Article:

Di Lonardo Burr, S. orcid.org/0000-0001-6338-9621 and LeFevre, J.-A. (2021) Fixated in more familiar territory: providing an explicit midpoint for typical and atypical number lines. Quarterly Journal of Experimental Psychology, 74 (3). pp. 523-535. ISSN 1747-0218

https://doi.org/10.1177/1747021820967618

Fixated in More Familiar Territory:

Providing an Explicit Midpoint for Typical and Atypical Number Lines

Sabrina M. Di Lonardo[1]

Jo-Anne LeFevre[1,2]

[1]Carleton University, Institute of Cognitive Science

[2]Carleton University, Department of Psychology

This manuscript was accepted for publication in the *Quarterly Journal of Experimental Psychology* on September 13, 2020. 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

Does providing an explicit midpoint affect adults' performance differently for typical and atypical number line tasks? Participants ($N = 29$) estimated the location of target numbers on typical (i.e., 0 – 10,000) and atypical (i.e., 0 – 7,000) number lines with either an explicitly labelled midpoint or no midpoint. For the typical number line, estimation accuracy did not differ for the explicit- and implicit-midpoint conditions. For the atypical number line, participants in the explicit-midpoint condition were more accurate than those in the implicit-midpoint condition and their pattern of error was similar to that seen for typical number lines (i.e., M-shaped). In contrast, for participants in the implicit-midpoint condition, the pattern of error on the atypical line was tent-shaped, with less accurate estimates around the midpoint and quartiles than the endpoints. Eye-tracking data showed that, for all number lines, participants used the middle of the line to guide their estimates, but participants in the explicit-midpoint condition were more likely to make their first fixation around the true midpoint than those in the implicit-midpoint condition. We conclude that adults have difficulty in estimating on atypical number lines because they incorrectly calculate the numerical value of the midpoint.

Abstract Word count: 193 words

Keywords: number line, estimation, eye-tracking, mathematical cognition, midpoint, benchmark strategy

Fixated in More Familiar Territory: Providing an Explicit Midpoint for Typical and Atypical Number Lines

Adults rely on estimation in their everyday lives. For example, people may estimate the total cost of groceries to determine if they have sufficient funds in their bank account or estimate the area of a room to figure out how much paint to buy. These estimates rely on their ability to make approximate calculations and to compare the results with known benchmarks. Everyday estimation skills are developed when students are taught mathematics in school. Number line tasks are often used to teach and to measure estimation skills. Consistent with the view that number line tasks reflect relevant skills, performance on number line tasks are strongly related to current and future math achievement (Schneider et al., 2018). Much of the research on number line estimation in adults has focused on typically bounded number lines that have a left endpoint of “0” and a right endpoint that is a power of 10 (see Figures 1a and 1b for examples; Ashcraft & Moore, 2012; Barth & Paladino, 2011; Siegler & Opfer, 2003). However, *atypical* number lines are more challenging, and produce variability in performance even for older children and adults. Atypical number lines may have an endpoint that is not a power of 10 (e.g., 0-7,000; Di Lonardo et al., 2020) or start at a value other than 0 (e.g., 367-1,367; Luwel et al., 2018). See Figures 1c and 1d for examples. The goal of the present research was to better understand adults’ strategy use on typical and atypical number line tasks. Specifically, we varied the presence of a labelled midpoint to determine whether providing this benchmark would improve adults’ number line performance.

(Insert Figure 1 approximately here)

The strategies used on typical number line tasks vary with age and with number line characteristics, but usually involve the use of explicit benchmarks such as the endpoints or implicit benchmarks such as the midpoint. On typical number lines, young children (i.e., 4-7 years old) use counting strategies that start from either the origin or the upper endpoint (Petitto, 1990; Xu & LeFevre, 2016). With training, children in this age range make use of an explicit midpoint (Xu & LeFevre, 2016). On a 0-1000 number line, older children (i.e., 9-10 years old) use some benchmark-based strategies without training, such as relying on both endpoints, but only use the midpoint when it is explicitly provided (Peeters, Sekeris, et al., 2017). Adults' benchmark-based strategies include the use of the implicit midpoint on typical number lines (e.g., Ashcraft & Moore, 2012; Luwel et al., 2018 Di Lonardo et al., 2020; Sullivan et al., 2011). For atypical number lines, Luwel et al. (2018) found that adults made less accurate estimates (i.e., 367-1,367) and concluded that they were less likely to use benchmark strategies than on typical lines. However, using eye tracking (see also Sullivan et al., 2011), Di Lonardo et al. (2020) found that adults also use benchmark strategies on atypical number lines They suggested that participants' worse performance on atypical number lines (i.e., 0-7,000) occurrs because they incorrectly calculate the value of the implicit midpoint.

