Furthermore, our finding that adults abruptly shift to linear performance when shown a single calibration slide parallels Opfer and Siegler’s (2007) finding that children can abruptly shift to linear performance when given a single point on a line: Being given the correct placement of one stimulus value aids in the correct placement of all other values. This opens the possibility that Opfer and Siegler’s shift similarly resulted from symbolic issues separate from any internal log-to-linear shift in magnitude representations. However, the parallel between adult and child performance is imperfect: Adults are familiar with decimal notation and the base-10 system that underlies it. Thus, even given confusing stimuli, their misconceptions do not go very deep. Children’s problems may be more wide ranging, heaping many unfamiliar symbolic numbers into a general class of “big” numbers—grouping the larger stimuli along with upper anchors.

Here, we offer a speculative account for Siegler and Opfer’s (2003) findings that children show logarithmic-like number/position mapping on number-line placement tasks. Children may map “known” values in a roughly linear fashion, on the leftward “known” end of a line, while placing all “big” numbers without concern for specific value on the rightward “big” end of the line. As long as the portion of the line that a child happens to assign to “known” numbers is larger than it should be on the basis of the true values (as among our participants on the precalibrated Dec-Exp tasks), logarithmic-looking performance is predicted. This account is largely compatible with the bilinear models of Ebersbach et al. (2008) and Moeller et al. (2009). Bilinear number-to-location mapping should yield logarithmic-looking performance if the line segment assigned to lower numbers is overly large. Again, we note that this is speculation. We suspect that the final word will not come from experiments using number-line estimation alone. Moreover, manipulations based on perceptual analogues to number lines (e.g., Anobile, Cicchini, & Burr, 2012; Dehaene et

al., 2008) may have only limited applicability, because their stimuli are mainly nonsymbolic. Rather, several numeracy tasks that combine these techniques may be needed to cross-validate the findings. For instance, number comparison tasks can assess the extent to which children's internal representations match performance functions seen for number-line estimation tasks.

Conclusions

Together, these experiments serve as an existence proof that performance on number-line estimation tasks does not always reflect the functional forms of underlying basal magnitude representations. A divergence between performance functions and underlying representations seems particularly likely when misunderstandings of symbol systems prevent accurate mapping between symbolic stimuli and appropriate reference magnitudes. This suggests that number-line estimation tasks may be inappropriate gauges of underlying magnitude representations when symbols are not properly understood, as is likely the case for young children. We therefore urge caution when interpreting the results of number-line estimation tasks, particularly when making inferences about underlying magnitude representations. Investigators should bear in mind that discontinuous or incomplete mappings between symbolic numbers and numerical magnitudes may result in performance that is difficult to interpret. It may be that analogous number-line tasks relying on nonsymbolic numerosity representations (e.g., dot arrays) would be more appropriate, particularly when peoples' understanding of the symbol system is in question.

Author Note P.G.M. was at the Department of Psychology at the University of Notre Dame during the study and is now at the University of Wisconsin–Madison. This research was made possible in part by support from the Moreau Academic Diversity Postdoctoral Fellowship Program of the University of Notre Dame. We thank Nicole McNeil, the members of the Cognition, Learning and Development Lab at the University of Notre Dame, Mike Villano, and Natalie Obrecht for their help and support.

References

- Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, 122, 454–459. doi:10.1016/j.cognition.2011.11.006
- 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
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 42, 189–201. doi:10.1037/0012-1649.41.6.189
- Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review*, 33, 234–249.
- 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
- 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:10.1016/j.jecp.2007.08.006
- Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, 106, 1221–1247. doi:10.1016/j.cognition.2007.06.004
- 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
- Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, 55, 169–195. doi:10.1016/j.cogpsych.2006.09.002
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, 75, 428–444. doi:10.1111/j.1467-8624.2004.00684.x
- Siegler, R. S., & Booth, J. L. (2005). Development of numerical estimation: A review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 197–212). New York, NY: Psychology Press.
- 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
- 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:10.1111/j.1751-228X.2009.01064.x
- Wright, B. (1991). What number knowledge is possessed by children beginning the kindergarten year of school. *Mathematics Education Research Journal*, 3, 1–16.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.