The University Of Sheffield.

This is a repository copy of *Fixated in unfamiliar territory: mapping estimates across typical and atypical number lines*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/206226/

Version: Accepted Version

Article:
Di Lonardo, S.M. orcid.org/0000-0001-6338-9621, Huebner, M.G., Newman, K. et al. (1 more author) (2020) Fixated in unfamiliar territory: mapping estimates across typical and atypical number lines. Quarterly Journal of Experimental Psychology, 73 (2). pp. 279-294. ISSN 1747-0218

https://doi.org/10.1177/1747021819881631

Di Lonardo, S. M., Huebner, M. G., Newman, K., & LeFevre, J.-A. (2020). Fixated in unfamiliar territory: Mapping estimates across typical and atypical number lines. Quarterly Journal of Experimental Psychology, 73(2), 279-294. Copyright © 2019 Experimental Psychology Society. DOI: https://doi.org/10.1177/1747021819881631. Article available under the terms of the CC-BY-NC-ND licence (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Fixated in Unfamiliar Territory:

Mapping Estimates Across Typical and Atypical Number Lines

Sabrina M. Di Lonardo[1]

Matthew G. Huebner[2]

Katherine Newman[3]

Jo-Anne LeFevre[1,2]

[1]Carleton University, Institute of Cognitive Science

[2]Carleton University, Department of Psychology

[3]York University, Department of Psychology

This manuscript was accepted for publication in the *Quarterly Journal of Experimental Psychology* on August 8, 2019. This preprint is the peer-reviewed accepted version but has not yet been copyedited and may differ from the final version published in the journal.

Abstract

Adults (N = 72) estimated the location of target numbers on number lines that varied in numerical range (i.e., typical range 0 – 10,000 or atypical range 0 – 7,000) and spatial orientation (i.e., the 0 endpoint on the left (traditional) or on the right (reversed)). Eye-tracking data was used to assess strategy use. Participants made meaningful first fixations on the line, with fixations occurring around the origin for low target numbers and around the midpoint and endpoint for high target numbers. On traditional direction number lines, participants used left-to-right scanning and showed a leftward bias; these effects were reduced for the reverse direction number lines. Participants made fixations around the midpoint for both ranges but were less accurate when estimating target numbers around the midpoint on the 7,000-range number line. Thus, participants are using the internal benchmark (i.e., midpoint) to guide estimates on atypical range number lines, but they have difficulty calculating the midpoint, leading to less accurate estimates. In summary, both range and direction influenced strategy use and accuracy, suggesting that both numerical and spatial processes influence number line estimation.

Abstract Word count: 181 words

Keywords: number line, estimation, eye-tracking, mathematical cognition, adults

Fixated in Unfamiliar Territory: Mapping Estimates Across Typical and Atypical Number Lines

Imagine that you are preparing to parallel park your car. You notice an available parking spot and ask yourself, “Can I fit?”. This is just one example of how adults rely on estimation abilities in day-to-day life. Estimation involves the approximate calculation of a value, quantity, duration, or size (Levine, 1982; Reys & Bestgen, 1981). Number line estimation, where participants locate a symbolic quantity in physical space (Petitto, 1990; Siegler & Opfer, 2003), is of particular interest because performance is strongly associated with current and future mathematics achievement (Schneider et al., 2018). The goal of the present research was to better understand adults’ strategy use on the number line task by varying numerical and spatial characteristics of number lines and by collecting behavioural observations in the form of eye-tracking data.

In the classic number-to-position version of the number line task, a horizontal line is presented with vertical markers at each end (Siegler & Opfer, 2003; see Figure 1 for examples). Participants are asked to estimate the location of a target number by marking its location on the line. Typically, the line is labelled “0” at the left end and “100”, “1,000”, or “10,000” at the right end. Number lines with base-10 scales and two fixed endpoints are referred to as *typical bounded* number lines (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Siegler & Opfer, 2003). Despite the large amount of research on typical bounded number lines, very few researchers have directly assessed the strategies that people use to estimate locations. Petitto (1990) and Xu and LeFevre (2016) described the use of counting strategies with four to seven-year-old children, and Peeters, Sekeris, Verschaffel, and Luwel (2017) described the use of benchmark strategies in children in grades 3 and 6 when benchmarks were explicitly labelled, however, little is known about strategies used by older children or adults. Furthermore, most research focuses on the outcomes of the estimation process (i.e., the final estimates) and does not explore the cognitive processes people use to make estimates.

Examining outcomes of the estimation process only relays part of the story: We also need information on *how* people estimate (i.e., what strategies they use) to understand *why* they make estimation errors. For example, if an individual is asked to estimate the number 3,500 on a 0 to 7,000 number line and they place their estimate at 3,000, we need to know what strategy they used to understand their performance. Did they miscalculate the benchmark? Were they unable to accurately place the number because of spatial biases? Another thing to consider is what exactly constitutes a strategy. In the present study, we define strategy as “a procedure that is invoked in a flexible, goal-oriented manner and that influences the selection and implementation of subsequent procedures” (Bisanz & LeFevre, 1990, p. 236).

In general, ignoring strategy use may have serious consequences for interpretations of performance patterns in mathematical tasks. For example, research on adults’ use of procedures in mental addition (e.g., LeFevre, Sadesky, & Bisanz, 1996) and on children’s strategy selection on subtraction (Siegler, 1989) shows that it is critical to examine the strategy and procedures that people used to perform the task. Furthermore, averaging across different strategies may lead to erroneous conclusions about the causes of performance differences across items or individuals (Siegler, 1989). Asking participants to report their strategies is one way to explore estimation processes in more detail (e.g., Smith-Chant & LeFevre, 2003). However, self-reports have a variety of limitations (Kirk & Ashcraft, 2001). An alternative that may be more suited to studying strategy use in estimation is to examine patterns of eye tracking (e.g., Sullivan, Juhasz, Slattery, & Barth, 2011). In the present research, eye-tracking data was used to validate and extend conclusions about strategic processes during number line processing.

Benchmark Use in Number Line Estimation

Performance improves with age on the number line task (Booth & Siegler, 2006, 2008; Siegler & Booth, 2004; Schneider et al., 2018). One possibility is that the source of improvement is a result of a shift in children's mental magnitude representation from logarithmic to linear (Booth & Siegler, 2006; Siegler & Booth, 2004; Siegler & Opfer, 2003). When children's estimates are plotted as a function of their actual position, a logarithmic pattern emerges (Dehaene, 2003; Siegler & Booth, 2004; Siegler & Opfer, 2003). However, as their number knowledge improves, and when the values on the scale are within the range of their number knowledge, the relation between estimates and locations becomes linear (Ashcraft & Moore, 2012; LeFevre et al., 2013; Siegler & Booth, 2004).

More recently, it has been suggested that changes in performance as children age reflects improvements in proportional estimation abilities involving the use of internal (i.e., implicit) and external (i.e., explicit) numerical benchmarks (Barth & Paladino, 2011; Huber, Moeller, & Nuerk, 2014; Link, Nuerk, & Moeller, 2014; Slusser & Barth, 2017; Sullivan et al., 2011). Young children start by using one benchmark, the origin or low endpoint (i.e., the "0" start point in typical bounded number lines). Somewhat older children use two benchmarks, that is the low and high endpoints, and eventually, solvers use the low endpoint, the high endpoint, and the internal midpoint to more accurately locate the target number on the line (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Newman & Berger, 1984; Petitto, 1990; Xu & LeFevre, 2016). The increased accuracy in estimation that occurs with age is, in this view, linked to the number of benchmarks that an individual uses while placing an estimate on the number line (Peeters et al., 2017; Slusser, Santiago, & Barth, 2013).

As accuracy increases with age and experience, patterns of error indicate that both external (e.g., endpoints) and internal (e.g., midpoint) benchmarks are used as anchors from which estimates are adjusted (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Peeters et al., 2016; Schneider et al., 2008; Siegler & Opfer, 2003; Sullivan et al., 2011). Therefore, people are expected to show the greatest amount of error on targets located the farthest from commonly used benchmarks (e.g., the low and high endpoints and midpoint; Ashcraft & Moore, 2012; Link et al., 2014; Petitto, 1990; Rouder & Geary, 2014; Siegler & Opfer, 2003; Slusser & Barth, 2017). When estimation error is graphed, use of benchmarks is reflected in the M-shaped pattern of reduced error in the endpoint and midpoint regions (Ashcraft & Moore, 2012). As numerical, spatial, and arithmetic skills increase with age, the "M" becomes progressively flatter (Ashcraft & Moore, 2012).

Barth and Paladino (2011) argued that people consider the location of target numbers in relation to benchmarks (i.e., endpoints and midpoint), relying on successive proportional judgements, rather than in isolation or in relation to the whole number range. Support for this claim was obtained by fitting the results for *n*-cycle power models typically used in proportion-judgement contexts (e.g., Hollands & Dyre, 2000; Spence, 1990) to the pattern of estimates (e.g., Barth & Paladino, 2011; Luwel, Peeters, Dierckx, Sekeris, & Verschaffel, 2018). For example, if a two-cycle power model fits the data for an individual's pattern of estimates, it indicates that the individual is using the low endpoint, high endpoint, and midpoint as benchmarks (Barth & Paladino, 2011). The model reflects a pattern of error such that performance is best around the two endpoints and midpoint and uniformly worse across the rest of the line. Although the use of benchmark-based strategies has been inferred from analyses of estimation error and model fit, relatively few studies have used direct measures, such as eye tracking, to examine whether model fits truly reflect strategy use. Thus, in the present research we used eye-tracking data to address this limitation of the literature.

Eye tracking in Number Line Research

Eye-tracking data provides behavioural evidence about how people use benchmarks to estimate on number lines. Schneider et al. (2008) found that children in grades 1 to 3 made the most fixations around the endpoints and midpoint of the line, consistent with the use of two external and one internal benchmark (see also Heine et al., 2010; Reinert, Huber, Nuerk, & Moeller, 2015). Sullivan et al. (2011) found that adults made more frequent and longer fixations around the midpoint, followed by the low and high endpoints. Hence, fixation frequencies form a W-shaped pattern that is complementary to the M-shaped error pattern observed by Ashcraft and Moore (2012). Thus, eye movement patterns have been used to infer the use of benchmarks on typical bounded number lines.