In the present study, we investigated participants' strategies on typical and atypical number lines in the presence or absence of an explicit midpoint (see Figure 1). We used eye tracking in combination with final estimates to infer participants' strategic use of benchmarks. Triangulation of eye tracking and performance data allowed us to make detailed conclusions about specific strategic decisions and provided evidence about participants' procedural errors.

Benchmark-Based Strategy Use

The proportional judgment view of number line estimation suggests that the number line task requires the conversion of a number to a proportion, which must then be translated into a spatial location (Barth & Paladino, 2011). In this view, solvers' estimates reflect proportional estimation abilities involving the use of benchmarks (Barth & Paladino, 2011; Huber et al., 2014; Link et al., 2014; Slusser & Barth, 2017). *Explicit* benchmarks are numerical values that are physically present on the line. For example, a typically bounded 0-100 number line consists of a left-endpoint or origin, labelled “0”, and a right endpoint, labelled “100”. *Implicit* benchmarks are not provided as part of the number line itself, but may be familiar proportions, such as one half or one quarter. For example, on a 0-100 number line, participants may use the unmarked location of the midpoint 50 (i.e., 50% benchmark) and the quartiles 25 and 75 (i.e., 25% and 75% benchmarks) to guide their reasoning. To use implicit benchmarks effectively, participants must be able to calculate the numerical value of the benchmark and physically locate it on the number line. Thus, to place 3,972 on a 0-7,000 number line, the solver may decide that this value is just a bit larger than the midpoint value of 3,500, locate the midpoint, and then adjust the target location to capture that the target is somewhat larger than 3,500.

From the proportional judgement view, the more benchmarks a solver uses to make an estimate, the higher their estimation accuracy because the target number should be closer to an available benchmark (Luwel et al., 2019; Peeters, Verschaffel, et al., 2017; Slusser et al., 2013). Thus, knowing how and when people use benchmark-based strategies may be crucial to understanding estimation. Some researchers have relied on self-reports of solvers’ strategy use (e.g., Luwel et al., 2019; Peeters, Verschaffel, et al., 2017). However, self-reports may be inaccurate, especially for relatively fast decisions, and they may be subject to instructional biases (Kirk & Ashcraft, 2001). Other researchers have fit *n*-cycle power models to participants’ estimates and inferred strategy use based on which models best fit the data (e.g., Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Reinert et al., 2015; Slusser et al., 2013). However, the fit of *n*-cycle power models is based entirely on final estimates and thus cannot be used to accurately infer the strategies people use before they make the final estimate. In contrast, eye tracking can be used to capture the entire estimation process (Di Lonardo et al., 2020; Sullivan et al., 2011).

Eye Tracking and Benchmark-Based Strategy Use

Eye tracking captures participants' estimation process from the presentation of the target number to the final estimate and thus provides information about their use of benchmarks. On typical number lines, participants fixate mainly around the endpoints and midpoint, suggesting that they use these benchmarks to guide their estimates (Heine et al., 2009; Reinert et al., 2015; Schneider et al., 2008; Sullivan et al., 2011). When graphed, these fixations therefore accumulate in a W-shaped pattern, concentrated around the endpoints and the midpoint (Schneider et al., 2008; Sullivan et al., 2011). This W-shaped pattern mirrors the M-shaped pattern of error found for typical number lines, with the least error around the endpoints and midpoint and more error around the quartiles (e.g., Ashcraft & Moore, 2012; Di Lonardo et al., 2020; Luwel et al., 2018). In summary, eye tracking results reflect behavioural findings regarding benchmark use and provide insight into participants' strategies.

Eye tracking data may be particularly useful as a tool to investigate participants' estimation process on atypical number lines. Di Lonardo et al. (2020) used eye tracking to investigate benchmark-based strategies on typical (i.e., 0-10,000) and atypical (i.e., 0-7,000) number lines. For typical number lines, errors showed the M-shaped pattern described by Ashcraft and Moore (2012), with the most accurate estimates occurring for targets around the endpoints and midpoint. In contrast, errors on atypical number lines showed the tent-shaped pattern described by Luwel et al. (2018), with more accurate estimates for targets around the endpoints than for those around the midpoint and quartiles. Luwel et al. concluded that participants were less likely to use internal benchmarks on atypical number lines. However, using eye tracking data, Di Lonardo et al. showed that participants used the centre of the line as part of their estimation process on both typical and atypical number lines. Di Lonardo et al. suggested that participants failed to correctly calculate the numerical value of the midpoint on the atypical number line. They therefore argued that the tent-shaped error pattern of final estimates reflected that the numerical value participants attributed to the midpoint was incorrect. This conclusion was supported by eye tracking data, which showed equivalent use (i.e., fixations and dwell time) of the centre of the line for both typical and atypical number lines, but less accurate placement of the actual midpoint value (i.e., 3,500) on the atypical number line. In the present study, we manipulated the presence of a labelled midpoint to directly test the hypothesis that accuracy differences between typical and atypical number lines reflect an incorrect calculation of the midpoint for atypical number lines.

Explicit Benchmarks and Number Line Estimation

Research on number lines has used two types of explicit benchmarks: unlabelled and labelled. Unlabelled benchmarks consist of a hatch mark but no numerical value. Labelled benchmarks consist of both a hatch mark and a numerical value. The addition of labelled and unlabelled benchmarks has been found to improve number line estimation accuracy for both adults and children (Luwel et al., 2019; Peeters et al., 2016; Peeters, Sekeris, et al., 2017; Peeters, Verschaffel, et al., 2017; Xu & LeFevre, 2016). However, Peeters, Sekeris, et al. (2017) found that benefits of including a labelled benchmark were different for Grade 3 and Grade 6 children. For children in Grade 3 (i.e., 8 years old), estimation accuracy was better on lines with labelled benchmarks than on those with unlabelled benchmarks. In contrast, for children in Grade 6 (i.e., 11 years old), accuracy was similar for lines with labelled versus unlabelled benchmarks. Peeters, Sekeris et al. suggested that midpoint-based strategies could only be used by third graders when a labelled midpoint was provided. In contrast, sixth graders could use midpoint-based strategies even when the midpoint was unlabelled. They concluded that the benefit of labelled benchmarks depends children's age and their familiarity with the number line range. With adults, there is evidence that benchmark-based strategies are used on both typical and atypical number lines (Di Lonardo et al., 2020). Thus, it is possible that adults are less accurate estimates on atypical number lines because they are less familiar with the number line range. More specifically, adults may incorrectly calculate the numerical value of the midpoint (Di Lonardo et al., 2020). Including a labelled midpoint may therefore support better performance for adults on atypical number lines.

Present Study

In the present study, the presence of a labelled explicit midpoint was manipulated across both typical and atypical number lines. Participants in the explicit-midpoint condition were shown lines with a labelled midpoint (i.e., Figure 1c and 1d) whereas participants in the implicit-midpoint condition were shown lines without a labelled midpoint (i.e., Figure 1a and 1b). The presence of the midpoint was manipulated between groups to minimize carryover effects. To provide a strong test of the importance of a labelled midpoint for more challenging, less familiar number lines, participants saw both a typical and an atypical number line (i.e., 7,000 vs. 10,000). Patterns of fixations, acquired with eye tracking, were used to examine strategy use whereas patterns of error were used to examine estimation performance.

Estimation Accuracy

We hypothesized that the presence of an explicit midpoint would result in more accurate estimates on atypical number lines only (Hypothesis I). Specifically, we hypothesized that the error patterns for the atypical (i.e., 0-7,000) number line would be tent-shaped in the implicit-midpoint condition and M-shaped in the explicit-midpoint condition, whereas the error patterns for the typical (i.e., 10,000) number line would be M-shaped for both conditions (Hypothesis II). For the atypical number line, providing the numerical value of the midpoint eliminates any need to calculate the value of the midpoint benchmark and thus should lead to more accurate estimates around the midpoint. In contrast, the numerical value of the midpoint is easy to calculate on a typical number line, so providing an explicit midpoint is unlikely to affect performance (Peeters, Sekeris, et al., 2017). Thus, we did not expect significant differences in estimation accuracy between the two midpoint conditions for the typical number line.