Eye-tracking data goes beyond just providing information on participants' final estimates, however. Although *n*-cycle power functions imply that participants are using benchmark-based estimation strategies, they are limited in two ways: i) the model fits are purely statistical and not based on behavioural strategy observation and ii) participants' *final* estimates are used to fit the models, even though strategic decisions begin long before a final estimate is made. In contrast, eye-tracking data can be used to obtain information on strategy use by examining participants' first fixations as well as their final estimates. Although eye-tracking data has been used in previous studies (e.g., Heine et al., 2010; Reinert et al., 2015; Schneider et al., 2008; Sullivan et al., 2011), these studies focused on final estimates and assessed performance on typical number lines. In the present study we manipulated the characteristics of the number lines to gain additional insight into the strategies used during the estimation process. More specifically, we manipulated the range and direction of the number lines to examine strategy use for numerical and spatial processes, respectively.

Atypical Number Lines

Atypical Range. Few researchers have examined people's performance on atypical range number lines in which the endpoints are not multiples of 10, 100, or 1,000. In Booth and Newton (2012), children aged 12-14 showed a less linear pattern of estimation on an atypical range number line (i.e., 0 – 6,257) than on a typical range number line (i.e., 0 – 10,000). In Hurst, Leigh Monahan, Heller, and Cordes (2014), adults estimated numbers on three different number lines: Typical range (i.e., 2,000 – 3,000), one-endpoint atypical range (i.e., 0 – 1,258), and two-endpoint atypical range (i.e., 1,639 – 2,897). Linear patterns of estimation were found for the typical and one-endpoint atypical range lines. However, for the two-endpoint atypical range line, a logarithmic model better fit the estimation pattern. Thus, as the number line range became less typical, adults showed estimation patterns similar to those of children. One reason for increased error on atypical range number lines is that people may find it more difficult to identify internal benchmarks, such as the numbers corresponding to the midpoint or quartiles. For example, on a 0 to 1,000 number line, internal benchmarks are either already known or are easy to calculate (i.e., the midpoint is 500 and the quartiles are 250 and 750). However, on a number line ranging from 1,639 to 2,897, it is difficult to calculate that the midpoint is 2,268 and the quartiles are 1,953.5 and 2,582.5; accordingly, people may not use these benchmark strategies.

Luwel et al. (2018) found that, for children in grade 5 and adults, error rates were higher on an atypical (i.e., 367 – 1367) versus a typical number line (i.e., 0 – 1,000). Furthermore, although they observed the classic M-shaped pattern of errors for the typical number line (Ashcraft & Moore, 2012), on the atypical number line errors had a tent-shaped pattern, with uniformly high error across the middle section of the line. These patterns suggested that participants had difficulty identifying the numerical value of the midpoint. Luwel et al. concluded that estimation on atypical number lines is hindered because of the increased difficulty in the mental calculations needed to locate internal benchmarks.

Atypical Direction. Although the majority of atypical number lines presented in the number line estimation task consist of an atypical range, to our knowledge only one study has explored the effects of a number line with an atypical direction (Ebersbach, 2015). Kindergartners and second graders were presented with a 0-100 number line in both the left-to-right and right-to-left orientation. Participants made less accurate estimates on the right-to-left number line compared to the left-to-right number line when they were presented with the right-to-left condition first. The performance of second graders was affected by the orientation to a lesser degree than the performance of kindergartners. These findings suggest that reversing the number line may affect spatial processes used in number line tasks (e.g., LeFevre et al., 2013).

We identified two possible influences on number line processing that may be related to spatial features of the number line that are changed when the endpoints are reversed. First, the effect of number line direction may be linked to culture-typical spatial-numerical associations (Cipora et al., 2019). Number lines are typically presented with smaller numbers on the left and larger numbers on the right. This orientation of number size is common in languages and cultures where text is read left-to-right and thus changing the location of the endpoints, that is, violating the typical left-to-right mapping between number size and location, may disrupt processing and influence participants' strategies (Ebersbach, 2015).

Another possible spatial influence on number line processing that is related to use of benchmarks is that when people are asked to bisect horizontal lines, they underestimate – placing the bisection to the left of the midpoint (Bowers & Heilman, 1980). This leftward bias in locating the midpoint is reminiscent of the neglect for one side of space shown by patients with hemispatial neglect on the contra-lateral side of the damage and therefore was termed ‘pseudo’ neglect. However, the leftward bias may be influenced by the direction in which people start scanning the line. Levander, Tegner, and Caneman (1993) found that the leftward bias was present when individuals started scanning from the left whereas a rightward bias was present when individuals started scanning from the right. To our knowledge, researchers have not assessed whether there is a leftward bias for midpoint estimations in number line tasks. To the extent that the leftward bias is related to scanning patterns, if participants start scanning from right-to-left on reversed number lines, then the midpoint may be estimated as closer to the right endpoint (a rightward bias). In summary, atypical direction number lines might be expected to influence processing by affecting participants’ use of midpoint benchmarks, changing scanning patterns, and/or influencing spatial biases, thus potentially leading to increased error and less linear patterns of performance.

The Present Study

To our knowledge, no studies have specifically examined benchmark strategy use on atypical range and direction number lines. Luwel et al. (2018) fits *n*-cycle power functions to the location of final estimates to infer participants’ use of benchmark-based strategies but they did not assess strategy use directly. Furthermore, they did not manipulate the direction of the number line. In the present study, participants completed a bounded number line estimation task where they saw one of four number lines, as shown in Figure 1 (a – d). A between-subjects design was selected so that participants could not carry strategies over from one type of number line to another. By manipulating range and direction, we examined both the numerical and spatial processing required to estimate on atypical number lines. Two different types of data were analyzed: patterns of fixations and patterns of estimation error (in relation to actual target locations). Analysis of the fixations on the line across trials was used to better understand strategy use. Statistical analyses of error patterns provided information about the relation between strategy use and performance.

(Insert Figure 1a – d approximately here)

Strategy. We expected that the pattern of first fixations would reflect strategy use and would be driven by features of the target number. Specifically, we hypothesized that participants would select a strategy that reflected target magnitude, and thus would make the largest proportion of first fixations around the low endpoint (0%) for low target numbers and the largest proportion of first fixations around the midpoint (50%) or endpoint (100%) for high target numbers. Given that previous work has found that adults tend to use the midpoint as a reference (e.g., Ashcraft & Moore, 2012; Luwel et al., 2018; Xu & LeFevre, 2016), we hypothesize that for high target numbers, individuals will start at the midpoint and adjust their estimates upwards. With the exception of high target numbers that are very close to the endpoint, we hypothesize that individuals will prefer to upwardly adjust (i.e., start at the midpoint and count up) their estimates as opposed to downwardly adjust (i.e., start at the endpoint and count down).

We further hypothesized that participants would use different strategies, dependent on the range and direction of the number line. Because the internal benchmarks are more difficult to calculate for the 7,000-range number line than the 10,000-range number line, we expected participants would make a greater proportion of fixations around the external benchmarks than the internal benchmarks, reflecting less sophisticated strategy choices. With respect to direction, we hypothesized that participants would frequently scan from left-to-right, and thus make a greater proportion of first fixations around the low endpoint for traditional direction number lines versus around the high endpoint (i.e., 10,000 or 7,000) for the reverse direction number line.

Estimation Accuracy. We hypothesized that participants presented with atypical range and reverse direction number lines would make less accurate estimates than those presented with the typical range and traditional direction number lines. Specifically, participants would show a flattened M-shaped pattern of error for typical range (i.e., 10,000) number lines and a tent-shaped pattern of error for atypical range (i.e., 7,000) number lines. These patterns would be consistent with those observed by Ashcraft and Moore (2012) and Luwel et al. (2018). Furthermore, participants were expected to be more accurate for low target numbers than for high target numbers for the 7,000 range number lines. These patterns reflect the reduced accuracy of numerical calculations of midpoint or target magnitudes. This pattern was anticipated for both the traditional and reverse direction because the location of the benchmarks did not physically change.

Method

Participants

Seventy-two undergraduate students participated in the study (M_{age} = 19.7 years; 52.8% female). All participants spoke fluent English, with 79.2% identifying English as their first language. Of the 72 participants, 69 reported first languages that are read from left-to-right. Participants reported having normal or corrected-to-normal vision. Participants received partial course credit for their participation. The study was approved by the Carleton University Ethics Review Board and all participants provided written consent for their participation in the study.

Materials

Demographic questionnaire. Participants were asked to provide demographic information on a paper-and-pencil questionnaire (i.e., age, sex, university major, primary language, handedness).

Calculation Fluency Task (CFT; Sowinski, Dunbar, & LeFevre, 2014). The CFT is a paper-and-pencil test of arithmetic fluency consisting of one page each of double-digit addition (e.g., 41 + 17), double-digit subtraction (e.g., 51 – 19), and multiplication (e.g., 28 x 7). Participants are first presented with six practice trials, two for each operation. Following this, participants are asked to quickly and accurately complete as many questions as possible for each operation. They are given one minute on each of the three pages. Scores are the total number correct across all three operations.

Bounded Number Line Task. Participants were shown one of the four number lines shown in Figure 1(a – d). For the 7,000 range conditions, 33 target numbers were selected to cover the entire number line scale as well as the range within each thousand-unit interval. For the 10,000 range conditions, 51 target numbers were selected for the same reasons as the 7,000 conditions. The targets for the 7,000 range conditions consisted of 16 numbers below the midpoint, 16 numbers above the midpoint, and the midpoint (3,500). The targets for the 10,000 range conditions consisted of 25 numbers below the midpoint, 25 numbers above the midpoint, and the midpoint (5,000). The target numbers were presented in a different random order for each participant. The target numbers for each condition can be found in Appendix A.

Apparatus

Number line stimuli were visually presented on a 21-inch Dell monitor attached to a Dell Precision PWS390 computer, using Windows XP Professional 2002 operating system to run EyeLink 1000 experiment software (SR-Research, Kanata, Ontario, Canada). The computer monitor logo and stand were covered to avoid cueing the participant to the middle of the screen. A specialized camera was used to record participants' eye fixations and saccades for one (i.e., right) eye. The camera emitted a beam of infrared light that was reflected off the participants' cornea. The computer sampled the location of the eye at a rate of 1,000 Hz.

Procedure

Participants were assessed individually on all measures. After reading and signing the informed consent document, they completed the demographic questionnaire and calculation fluency task.

For the number line task, participants were seated approximately 68 cm from the infrared camera and 110 cm from the monitor. The chair height and chin-rest were adjusted for the comfort of each participant and to ensure that the camera had a full view of the participant's eye. Before the experiment began, instructions were presented on the monitor as the experimenter read them aloud. A nine-point calibration was completed followed by two practice trials. When the practice trials were successfully completed, the calibration procedure was repeated, and the participant was instructed to try not to move his/her head until experimental trials were completed.

For the duration of testing, participants kept the index finger of their dominant hand on the spacebar. To begin each trial, participants were instructed to press the spacebar. A black square then appeared in the top right corner of the screen. Once a fixation was recorded in the square, the square would disappear, and the target number and number line were presented simultaneously. All target numbers were presented in the same location in the top left corner of the screen. Participants were instructed to look at the target number and then to quickly and accurately estimate the target's spatial location on the line. When the participant decided on the target's location, they were instructed to fixate on that location and press the spacebar. Upon completion of the number line task, participants were debriefed.

Results

Demographic information for each group is reported in Table 1. Participants' mean CFT scores did not vary across conditions, $F(3, 71) = .81$, $p = .50$, suggesting that groups did not vary in multi-digit arithmetic ability. Notably, the means were lower compared to the norming sample of 242 Canadian participants, where the mean was 31.60, SD = 13.53 (Sowinski et al., 2014); the means for the traditional 10,000 and reverse 7,000 conditions were significantly different than the norming group.

Analysis of Estimation Performance

Of the 3,024 trials, 29 (0.96%) trials were missing fixation data (i.e., the participant pressed the spacebar before fixating on the number line). These trials were not included in any analyses. Estimation accuracy for each trial was calculated as percentage of absolute error (PAE) using the formula: [|(Participant's estimate - Target location)| / Scale of estimate] x 100 (Petitto, 1990). For example, if a participant was presented with a target of 500 on the 7,000 number line, but their estimate corresponded to a location on the number line with a value of 1,000, the PAE score was 7.14% (i.e., |(1,000-500)| / 7,000 x 100). Three individuals had average PAE values that were more than two standard deviations from the average PAE across all participants. Furthermore, they had large standard deviations, suggesting a wide variation in their estimation accuracy. It is likely that when final estimates were made they moved their eyes before pressing the space bar. Thus, they were removed from PAE analyses. However, their first fixations were analyzed because these were unaffected by the final estimate movement. Overall, with the three individuals removed, mean final PAE across the four conditions was 6.95% (*SD* = 7.13). Mean PAE scores for each target number were used to examine accuracy patterns.

Post hoc comparisons of significant effects were done with Bonferroni corrections. Interactions shown in figures were interpreted using 95% inferential confidence intervals as recommended by Jarmasz and Hollands (2008).

(Insert Table 1 approximately here)

First Fixations

To understand strategy use, graphs of fixation patterns were made for first fixations. Similar to Sullivan et al. (2011), the number lines were divided into 20 regions, each representing 5% of the number line. The percentage of fixations made in each of the 20 regions for low and high numbers across the four conditions are shown in Figures 2a – d. Figures 2a and 2b display the pattern of first fixations for low (i.e., numbers below the midpoint) and high (i.e., numbers above the midpoint) target numbers for the 10,000 and 7,000 traditional conditions. Figures 2c and 2d display the pattern of first fixations for low and high target numbers for the 10,000 and 7,000 reverse conditions. Each of the quadrants shows the data separately for the traditional and reverse direction lines (separated by a vertical bar).

Consistent with our hypotheses**,** for traditional direction lines, first fixations were most frequent near the low endpoint for low numbers, with very few above the midpoint (see the left panels of Figures 2a and 2c). For high numbers, first fixations were more frequent around the midpoint than at either endpoint, although these distributions were less peaked at a single benchmark than those for low numbers (left panels of Figures 2b and 2d). These patterns suggest that participants used benchmarks in combination with left-to-right scanning. Reverse direction lines also showed evidence of the use of benchmarks and left-to-right scanning. On reverse lines for low numbers (right panel of Figures 2a and 2c), first fixations occurred at both the midpoint and low endpoint, whereas for high numbers (right panels of Figures 2b and 2d), first fixations occurred around the high (left) endpoint more (10,000) or almost (7,000) as frequently as around the midpoint. Furthermore, participants made some fixations around the left endpoint (i.e., the high endpoint) even on low target numbers, indicating that the left-to-right scanning strategy was persistent across number line directions.

To provide statistical tests of our hypotheses, the average number of first fixations made around the potential benchmarks were analyzed in a 2 (range: 10,000, 7,000) x 2 (direction: traditional, reverse) by 2 (target size: low, high) by 5 (benchmark: 0%, 25%, 50%, 75%, 100%) mixed ANOVA with repeated measures on the benchmark and target size variables. Five potential benchmarks were chosen to coincide with previous research on number line estimation (e.g., Ashcraft & Moore, 2012; Luwel et al., 2018; Sullivan et al., 2011). Target numbers below the midpoint were classified as "low" and target numbers above the midpoint were classified as "high". Fixations within 10% of the external benchmarks (i.e., 0%-10% and 90%-100%) and plus or minus 5% for the internal benchmarks (i.e., 20%-30%, 45%-55%, and 70%-80%) were used in this analysis.

The ANOVA results are presented in Table 2. There were significant main effects of target size, benchmark, and direction. There were significant two-way interactions for target size by direction, target size by benchmark, and benchmark by direction. There was also a significant three-way target size by benchmark by direction interaction. Main effects and interactions were qualified by the four-way interaction, $F(4, 272) = 9.44$, $p = .003$, $\eta^2 = .13$. As shown in Figures 3a – d, in general, individuals made a greater proportion of fixations around the low benchmarks for low target numbers and the high benchmarks for high target numbers[1], indicating that they were using features of the target number to direct their initial fixations. Moreover, for traditional direction number lines (Figures 3a and 3c), they made the greatest proportion of first fixations around the left (i.e., 0%) benchmark for low target numbers. This pattern supports our hypothesis that individuals would use a left-to-right scanning pattern when making estimates. Similarly, they made a greater proportion of fixations around the midpoint (i.e., 50%) for high than for low target numbers, consistent with a left-to-right strategy that was tuned to the size of the target number. Finally, there were relatively few fixations near the high endpoint for traditional number lines, supporting an upward adjustment (i.e., moving from smaller to larger numbers) even for high target numbers.

(Insert Table 2 approximately here)

For the reverse direction lines (see Figures 3b and 3d), the pattern of first fixations was different than that observed for the traditional direction lines and also varied as a function of range. In comparison to the traditional direction lines, the first fixations on the reverse direction lines did not have the steady pattern of decline from 0% to 50% for low target numbers or from 50% to 100% for high target numbers. The large proportion of first fixations around the high endpoint for high numbers on the 10,000 reverse line indicate that participants were often using a left-to-right scanning strategy. In contrast, for low numbers, first fixations were evenly distributed across the 50%, 75%, and 100% benchmarks. On the 7,000 reversed line, there was a similar pattern for both low and high target numbers, such that first fixations were more likely to correspond to the appropriate numerical 'half' of the line, with relatively fewer fixations around

[1] Even though the target number always appeared at the top left corner of the screen, participants did not always make their first fixations around the left endpoint. This finding suggests that participants' initial fixations were responsive to the size of the target number.

the 50% benchmark. Thus, participants' first fixations were less strongly tied to left-to-right scanning or benchmark location for the 7,000 reversed number lines.

(Insert Figures 3a – d approximately here)

In summary, on 10,000 traditional, 10,000 reverse, and 7,000 traditional direction number lines, there is evidence of a left-to-right scanning pattern starting at either the left-endpoint or the midpoint. In contrast, on the 7,000 reverse direction line, participants' first fixations suggested that their left-to-right scanning patterns were disrupted. Furthermore, for high target numbers, participants showed evidence of more frequent left-to-right scanning on the 10,000 line compared to the 7,000 line. Thus, the more familiar the number line, the more likely participants used a left-to-right scanning pattern similar to that seen in reading.

Analyses of Number of Fixations and Dwell Times

The overall mean number of fixations per trial was 4.85, and the overall mean dwell time was 2,238 ms. Mean fixations and mean dwell times on each trial were analyzed in separate 2 (range: 10,000, 7,000) by 2 (direction: traditional, reverse) ANOVAs. There were no significant effects, all ps > .22. Participants did not differ in their average number of fixations or average dwell time across conditions, suggesting that performance differences (see section on estimation accuracy) were not linked to differences in the mean number of fixations or dwell times.