Strategy

We hypothesized that the pattern of first fixations would reflect strategy selection. Specifically, following Di Lonardo et al. (2020), we expected that participants would make the greatest proportion of first fixations around the origin and midpoint for low target numbers (i.e., target numbers less than the midpoint) and around the midpoint and endpoint for high target numbers (i.e., target numbers greater than the midpoint; Hypothesis III). We also hypothesized that participants in the explicit-midpoint condition would make more first fixations around the midpoint than participants in the implicit-midpoint condition, regardless of number line range (Hypothesis IV). Participants in both midpoint conditions were expected to use the middle of the line as a benchmark, but in the explicit-midpoint condition, participants were expected to make a greater proportion of initial fixations around the midpoint because they would be drawn to the visual cue provided by the labelled reference point (see Figure 1).

Method

Participants

Twenty-nine undergraduate students (M_{age} = 20.0 years, SD_{age} = 3.4; $Median_{age}$ = 19; Range = 18-30; 14 women) from a Canadian university participated in the study. All participants spoke fluent English, with 86.2% identifying English as their first language. Participants were required to have normal or corrected-to-normal vision. Participants were enrolled in an introductory Psychology or Cognitive Science course and received partial course credit as compensation for their participation. Students were in a variety of different programs, including, for example, cognitive science, psychology, biology, law, communications, and statistics. The study was approved by the Carleton University Research Ethics Board. Written consent was obtained from all participants. Fourteen participants were in the implicit-midpoint condition (M_{age} = 20.0 years; 6 women) and 15 participants were in the explicit-midpoint condition (M_{age} = 19.8 years; 8 women).

Materials

Demographic Questionnaire

Participants indicated their age, gender, visual acuity, program of study, and language.

Calculation Fluency Test

Arithmetic fluency was measured with the Calculation Fluency Test (CFT; Sowinski et al., 2014). This paper-and-pencil measure consists of three pages, each containing 60 arithmetic problems. The first page consists of double-digit addition problems (e.g., 45 + 91), the second page consists of double-digit subtraction problems (e.g., 91 – 64), and the third page consists of double- by single-digit multiplication problems (e.g., 49 x 2). Before completing the task, participants completed two practice problems for each operation (i.e., a total of six). Then, participants were given 60 seconds per page to complete problems as quickly and accurately as possible. The total number of correct answers across all operations was used as a measure of participants' arithmetic fluency.

Bounded Number Line Task

A number line with a length of 1720 pixels was used for the number line task. The endpoints of this line were marked for all participants (see Figure 1a-d). Participants saw both a typical (i.e., 0-10,000) and an atypical (i.e., 0-7,000) number line but were assigned to one of two benchmark conditions: (i) 10,000 implicit-midpoint and 7,000 implicit-midpoint (see Figure 1a and 1b; n = 14) or (ii) 10,000 explicit-midpoint and 7,000 explicit-midpoint (see Figure 1c and 1d; n = 15). The trials for typical and atypical lines were completed in separate blocks so that the value of the right endpoint did not change on a trial-by-trial basis. Thus, 14 participants saw the typical number line first and 15 participant saw the atypical number line first.

The target numbers used for the typical and atypical number lines are included in the Appendix. The target numbers were chosen such that there was one stimulus for each numerical century (e.g., 2,000-2,100). Thus, 100 stimuli were created for the typical number line, and 70 were created for the atypical number line. The target list included the exact quartiles and midpoint (e.g., on the atypical number line, 1,750, 3,500, and 5,250). Each participant saw a different random order of the stimuli within each condition. Participants in the implicit- and explicit-midpoint conditions were asked to estimate the same target numbers.

Apparatus

All number line stimuli were visually presented on a 23-inch 1920 x 1080 pixels desktop monitor. Participants' eye fixations and saccades were recorded using an SR Research EyeLink 1000 eye tracker, running at a sampling rate of 1000 Hz. Eye fixations and saccades for the right eye were recorded using a specialized camera. The right eye was used by default, but the left eye was used if calibration failed for the right eye. To keep the head as still as possible, participants completed the experiment in a chin rest. Their head was approximately 70 cm away from the eye tracker's camera and 110 cm from the computer monitor.

Procedure

All tasks were completed individually. After providing written informed consent, participants completed the demographic questionnaire on a laptop computer. Then, participants were instructed to place their head in the chin rest. The height of the chin rest and chair were adjusted to maximize participant comfort and eye tracker accuracy.