Analyses of Number of Fixations and Dwell Times Across Benchmarks

To see if and how participants' eye movements were related to benchmarks after their first fixations, the mean number of fixations and mean dwell time per condition were analyzed for locations around five potential benchmarks: origin (0%), first quartile (25%), midpoint (50%), third quartile (75%), and endpoint (100%). All values reported are Greenhouse-Geisser corrected. There were significant effects of target size for both mean fixations, $F(1, 207.23)$ =

9.44, $p = .003$, $\eta^2 = .13$, and mean dwell times, $F(1, 176.93) = 5.62$, $p = .02$, $\eta^2 = .08$. There were also significant effects of benchmarks for both mean fixations, $F(4, 207.23) = 22.91$, $p < .001$, $\eta^2 = .26$, and mean dwell times, $F(4, 176.93) = 22.38$, $p < .001$, $\eta^2 = .25$. Overall, participants had the highest mean fixations and dwell times around the midpoint. Fixations and dwell times around the first and third quartile did not significantly differ. Participants made slightly more fixations around the origin than the endpoint, in line with the left-to-right scanning pattern seen in first fixations. The main effects were qualified by a target size by benchmark interaction for both mean fixations, $F(4, 207.23) = 108.10$, $p < .001$, $\eta^2 = .62$, and mean dwell times, $F(4, 176.93) = 88.74$, $p < .001$, $\eta^2 = .57$. As shown in Table 3, participants made more fixations and had longer dwell times around the low benchmarks for low target numbers and around the high benchmarks for high target numbers. There were no other significant effects for dwell times.

(Insert Table 3 approximately here)

For mean number of fixations, there was a significant interaction of target size, benchmark, and direction, $F(3.14, 207.23) = 2.63$, $p = .049$, $\eta^2 = .038$. As shown in Figure 4a, for low target numbers, participants in the traditional direction conditions made more fixations around the 0% benchmark and fewer fixations around the 75% benchmark than participants in the reverse direction conditions; otherwise the patterns were similar. These data suggest that participants were somewhat more likely to direct their attention to the left side of the number line in the reverse direction, even though low target numbers were not located there. For high target numbers, as shown in Figure 4b, participants in the traditional direction conditions made more fixations around the midpoint and fewer around the high endpoint than participants in the reversed conditions. Again, these results suggest that the participants focused more attention on the higher half of the line for both traditional and reversed lines, but were more likely to use the high (left) endpoint in the reversed condition. This pattern is similar to the pattern seen in first fixations (Figures 2b and 3b). Individuals in the reverse conditions may be directly targeting the benchmark closest to the target number and then adjusting their estimates.

(Insert Figures 4a – b approximately here)

Finally, the interaction of target size, range, and direction was significant for mean number of fixations, $F(1, 207.23) = 5.21$, $p = .026$, $\eta^2 = .073$. As shown in Figure 5, participants in the 10,000 traditional, 10,000 reverse, and 7,000 traditional did not differ in the average number of fixations they made for low and high target numbers. However, participants in the 7,000 reverse condition made significantly more fixations for high target numbers than low target numbers. Because the 7,000 reverse number line consists of both an atypical range and direction and high target numbers tend to require more effort to estimate, participants may have needed to make more fixations to settle on an accurate estimate.

(Insert Figure 5 approximately here)

In summary, mean number of fixations and dwell time patterns across conditions supported the hypothesis that participants relied on left-to-right scanning. Notably, these analyses did not provide additional information about strategy use beyond that available from the analysis of first fixations. Furthermore, the first fixation analyses provided more detail about how initial eye fixations were influenced by condition and benchmarks, providing useful strategy information. Thus, in the present study, first fixations were the most meaningful aspects of eye-tracking data with respect to understanding strategy choice and use of benchmarks.

Estimation Accuracy

Overall performance. Percentage Absolute Error (PAE) was analyzed in a 2 (target size: low, high) x 2 (range: 10,000, 7,000) x 2 (direction: traditional, reverse) mixed ANOVA with repeated measures on target size. Consistent with our hypotheses, participants were more accurate on 10,000 (M = 5.89%) than on 7,000 number lines (M = 8.47%), $F(1, 65) = 8.28$, p = .005, $\eta^2 = .11$. However, contrary to our hypotheses, they were *less* accurate on the traditional direction (M = 7.57%), than on the reversed direction number lines (M = 6.13%), $F(1, 65)$ = 3.90, $p = .05$, $\eta^2 = .06$. There was no significant effect of target size on PAE, however, target size interacted with range, $F(1, 65) = 8.18$, $p = .006$, $\eta^2 = .11$. As shown in Figure 6, performance on low target numbers did not vary with range whereas for high target numbers, participants in the 7,000 condition were significantly less accurate than participants in the 10,000 condition. Thus, consistent with our hypotheses, as the target numbers became larger, estimation error increased for the atypical range number line.

(Insert Figure 6 approximately here)

Error performance in relation to benchmarks. We examined accuracy around five potential benchmarks: origin (0%), first quartile (25%), midpoint (50%), third quartile (75%), and endpoint (100%). The PAE for the two target numbers closest to each of the five benchmarks were averaged for each participant. For the midpoint benchmark, the average was calculated based on the midpoint trial and the trials of the two target numbers closest to either side of the midpoint.

PAE was analyzed in a 2 (range: 10,000, 7,000) by 2 (direction: traditional, reverse) by 5 (benchmark: 0%, 25%, 50%, 75%, 100%) mixed ANOVA with repeated measures on the benchmark variable. As in the overall analysis, PAE varied with range, $F(1, 63) = 9.04$, $p = .004$, $\eta^2 = .13$, and direction, $F(1, 63) = 4.49$, $p = .04$, $\eta^2 = .07$. Performance also varied across benchmarks, $F(4, 252) = 16.85$, $p < .001$, $\eta^2 = .21$, with participants making more accurate estimates around the low and high endpoints than at the midpoint and quartiles (all *p*s < .001).

The main effect is qualified by the significant benchmark by range interaction, $F(4, 252) = 3.82$, $p = .01$, $\eta^2 = .06$. As shown in Figure 7, the 10,000-range number line shows a flattened M-shaped pattern with reduced error at the midpoint relative to the quartiles. In contrast, the 7000-range number line showed a tent-shaped pattern of performance, where midpoint performance was equivalent to that at the quartiles. These results support our hypotheses and suggest that participants did not calculate the midpoint correctly (see below for an analysis of error performance on the midpoint value). Consistent with the view that calculation difficulties contribute to the loss of a midpoint advantage, participants in the 10,000-range condition had lower error at the midpoint ($M = 3.64\%$) than participants in the 7,000-range condition ($M = 9.25\%$, $p < .001$). Error was also lower for the 10,000 condition at the first (25%) quartile.

(Insert Figure 7 approximately here)

Midpoint bias. To test the possibility that participants experienced a form of pseudoneglect (i.e., a leftward bias when locating the midpoint), we calculated the percentage of people who underestimated versus those who overestimated the location of the midpoint on the midpoint trial. Consistent with the pseudoneglect phenomenon observed in line bisection research, we found a leftward bias on the traditional direction number lines when individuals were asked to estimate the midpoint, with 69.2% of participants underestimating and only 30.8% of participants overestimating. In contrast, the pseudoneglect phenomenon was not found for the reverse direction number lines: 50.0% of participants underestimated and 50.0% of participants overestimated on the midpoint trial. Thus, reversing the number line may have reduced a leftward bias for estimating the midpoint that is found on canonical number lines that are oriented from left to right.

To determine whether the accuracy of midpoint location varied across conditions (independently of the direction of the error), PAE on the midpoint trial was analyzed in a 2 (direction: traditional, reverse) by 2 (condition: 10,000, 7,000) ANOVA. Consistent with our hypotheses, participants in the 7,000 range conditions were less accurate on the midpoint trial than participants in the 10,000 range conditions (9.05% vs. 3.04%), $F(1, 65) = 8.97$, $p = .004$, $\eta^2 = .12$, however, there was no effect of direction. This result suggests that individuals in the 7,000 range conditions may have miscalculated the numerical value of the midpoint and thus reduced the accuracy of locating targets across the number line. Thus, considering both first fixations (i.e., initial spatial processes and approximate locations) and final PAE (i.e., precise target locations), suggests that numerical processes limit accuracy on atypical number lines.

Model Fit for Benchmark Estimation Patterns

Following Luwel et al. (2018), we fitted a series of models to participants' final estimates. Power cycle models can be used if the number line estimation task is assumed to be a proportion judgment task (Barth & Paladino, 2011). When a participant makes an estimate using only the low and high endpoints as benchmarks, a one-cycle model fit is expected (Slusser et al., 2013). When a participant makes an estimate using the low and high endpoints and the midpoint, a two-cycle model fit is expected (Slusser et al., 2013). To compare model fits, we used the Akaike Information Criterion corrected for small samples (AICc). Smaller AICc values suggest better model fit. Delta AICc values are used to determine if there is support for one model in comparison to another. A delta AICc between 0 and 4 suggests that one model cannot be considered superior to another (Burnham & Anderson, 2002). Table 4 displays the percentage of individuals whose data were best fit with either the one-cycle or two-cycle model. Interestingly, for each condition, the majority of fits had a delta AICc of less than 4 between the one- and two-cycle fits, meaning a superior model could not be determined. Across the four conditions, the one-cycle model fit a higher percentage of participants than the two-cycle model. These results suggest that the majority of individuals were using only the low and high endpoints as benchmarks to make their estimates. However, behavioural data (i.e., first fixations) and PAE suggest that individuals in all four conditions use the midpoint to guide estimates, especially for high target numbers. The statistical models do not support the behavioural data. Ashcraft and Moore (2012) also found that model fits did not adequately reflect their analysis of PAE data. Thus, model fits dot not appear to accurately reflect strategy use.

(Insert Table 4 approximately here)

Discussion

In the present study, we examined how atypical range and reverse direction influenced strategy choice and number line performance. Participants made estimates on one of four number lines as shown in Figure 1. Fixations on the number line were recorded using eye-tracking equipment and analyzed for evidence that participants used benchmarks. Specifically, fixations made within 5% of each potential benchmark location (i.e., 0%, 25%, 50%, 75%, and 100%) were considered as evidence for use of benchmarks. In previous studies, analyses of the location of final estimates and model fits were used to infer strategy use (e.g., Ashcraft & Moore, 2012; Barth & Paladino, 2011; Luwel et al., 2018). However, analyses of final estimates do not provide information about the process of estimation or initial strategy selection. In the present study, we used eye-tracking data to better understand the process of estimation, and more specifically, how and when individuals use benchmarks to make estimates.

What Can Eye-Tracking Tell Us About Number Line Strategies?

What strategies are individuals using when estimating on a number line? For both typical and atypical range and direction number lines, participants relied on benchmarks and target size to help guide their estimates. We found evidence of a left-to-right scanning strategy, with participants making a greater proportion of fixations on the left than on the right endpoint for both the traditional and reverse direction lines. We also found evidence for use of benchmark strategies: For low number trials, participants frequently made a first fixation near the origin benchmark (i.e., 0%) whereas, for the high number trials, participants more frequently made first fixations at the midpoint (i.e., 50%) or high endpoint (i.e., 100%) benchmarks. The analyses of first fixations provided clear support for these two strategies.

Do participants' strategies vary across number line conditions? We hypothesized that the use of internal benchmarks, specifically the midpoint, would be most evident for 10,000 range number lines because the numerical value of the benchmark is easier to retrieve or calculate (i.e., 5,000) than for the 7,000 range (i.e., 3500). In contrast, we found that participants used the midpoint for both 10,000 and 7,000 numbers. The finding that the midpoint was used for both traditional and atypical range lines was contrary to previous literature on atypical number lines (e.g., Barth & Paladino, 2003; Luwel et al. 2018) where PAE analyses and model fits have suggested that for atypical range number lines, individuals do not use internal benchmarks. In fact, eye-tracking data in the present study suggests that participants routinely use benchmark locations to guide their estimates, as shown by the similar patterns for 10,000 and 7,000 number lines (Figure 2). Difficulties occurred, however, because of inaccurate numerical processing – for example, inaccurate calculation of the midpoint, which resulted in less accurate estimates. Thus, the present research suggests that people use benchmark strategies on both traditional and atypical range number lines, but poor numerical processing reduces the efficiency of these strategies on the latter.

We examined the influence of spatial processing on strategy use by manipulating the direction of the number line. For all four conditions we found evidence of left-to-right scanning patterns. This pattern was evident for low target numbers on traditional number lines (e.g., Figures 2a and 2c) and for high target numbers on reverse number lines (e.g., Figures 2b and 2d). However, there were higher peaks around the left endpoint for low numbers on traditional number lines than high numbers on reverse number lines. This pattern suggests that left-to-right scanning starting at the endpoint may be preferred when numbers increase from left-to-right (i.e., 0 to 10,000/7,000) than when they decrease from left-to-right (i.e., 10,000/7,000 to 0). Using left-to-right scanning would require participants to adjust downwards to the target, whereas adjusting upwards appears to be preferred (see also Sullivan et al., 2011).

In general, the patterns of first fixation were similar for low target numbers in both the traditional and reverse direction conditions. In contrast, for the high target numbers, different patterns emerged across traditional and reverse direction lines. For the 10,000 and 7,000 range, participants who estimated on traditional direction lines used the midpoint for their first fixation (Figures 2b and 2d). For the 10,000 reverse direction, participants used the left endpoint (i.e., 100%). Again, these patterns indicate that participants use left-to-right scanning. Because 10,000 is a familiar range, participants may have decided that the left-to-right scanning was the most efficient strategy for the reverse number line, even though it required a downward adjustment. In contrast, for the 7,000 reverse number line, adjusting downward is more difficult as the proportions of the number line are not as easy to identify (e.g., 2,500 vs. 1,750 for the 25% benchmark), thus participants were more likely to make their first fixation around the midpoint and move upwards. We also saw a more uniform pattern of first fixations for high target numbers on the 7,000 reverse line. This finding probably reflects the difficulty of estimating on a number line that is both atypical in range and direction resulting in greater variability in participants' strategy choices.

The first fixations were the best indicators of strategy choice in this study. Information provided by overall fixation and dwell time patterns did not provide much additional insight into strategy use. Participants did not make more fixations nor did they spend more time fixating on the number line or around the benchmarks as they interacted with the atypical number lines. After first fixations, participants made similar use of benchmarks on all of the number lines. Adults are quite accurate when estimating on a number line. They make a small number of fixations on each trial and choose a final estimate quickly. Thus, they probably select a strategy immediately after seeing the target number, as evidenced by the meaningful first fixations made on the number line that corresponded to target size. Given that strategies are used to guide behaviours, we would expect participants to select a strategy well before final estimates are made, and thus first fixations should be most reflective of strategy choice.

Unexpected Effects on Atypical Number Lines

The direction of the number line influenced overall performance in an unexpected way. Contrary to our hypotheses, participants who estimated on reversed number lines were more accurate than those who estimated on traditional direction lines. One possible explanation for these results is that when the number line is reversed, participants are less likely to use left-to-right visual scanning (e.g., Behrmann, Watt, Black, & Barton, 1997; Foulsham, Gray, Nasiopoulos, & Kingstone, 2013), instead using target information to direct their fixations more precisely. We also identified a spatial bias such that people in the traditional conditions were more likely to place the midpoint to the left of the centre of the line. This spatial bias has been previously identified in line bisection tasks but not in number line performance. Consistent with a leftward spatial bias, when participants were asked to estimate the midpoint (i.e., 3,500 for the 7,000 range conditions and 5,000 for the 10,000 range conditions) they were more likely to underestimate on the traditional direction number lines, whereas for the reverse conditions they were equally likely to over- or underestimate. In research on line bisection, the leftward bias is related to scanning patterns (left-to-right), consistent with the general assumption that number magnitude is associated with the left side of space in cultures with left-to-right reading directions (Cipora et al., 2019). Thus, reversing the number line created a novel situation which reduced the bias resulting from years of left-to-right reading and left-to-right numerical magnitude orientation.

The present research indicates that both range and direction influence number line estimation strategy and performance but are tied to different cognitive processes. First, number knowledge is linked to the effects of varying the range of the number line (e.g., Barth & Paladino, 2011): Participants need to understand the relation between the symbolic representation of the number and the proportion that number represents on a number line in order to make their estimates. In contrast, the direction of the number line may be more strongly linked to spatial skills (e.g., Cornu, Hornung, Schiltz, & Martin, 2017; LeFevre et al., 2013). The unfamiliar direction may have forced participants to more carefully consider the link between number magnitude and spatial location. Accordingly, even though participants in the 7,000-range conditions were significantly less accurate than participants in the 10,000-range conditions, for both range conditions participants who estimated on reversed number lines performed better than those who estimated on traditional lines. This pattern suggests that spatial skills and numerical knowledge are separately influential in number line performance. Nevertheless, the effects of range were larger than the effects of direction. Thus, although both number knowledge and spatial processes are required to make estimates on a number line, numerical magnitude (associated with range) plays a larger role in strategy selection and estimation accuracy than spatial processing (associated with direction).

Cyclical Power Models as Evidence for Strategies

Finally, we examined whether performance on the number lines were best fit by a one-cycle or two-cycle model in order to connect with previous research. Ashcraft and Moore (2012) used the beta values of *n*-cycle power functions to examine model fit and found that these models were inconsistent with the behavioural data. Similarly, in the present research, the inferences about performance based on model fit were inconsistent with eye-tracking data. Given that model fit analyses rely on final estimates only, we argue that model fits should not be used to infer strategy use. A participant must select a strategy before their final estimate is made. Thus, although model fits provide evidence of over- and under-estimation patterns in number line estimation, they do not directly reflect benchmark-based strategy use. Accordingly, it is critical to explore strategy use using multiple measures to understand *how* people approach number line estimation.

Conclusions and Future Directions

The present research makes several important contributions to the literature on number line processing. First, our novel use of eye tracking on atypical number line estimation corroborates previous findings based on accuracy, providing information about why participants are less accurate when estimating on atypical-range number lines. Specifically, this pattern is probably a result of participants' miscalculating the numerical value of the midpoint (Link et al., 2014; Luwel et al., 2018). Second, we found that reversing the number line reduced the left-to-right processing bias found on traditional direction lines that is presumably related to routine left-to-right visual scanning in situations with numbers (Behrmann et al., 1997; Foulsham et al., 2013). Third, we showed that eye-tracking data is important for understanding benchmark-based strategy use from the start of the trial because first fixations are a better index of strategy use than measures that only reflect final estimates. Accordingly, we found evidence that participants used benchmark-based strategies on all number lines. Poorer performance occurred when participants translated numbers incorrectly (e.g., estimating the midpoint of the 0-7,000 line as 3,000 rather than 3,500). Furthermore, when the number line direction was reversed, the leftward bias present for canonical number lines was minimized, especially with the 7,000 reverse condition where individuals showed no preference to fixate first on left or right endpoints (see Figures 3c and 3d). In sum, these findings provide important corroboration and new information about strategy processes in number line tasks.

Despite these important contributions, further research is needed on strategy use in number line tasks. For example, if participants reported their strategy use for each trial, comparisons between eye-tracking data and verbal reports of strategy use could enhance the interpretation of both measures and provide further insight into the overall process of number line estimation. Moreover, use of a within-subjects design would allow researchers to observe how participants adjust their strategies when presented with both traditional and atypical number lines. Including ‘easy’ numbers such as 1,000 or 7,500 (for a 10,000 number line) would also help to identify error that is the result of rounding versus application of benchmark strategies.

In concluding, let us return to our parking scenario, with one small change. Imagine you are driving and preparing to park in a country where you must drive on the opposite side of the road. You prepare to parallel park and ask yourself, “Can I fit?” This situation is atypical for you, so you are forced to adjust your strategies, potentially pulling in and out a few extra times, or more carefully choosing the starting point of your trajectory. This example illustrates the importance of understanding how individuals develop and adjust their estimation strategies under atypical conditions. Accordingly, the findings of the present research provide new evidence about the effects of number range and spatial orientation on benchmark strategy use in number line estimation.

Funding: This project was funded by Natural Sciences and Engineering Research Council of Canada through a Discovery Grant to Jo-Anne LeFevre.

Acknowledgements: The authors thank Michelle Hosang-Roy for assisting with data collection and Drs. Chang Xu and Rebecca Merkley for their feedback on previous versions of the manuscript.