Next, participants began the number line task. The experiment was conducted using EyeLink Experiment Builder software (SR Research, Kanata, Ontario, Canada). Prior to beginning the task, instructions appeared on the computer monitor and the experimenter read them aloud. The instructions specified that participants were to make their estimates as quickly and accurately as possible. Prior to estimating on each number line, a nine-point calibration was completed followed by two practice trials. Each trial began with the presentation of a black fixation square, in the top right corner of the screen, for 2,000 ms. After this fixation square disappeared, the target number appeared in the top left corner of the screen and the number line appeared in the middle of the screen, simultaneously. Participants had up to 10 seconds to fixate on the location where they believed the number should go on the line. They were instructed to press the spacebar once they were fixating on this location. If they failed to press the spacebar within 10 seconds, their last fixation was recorded as their estimate and the experiment automatically proceeded to the next trial. In between the two number lines, participants completed the CFT. After completing the second number line, participants were debriefed.

Results

Arithmetic fluency, as measured by the CFT, did not significantly differ between participants in the implicit-midpoint (*M* = 27.0, *SD* = 17.5) and explicit-midpoint conditions (*M* = 30.3, *SD* = 15.6), $t(27) = 0.53$, $p = .60$, $d = 0.20$. Furthermore, total CFT scores in the present sample (*M* = 28.7; *SD* = 16.3) were similar to those of a norming sample of 242 Canadian participants (*M* = 31.6; *SD* = 13.5; Sowinski et al., 2014), $t(28) = 0.97$, $p = .34$. Thus, the participants had average arithmetic fluency and there was no evidence that fluency differed across the benchmark conditions.

Analysis of Estimation Performance

Of the 29 participants, three participants were missing fixation data for their final estimates for a large portion of trials (> 33%), so their data were removed from all estimation accuracy analyses. These individuals failed to fixate on the line when making final estimates, likely due to calibration issues or moving their eye away from the line too quickly. Thus, their final estimates were not accurate. However, their first fixations were analyzed, as first fixations would be unaffected by the final estimate movement. For the remaining 26 participants, 56 of the 4,420 trials (1.27%) were missing final fixation data, likely because the participant accidentally pressed the spacebar too early, before making an estimate. These trials were excluded from analyses.

Estimation accuracy was calculated as percentage absolute error (PAE) based on Petitto's (1990) formula: (|[Participant's estimate – Target Number]| / Scale of estimate) × 100. For example, on a 0-7,000 number line, if a participant was presented with a target of 3,000, but their estimate corresponded to the position for the number 2,500, the resulting PAE value would be 7.14% (|[2,500 – 3,000]| / 7,000 × 100). To account for estimates that may have reflected a shift in fixation prior to pressing the spacebar as opposed to a true reflection of estimation location, the mean PAE for each of the four number lines (i.e., 10,000 implicit-midpoint, 7,000 implicit-midpoint, 10,000 explicit-midpoint, and 7,000 explicit-midpoint) was calculated, and trials where the PAE was greater than two standard deviations above the mean were treated as outliers and removed from the analyses. This trimming resulted in the removal of 175 (4.0%) of the 4,364 trials. After removing these outliers, mean PAE across all conditions and number lines was 5.94% (*SD* = 5.18). Accuracy patterns were examined using mean PAE for each target number.

In addition to frequentist statistics, Bayes factors are reported, allowing for the evaluation of the fit of the data under both the null and alternative hypotheses. The Bayes factor, BF_{01}, is "a ratio that contrasts the likelihood of the data fitting under the null hypothesis with the likelihood of fitting under the alternative hypothesis" (Jarosz & Wiley, 2014, p. 3). The inverse, BF_{10}, puts the ratio in terms of the alternative hypothesis. For the present study, BF_{01} is reported when the Bayes factor is in favour of the null hypothesis; BF_{10} is reported when the Bayes Factor is in favour of the alternative hypothesis. The inclusive Bayes Factor, BF_{inc}, is used for analyses with more than two factors. The inclusive Bayes Factor is similar to BF_{10}, but it represents the change from prior to posterior inclusion odds. That is, it compares two classes of models – one with the factor of interest and one without – and presents the odds in favour of the models that include the factor. Bayes Factor values can range from 0 to infinity. For all Bayesian analyses, the strength of the evidence for the null or alternative hypothesis was interpreted in accordance with Jeffreys' (1961) guidelines (see Table 4 of Jarosz & Wiley, 2014). Bonferroni corrections were used for post-hoc comparisons of significant effects. Error bars in figures represent 95% confidence intervals, as recommended by Jarmasz and Hollands (2009). All analyses are Greenhouse-Geisser corrected, where appropriate.