Declaration of Conflicting Interests: The authors declare that there is no conflict of interest.

References

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology, 111*, 246-267. doi:10.1016/j.jecp.2011.08.005

Baayen, R. H. (2008). Analyzing linguistic data: A practical introduction to statistics. Cambridge: Cambridge University Press.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14*(1), 125-135. doi:10.1111/j.1467-7687.2010.00962.x

Behrmann, M., Watt, S., Black, S. E., & Barton, J. J. S. (1997). Impaired visual search in patients with unilateral neglect: An oculographic analysis. *Neuropsychologia, 35*(11), 1445-1458. doi: 10.1016/S0028-3932(97)00058-4

Bisanz, J. & LeFevre, J. (1990). Strategic and nonstrategic processing in the development of mathematical cognition. In D. F. Bjorkland (Ed.), *Children's strategies: Contemporary views of cognitive development* (pp. 213-244). Hillsdale, New Jersey, US: Lawrence Erlbaum Associates, Inc.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*(1), 189-201. doi:10.1037/0012-1649.42.1.189

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*(4), 1016-1031. doi:10.1111/j.1467-8624.2008.01173.x

Booth, J.K., & Newton, K.L. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology, 31,* 247-253. doi:10.1016/j.cedpsych.2012.07.001

Bowers, D., & Heilman, K. M. (1980). Pseudoneglect: Effects of hemispace on a tactile line bisection task. *Neuropsychologica, 18*, 491-498. doi:10.1016/0028-3932(80)90151-7

Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel interference: A practical information-theoretic approach (2nd Ed.). New York: Springer.

Cipora, K., Schroeder, P. A., Soltanlou, M., & Nuerk, H.-C. (2019). More space, better math: Is space a powerful tool or a cornerstone for understanding arithmetic? In K. S. Mix & M. T. Battista (Eds.), *Visualizing Mathematics: The Role of Spatial Reasoning in Mathematical Thought.*Springer Nature.

Cornu, V., Hornung, C., Schiltz, C., & Martin, R. (2017). How do different aspects of spatial skills relate to early arithmetic and number line estimation? *Journal of Numerical Cognition*. doi: 10.5964/jnc.v3i2.36

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences, 7*(4), 145-147. doi:10.1016/S1364-6613(03)00055-X

Ebersbach, M. Evidence for a spatial-numerical association in kindergartners using a number line task. *Journal of Cognition and Development, 16*(1), 118-128. doi: 10.1080/15248372.2013.805134

Foulsham, T., Gray, A., Nasiopoulos, E., & Kingstone, A. (2013). Leftward biases in picture scanning and line bisection: A gaze-contingent window study. *Vision Research, 78,* 14-25. doi: 10.1016/j.visres.2012.12.001

Grant, E. R. & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14*(5), 462-466. doi: 10.1111/1467-9280.02454

Heine, A., Thaler, V., Tamm, S., Hawelka, S., Schneider, M., Torbeyns, J., . . . Jacobs, A. M. (2010). What the eyes already "know": Using eye movement measurement to tap into children's implicit numerical magnitude representations. *Infant and Child Development, 19*,175-186. doi:10.1002/icd.640

Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review, 107*(3), 500-524. doi: 10.1037/0033-295X.107.3.500

Huber, S., Moeller, K., & Nuerk, H. (2014). Dissociating number line estimations from underlying numerical representations. *The Quarterly Journal of Experimental Psychology, 67*(5), 991-1003. doi:10.1080/17470218.2013.838974

Hurst, M., Leigh Monahan, K., Heller, E., & Cordes, S. (2014). 123s and ABCs: Developmental shifts in logarithmic-to-linear responding reflect fluency with sequence values. *Developmental Science, 17*, 892-904. doi:10.1111/desc.12165

Jarmasz, J. & Hollands, J. G. (2009). Confidence intervals in repeated-measures designs: The number of observations principle. *Canadian Journal of Experimental Psychology, 63*(2), 124-138. doi: 10.1037/a0014164

Kirk, E. P. & Ashcraft, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27,* 157-175. doi: 10.1037/0278-7393.27.1.157

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive Science, 37*, 775-799. doi:10.1111/cogs.12028

LeFevre, J., Jiménez Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, J.L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology, 4*, 1-9. doi:10.3389/fpsyg.2013.00641

LeFevre, J., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 216-230. doi: 10.1037/0278-7393.22.1.216

Levander, M., Tegner, R., & Caneman, G. (1993). Tactile line bisection in normal subjects. *Perceptual and Motor Skills, 76*, 831-836. doi:10.2466/pms.1993.76.3.831

Levine, D. R. (1982). Strategy use and estimation ability of college students. *Journal for Research in Mathematics Education, 13*(5), 350-359.

Link, T., Nuerk, H.C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *The Quarterly Journal of Experimental Psychology, 67*(8), 1597-1613. doi:10.1080/17470218.2014.892517

Luwel, K., Peeters, D., Dierckx, G., Sekeris, E., & Verschaffel, L. (2018). Benchmark-based strategy use in atypical number lines. *Canadian Journal of Experimental Psychology.* Advance online publication. doi:10.1037/cep0000153

Mock, J., Huber, S., Klein, E., & Moeller, K. (2016). Insights into numerical cognition: considering eye-fixations in number processing and arithmetic. *Psychological Research, 80*(3), 334-359. doi:10.1007/s00426-015-0739-9

Newman, R.S., & Berger, C.F. (1984). Children's numerical estimation: flexibility in the use of counting. *Journal of Education Psychology, 76*, 55-64. doi:10.1037/0022-0663.76.1.55

Peeters, D., Degrande, T., Ebersbach, M., Verschaffel, L., & Luwel, K. (2016). Children's use of number line estimation strategies. *European Journal of Psychology of Education, 31*(2), 117-134. doi:10.1007/s10212-015-0251-z

Peeters, D., Sekeris, E., Verschaffel, L., & Luwel, K. (2017). Evaluating the effect of labeled benchmarks on children's number line estimation performance and strategy use. *Frontiers in psychology, 8*, 1082. doi: 10.3389/fpsyg.2017.01082

Peeters, D., Verschaffel, L., & Luwel, K. (2017). Benchmark-based strategies in whole number line estimation. *British Journal of Psychology, 108*(4), 668-686. doi:10.1111/bjop.12233

Petitto, A. L., (1990). Development of number line and measurement concepts. *Cognition and Instruction, 7*, 55-78. doi:10.1207/s1532690xci0701_3

Reinert, R. M., Huber, S., Nuerk, H.-C., & Moeller, K. (2015). Strategies in unbounded number line estimation? Evidence from eye-tracking. *Cognitive Processing, 16,* 359-363. doi:10.1007/s10339-015-0675-z

Reys, R., & Bestgen, B. (1981). Teaching and assessing computational estimation skills. *Elementary School Journal, 82*(2), 116-127.

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science, 17*(4), 525-536. doi:10.1111/desc.12166

Sagiv, N., Simner, J., Collins, J. Butterworth, B., & Ward, J. (2006). What is the relationship between synaesthesia and visuo-spatial number forms. *Cognition, 101,* 114-128. doi:10.1016/j.cognition.2005.09.004

Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smet, B., Verschaffel, L., Jacobs, A., & Stern, E. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development, 23*, 424-437. doi:10.1016/j.cogdev.2008.07.002

Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: a meta-analysis. *Child Development, 89*(5), 1467-1484. doi: 10.1111/cdev.13068

Siegler, R. S. (1989). Hazards of mental chronometry: An example from children's subtraction. *Journal of Educational Psychology, 81*(4), 497-506. doi: 10.1037%2F0022-0663.81.4.497

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*(2), 428-444. doi:10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*(3), 237-250. doi:10.1111/1467-9280.02438

Siegler, R. S., & Thompson, C. A. (2014). Numerical landmarks are useful – except when they're not. *Journal of Experimental Child Psychology, 120,* 39-58. doi:10.1016/j.jecp.2013.11.014

Simms, V., Clayton, S., Cragg, L., Gilmore, C., & Johnson, S. (2016). Explaining the relationship between number line estimation and mathematical achievement: The role of visuomotor integration and visuospatial skills. *Journal of Experimental Child Psychology, 145,* 22-33. doi:10.1016/j.jecp.2015.12.004

Slusser, E., & Barth, H. (2017). Intuitive proportion judgment in number-line estimation: Converging evidence from multiple tasks. *Journal of Experimental Psychology, 162*, 181-198. doi:10.1016/j.jecp.2017.04.010

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

Sowinski, C., Dunbar, K., & LeFevre, J. (2014). *Calculation Fluency Test,* unpublished technical report, Math Lab, Carleton University, Ottawa, Canada.