Estimation Accuracy

Overall Performance

Estimation accuracy (PAE) was analyzed in a 2 (midpoint condition: implicit, explicit) x 2 (range: typical, atypical) mixed ANOVA, with repeated measures on range. There was a significant effect of range, with participants making more accurate estimates on the typical 10,000 number line ($M = 5.53\%$) than on the atypical 7,000 number line ($M = 6.77\%$), $F(1, 24) = 12.23$, $p = .002$, $\eta_P^2 = .34$. The estimated Bayes factor, $BF_{10} = 10.04$ indicates strong support for the effect of range. The effect of condition was not significant, $F(1, 24) = 3.18$, $p = .087$, $\eta_P^2 = .12$, $BF_{10} = 1.12$. However, the range by condition interaction was significant, $F(1, 24) = 6.66$, $p = .016$, $\eta_P^2 = .22$. The estimated Bayes factor, $BF_{10} = 3.68$, indicates substantial support for a model that includes the range by condition interaction (see Figure 2). In support of Hypothesis I, on the atypical number line, participants made more accurate estimates when the midpoint was explicitly marked ($M = 5.68\%$) than when the midpoint was not provided ($M = 7.87\%$). In contrast, on the typical number line, participants' accuracy did not differ in the presence ($M = 5.35\%$) or absence ($M = 5.70\%$) of the midpoint.

(Insert Figure 2 approximately here)

Error Performance on the Midpoint Trial

Accuracy on the midpoint trial was analyzed in a 2 (midpoint condition: implicit, explicit) x 2 (range: typical, atypical) mixed ANOVA, with repeated measures on range. There was a significant effect of range, with participants more accurately estimating the location of the midpoint on the typical 10,000 number line ($M = 2.63\%$) than on the atypical 7,000 number line ($M = 3.87\%$), $F(1, 22) = 4.40$, $p = .048$, $\eta_P^2 = .17$, $BF_{01} = 1.32$. The effect of condition was significant, with better performance in the condition with the explicit midpoint ($M = 1.28\%$) than the implicit midpoint (*M* = 5.58%), $F(1, 22) = 8.76$, $p = .007$, $\eta_P^2 = .29$, $BF_{10} = 6.76$. The main effects were qualified by the range by condition interaction, $F(1, 22) = 8.79$, $p = .007$, $\eta_P^2 = .29$. The estimated Bayes factor, $BF_{10} = 5.72$, indicates substantial support for a model that includes the range by condition interaction. On the atypical number line, participants less accurately estimated the midpoint in the implicit-midpoint than in the explicit-midpoint condition (*M* = 7.27% vs. *M* = 0.99%, $p = .001$). In contrast, on the typical number line, performance on the midpoint trial did not significantly differ for those in the implicit- vs. explicit-midpoint conditions (*M* = 3.89% vs. *M* = 1.57%, $p = .14$). Thus, participants in the implicit-midpoint condition were less accurate at placing the midpoint on the atypical number line than participants in the explicit-midpoint condition. This pattern suggests that participants in the implicit-midpoint condition did not know the value of the midpoint.

Error Performance in Relation to Benchmarks

Accuracy was examined around five potential benchmarks: origin (0%), first quartile (25%), midpoint (50%), third quartile (75%) and endpoint (100%). These potential benchmarks were chosen to coincide with those used in previous research (e.g., Ashcraft & Moore, 2012; Di Lonardo et al., 2020; Luwel et al., 2018; Sullivan et al., 2011). The PAEs for the five target numbers closest to the endpoints (0%, 100%) and the PAEs for the exact internal benchmarks (25%, 50%, 75%) and the two target numbers closest to either side of these benchmarks, were averaged for each participant, resulting in 25 target numbers per participant for the analyses.