Smith-Chant, B. & LeFevre, J. (2003). Doing as they are told and telling it like it is: Self-reports in mental arithmetic, *Memory and Cognition, 31*(4), 516-528. doi: 10.3758/BF03196093

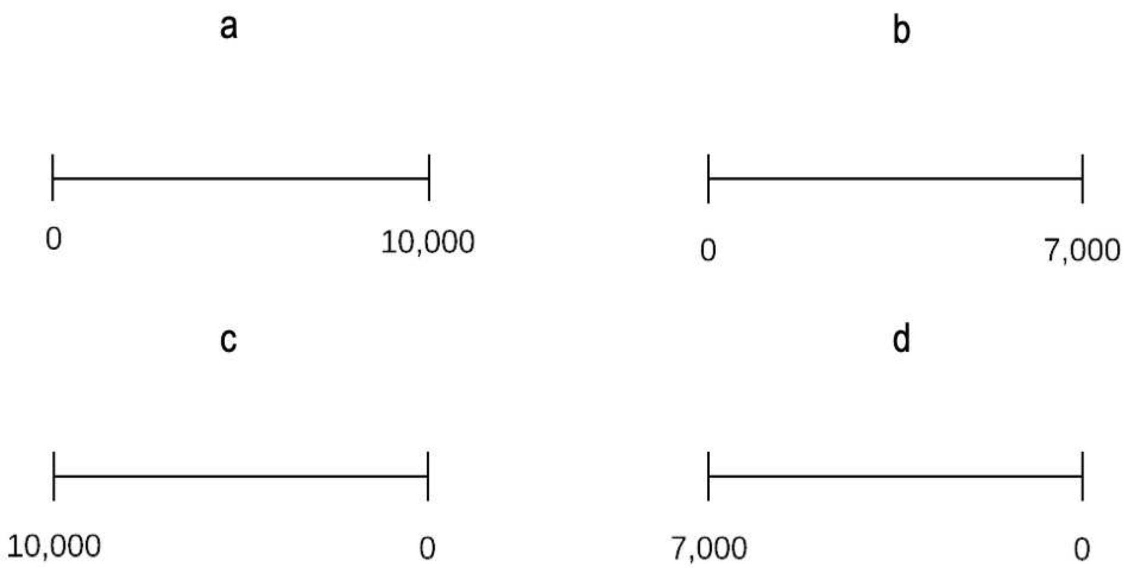
Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 683-692. doi:10.1037/0096-1523.16.4.683

Sullivan, J. L., Juhasz, B. J., Slattery, T. J., & Barth, H. C. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review, 18*(3), 557-563. doi:10.3758/s13423-011-0081-1

Tsao, Y. L. & Pan, T. R. (2010). The study of the computational estimation performance and computational estimation strategy. *New Wave-Educational Research & Development, 13*(1), 12-42.

Viarouge, A., Hubbard, E. M., & McCandliss, B. D. (2014). The cognitive mechanisms of the SNARC effect: An individual differences approach. *PLoS One, 9*(4), e95756. doi:10.1371/journal.pone.0095756

Xu, C., & LeFevre, J.-A. (2016). Training young children on sequential relations among numbers and spatial decomposition: Differential transfer to number line and mental transformation tasks. *Developmental Psychology, 52*(6), 854-866. doi:10.1037/dev0000124

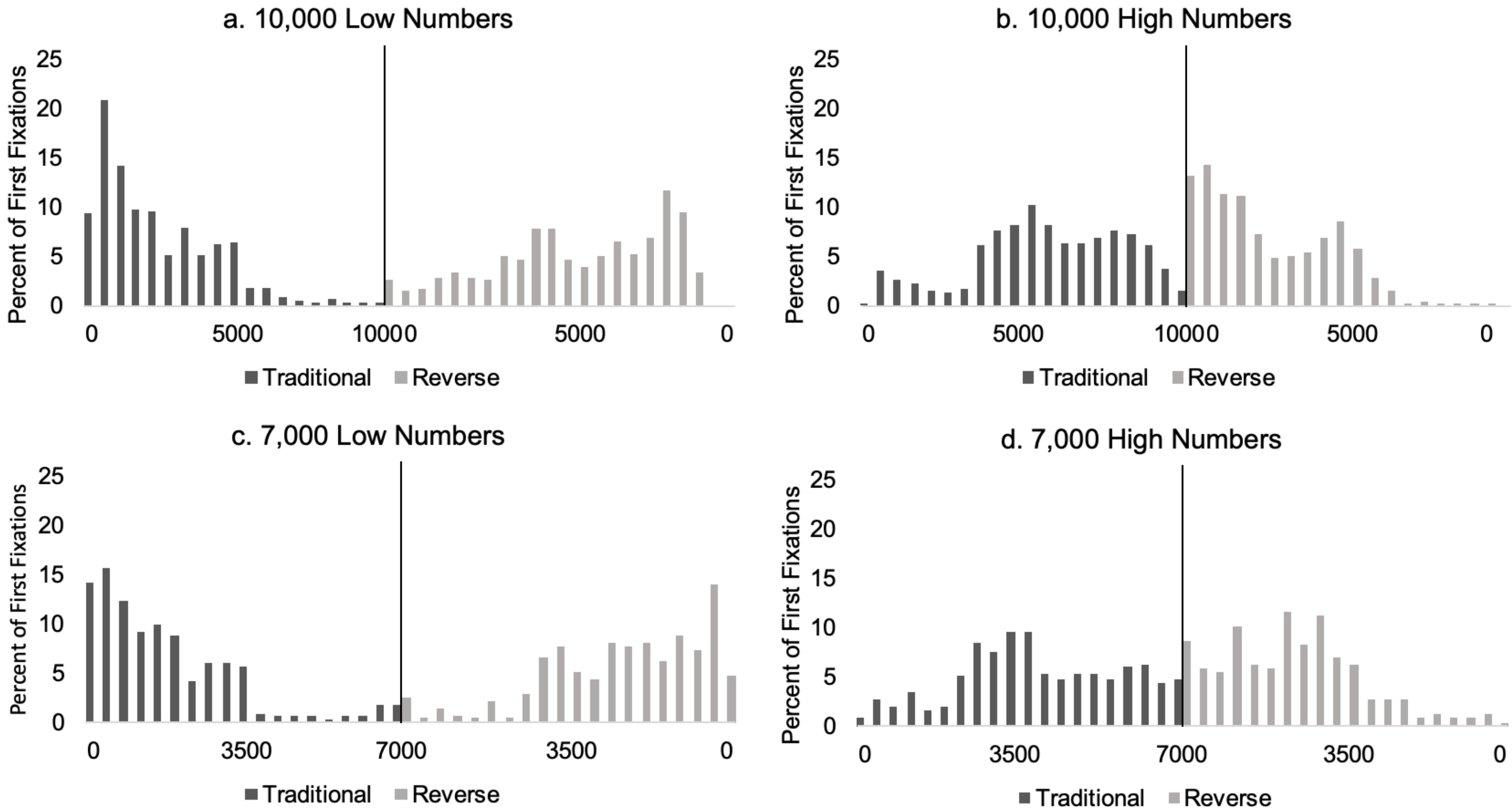


Figures 1a – d. The four number line conditions: a) 10,000 traditional, b) 7,000 traditional, c) 10,000 reverse, d) 7,000 reverse.

Table 1

Demographic Information for Each Number Line Condition

Condition	n(females)	Mean age (years)	Calculation Fluency Test	
			M	*SD*
0 to 10,000	20(12)	19.75	23.40	14.84
0 to 7,000	20(11)	19.50	26.75	14.47
10,000 to 0	16(8)	20.13	26.06	16.02
7,000 to 0	16(7)	19.31	20.00	9.99



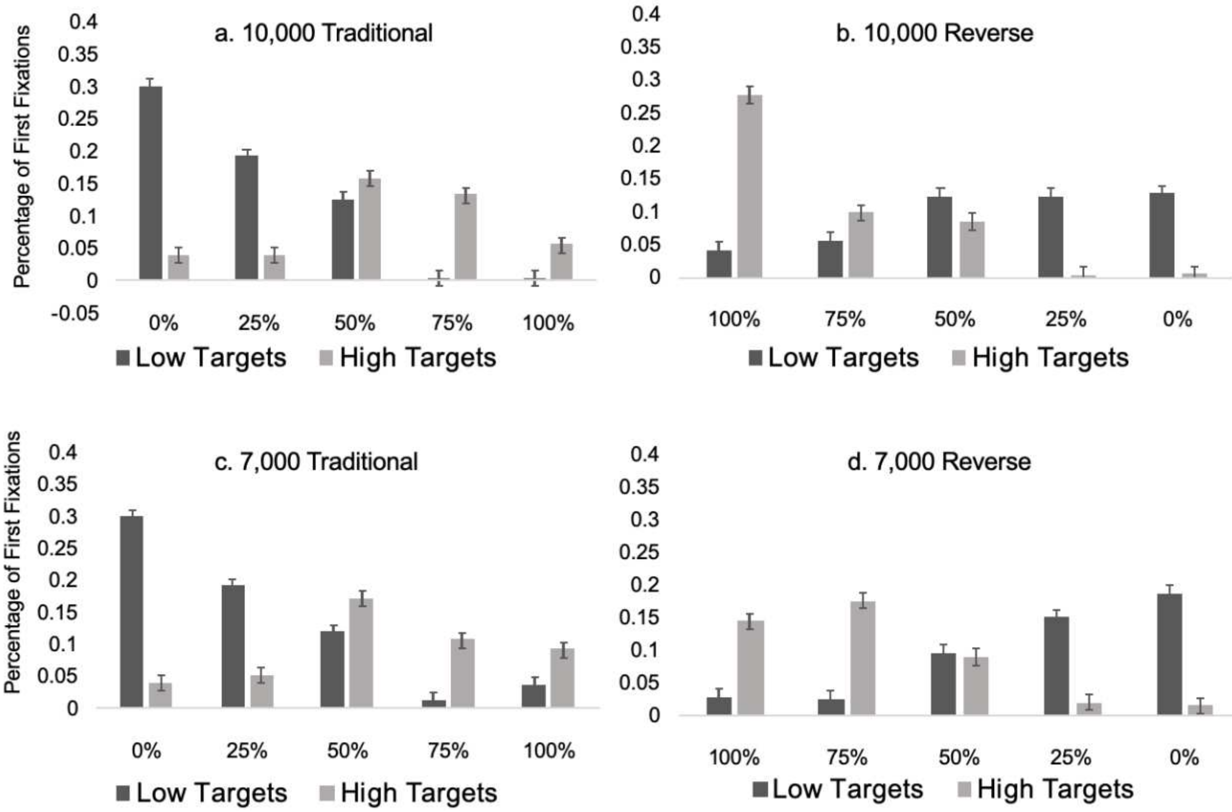
Figures 2 a – d. Percentage of first fixations observed in each interval for the 10,000 and 7,000 ranges as a function of direction: a) Traditional direction low numbers, b) Traditional direction high numbers, c) Reverse direction low numbers, and d) Reverse direction high numbers.