Mean PAE was analyzed in a 2 (midpoint condition: implicit, explicit) x 2 (range: 10,000, 7,000) x 5 (benchmark: 0%, 25%, 50%, 75%, 100%) mixed ANOVA with repeated measures on range and benchmark. Similar to the overall analysis, there was a significant main effect of range, $F(1, 23) = 11.24$, $p = .003$, $\eta_P^2 = .33$, with participants making more accurate estimates on the typical than atypical number line (M = 4.58% vs. M = 6.09%). The estimated Bayes factor, BF_{10} = 37.76 indicates very strong support for the effect of range. There was also a significant main effect of benchmark, $F(3.25, 74.79) = 15.70$, $p < .001$, $\eta_P^2 = .19$. The estimated Bayes factor, BF_{10} = 1.77e+10, indicates decisive support for the effect of benchmark. In general, participants made significantly more accurate estimates around the origin (M = 2.78%), midpoint (M = 5.06%), and endpoint (M = 4.75%) than around the quartiles (M_{25} = 7.86%, M_{75} = 6.33%). The main effect of midpoint condition was not significant, $F(1, 23) = 1.93$, $p = .178$, $\eta_P^2 = .08$, BF_{01} = 2.19.

The two-way interactions were not significant, but the main effects were qualified by a significant three-way range by condition by benchmark interaction, $F(2.87, 65.93) = 3.22$, $p = .03$, $\eta_P^2 = .12$. The Bayes factor, BF_{10} = 1.27, suggests that there is anecdotal evidence for the inclusion of the three-way interaction in the model. The interaction is shown in Figure 3.

(Insert Figure 3 approximately here)

Consistent with Hypothesis II, for the typical number line the error pattern around the benchmarks was M-shaped in both the explicit- and implicit-midpoint conditions, with more accurate estimates around the midpoint and endpoints than the quartiles (Figure 3, right panel). Post-hoc analyses revealed no significant differences between explicit- and implicit-midpoint conditions at any of the benchmarks (ps > .05).

In contrast, for the atypical number line, the error pattern around the benchmarks differed based on condition (Figure 3, left panel). The pattern of error was tent-shaped for participants in the implicit-midpoint condition and M-shaped for those in the explicit-midpoint condition. Across the five benchmarks for atypical number lines, post-hoc analyses revealed that accuracy was only significantly different around the midpoint benchmark (p = .01), with participants in the explicit-midpoint condition making more accurate estimates around the midpoint (*M* = 3.42%) than participants in the implicit-midpoint condition (*M* = 8.31%). Thus, providing an explicit midpoint helped participants make more accurate estimates only on trials near the midpoint.

Analyses of Fixations

The estimation accuracy analyses provide useful information about how final estimates differ across range and condition, but they do not provide information about strategy use. Thus, we turn to eye-tracking to investigate potential differences in strategy selection between the conditions.

First Fixations

To investigate participants' strategy use, graphs of fixation patterns were made for participants' first fixations. Following Sullivan et al. (2011) and Di Lonardo et al. (2020), each number line was divided into 20 regions, with each region containing 5% of the number line. Target numbers below the midpoint were classified as "low numbers" and target numbers above the midpoint were classified as "high numbers". Low and high target numbers are contrasted to reflect participants' initial decisions relating to strategy use as a function of target size. The proportion of first fixations made in each of the 20 regions, separated by low and high target numbers, are shown in Figure 4a to 4d.

(Insert Figure 4 approximately here)

Replicating the findings of Di Lonardo et al. (2020), and consistent with Hypothesis III, participants made more fixations around the origin and midpoint for low target numbers and around the midpoint and endpoint for high target numbers. This pattern suggests that participants' first fixations were meaningful indications of their estimation processes. The peaks around the endpoints and midpoint across the four graphs are in accordance with the W-shaped pattern reported by Sullivan et al. (2011). This pattern suggests that, across ranges and midpoint conditions, participants made use of benchmark-based strategies.

Consistent with Hypothesis IV, participants in the explicit-midpoint condition made more first fixations around the midpoint than participants in the implicit-midpoint condition (Figures 4c and 4d). This pattern is clear in all situations except for the low targets for the 10,000 number line. Thus, for the atypical line, where the midpoint's numerical value was more difficult to calculate, participants looked at the midpoint more when the value was provided for them. In contrast, for the typical number line, participants were more likely to look at the explicit midpoint only for high target numbers.

Di Lonardo et al. (2020) suggested that for typical number lines, participants use a left-to-right scanning strategy. Our data support this conclusion for the implicit-midpoint condition. As shown in Figures 4a and 4b, participants made a high number of first fixations around the left endpoint for low target numbers. Even for high target numbers, which were all located on the right side of the line, first fixations were concentrated on the left side of the number line. In contrast, in the presence of an explicit midpoint, participants were more likely to make their first fixation around the midpoint (Figures 4c and 4d), suggesting they used it as a starting point regardless of target size. This pattern was especially noticeable for the atypical number line, where high peaks were present around the midpoint (Figure 4c).