Table 2

ANOVA summary table for first fixations as a function of target size, benchmarks, range, and direction

Source	*df*	MS	*F*	*p*	Effect Size
Target Size	1	0.089	34.07	<.001*	.33
Benchmark	4	0.070	5.36	<.001***	.07
Range	1	0.001	0.30	.59	.004
Direction	1	0.037	7.56	.01**	.10
Target Size x Benchmark	4	0.726	109.37	<.001***	.62
Target Size x Range	1	<0.001	0.14	.71	.002
Target Size x Direction	1	0.055	20.90	<.001***	.24
Benchmark x Range	4	0.008	0.59	.67	.009
Benchmark x Direction	4	0.148	11.41	<.001***	.14
Range x Direction	1	0.002	0.44	.51	.006
Target Size x Benchmark x Range	4	0.013	2.01	.10	.03
Target Size x Benchmark x Direction	4	0.059	8.86	<.001***	.12
Target Size x Range x Direction	1	0.001	0.40	.53	.006
Benchmark x Range x Direction	4	0.031	2.38	.05	.034
Target Size x Benchmark x Range x Direction	4	0.021	3.24	.02*	.045
Total	272				

Note: *p < .05, **p < .01, ****p < .001

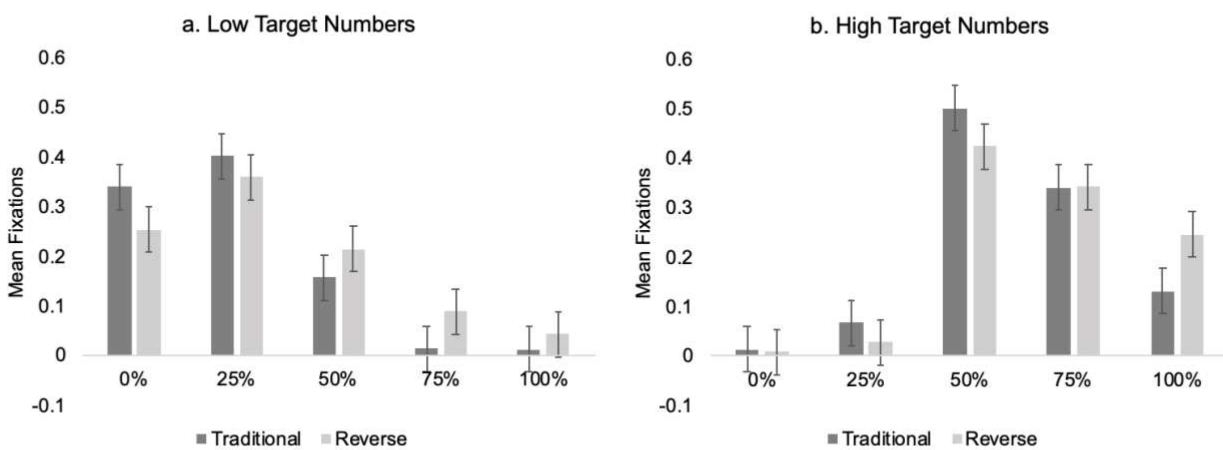


Figures 3 a – d. Percentage of first fixations observed around each benchmark (0%, 25%, 50%, 75%, 100%) for low and high target numbers as a function of range and direction: a) 10,000 traditional direction, b) 10,000 reverse direction, c) 7,000 traditional direction, and d) 10,000 reverse direction. Error bars represent 95% confidence intervals.

Table 3

Mean Fixations and Dwell Times Across Benchmarks for Low and High Target Numbers

Benchmarks	Mean Fixations		Mean Dwell Times	
	Low Targets	High Targets	Low Targets	High Targets
0%	.30	.01	156.76	2.72
25%	.38	.05	190.86	16.23
50%	.19	.46	90.06	227.92
75%	.05	.34	22.83	173.98
100%	.03	.19	6.46	96.06



Figures 4 a – b. Mean number of fixations for a) low target numbers and b) high target numbers for traditional and reverse direction number lines. Error bars represent 95% confidence intervals.

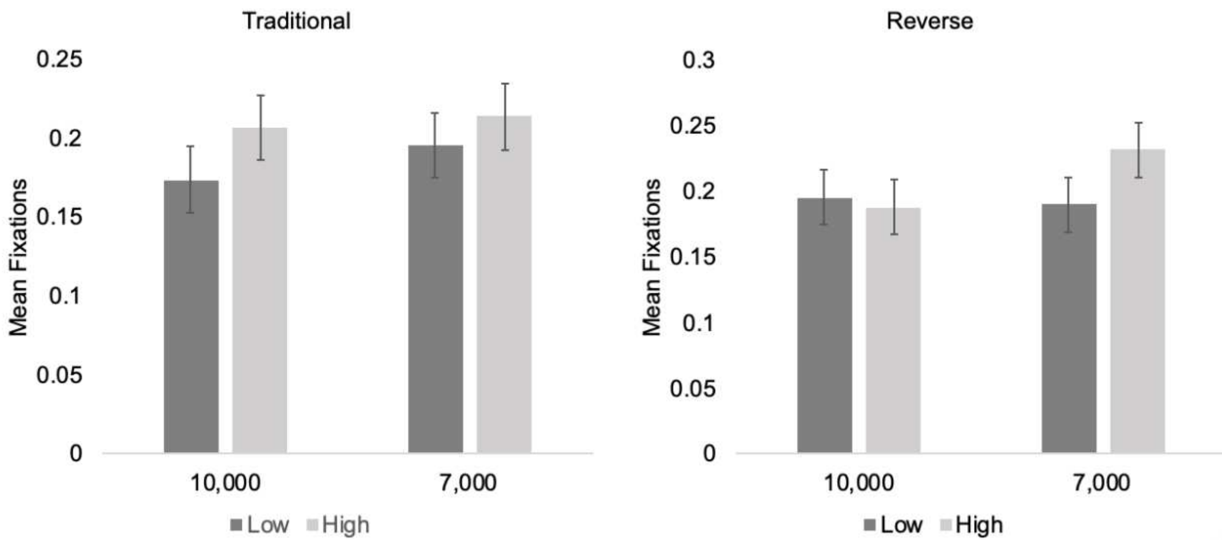


Figure 5. Mean number of fixations for traditional and reverse direction number lines for low and high target numbers. Error bars represent 95% confidence intervals.

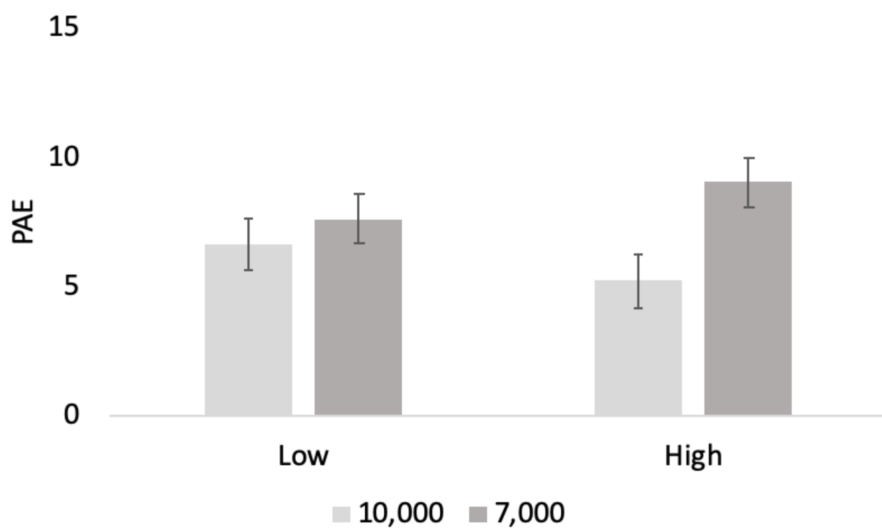


Figure 6. Mean percentage of error (PAE) for low and high target numbers for 10,000 and 7,000 range number lines. Error bars represent 95% confidence intervals.

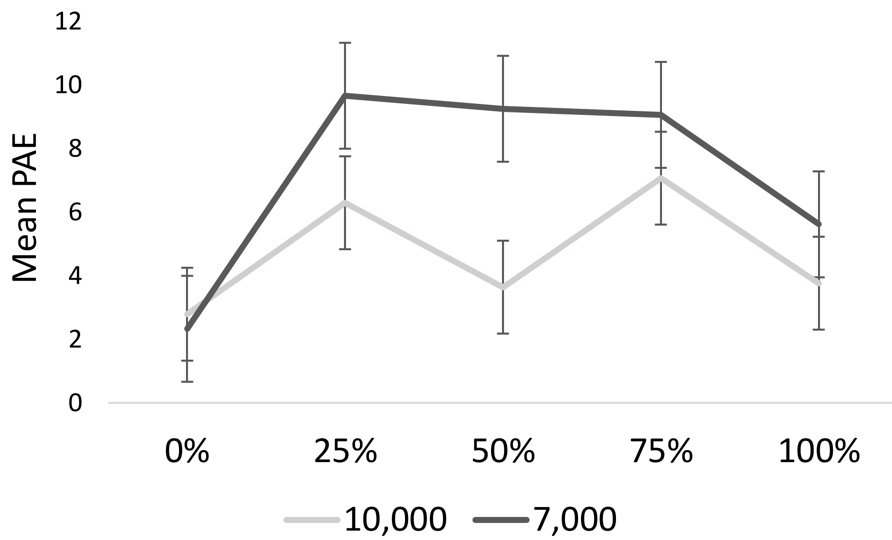


Figure 7. Mean percentage of error (PAE) at the five benchmarks for 10,000 and 7,000 range number lines. Error bars represent 95% confidence intervals.

Table 4

Percentage of participants in each condition whose data are best fit with one- and two-cycle model fits

Number Line	One-Cycle	Two-Cycle	Neither*
0 to 10,000	20.0	30.0	50.0
0 to 7,000	10.0	40.0	50.0
10,000 to 0	7.0	36.0	57.0
7,000 to 0	12.5	37.5	43.8

*The delta AIC values did not differ by more than 4