To statistically test the hypotheses about first fixations (Hypotheses III and IV), the proportion of first fixations made around five potential benchmarks were analyzed in a 2 (midpoint condition: implicit, explicit) by 2 (range: 10,000, 7,000) by 2 (target size: low, high) by 5 (benchmark: 0%, 25%, 50%, 75%, 100%) mixed ANOVA with repeated measures on range, target size, and benchmark. A fixation was assigned to a benchmark if it was within 10%

of the external benchmarks (i.e., 0-10% and 90-100%) and plus or minus 5% of the internal benchmarks (i.e., 20-30%, 45-55%, and 70-80%). For simplicity, the ANOVA results are presented in Table 1.

(Insert Table 1 about here).

There was a significant main effect of benchmark. In general, a greater proportion of first fixations were made around the left side of the number line (i.e., origin and midpoint) than the right side of the number line (i.e., third quartile and endpoint). The interaction of benchmark by condition was significant (see Figure 5). Consistent with Hypothesis IV, participants in the explicit-midpoint condition made a significantly higher proportion of first fixations around the midpoint than participants in the implicit-midpoint conditions ($p < .001$). None of the other differences were significant. Thus, participants were more likely to use the midpoint as an anchor if it was explicitly provided, regardless of target size and range. Furthermore, participants in the explicit-midpoint condition made a greater proportion of first fixations around the midpoint than around any other benchmark. In contrast, participants in the implicit-midpoint condition made more first fixations around the origin than around any other benchmark, indicating they were more likely to start on the left and scan to the right.

(Insert Figure 5 approximately here)

The two-way benchmark by condition and target size by benchmark interactions were also significant, as was the three-way interaction of target size by benchmark by range (see Figure 6). For low target numbers, the proportion of first fixations was only different between the typical and atypical number lines at the 25th percentile benchmark ($p = .04$). For high target numbers, the proportion of first fixations was only different at the 75th percentile benchmark ($p < .001$). Participants made a greater proportion of first fixations around the 25% and 75%

benchmarks when estimating on the typical number line in comparison to the atypical number line. Thus, even when the midpoint value was provided, participants were more likely to use the quartile as a starting point on typical than on atypical number lines. As hypothesized (Hypotheses III and IV), participants used the midpoint as a benchmark to guide their estimates both in the presence and absence of a labelled midpoint, but they were more likely to make their first fixations around the midpoint when it was explicitly labelled.

(Insert Figure 6 approximately here)

Analyses of Number of Fixations Per Trial

The overall mean number of fixations per trial was 5.77. The overall mean dwell time per trial was 2,274 ms. Mean fixations and dwell time per trial were analysed in separate 2 (range: 10,000, 7,000) by 2 (midpoint condition: implicit, explicit) mixed ANOVAs. The main effect of midpoint condition was significant for both fixations and dwell time. Participants in the implicit-midpoint condition made significantly more fixations per trial than participants in the explicit-midpoint condition (7.55 vs. 4.60), $F(1, 27) = 13.56$, $p < .001$, $\eta_P^2 = .33$. The Bayes factor, $BF_{10} = 26.43$, suggests that there is strong evidence for the inclusion of the effect of condition in the model. Similarly, participants in the implicit-midpoint condition had significantly longer dwell times per trial than participants in the explicit-midpoint condition (3118 ms vs. 1487 ms), $F(1, 27) = 9.83$, $p = .004$, $\eta_P^2 = .27$. The Bayes factor, $BF_{10} = 7.67$, suggests that there is substantial evidence for the inclusion of the effect of condition in the model. The main effect of range and the range by condition interaction were not significant for either fixations or duration. Thus, when participants were provided with a labelled midpoint, they made fewer fixations and had shorter dwell times per trial. The midpoint functioned as an additional reference to guide their estimates, allowing them to make fewer and shorter fixations before they chose a final location.

Di Lonardo et al. (2020) reported analyses of mean number of fixations and dwell times around the benchmarks. They concluded that the findings from those analyses did not provide additional information about strategy use beyond first fixations. We conducted a similar analysis, and, consistent with Di Lonardo et al., found that participants in both conditions used the benchmarks to guide estimates. In the absence of a labelled midpoint, participants had longer dwell times around the benchmarks, suggesting participants needed more time to strategize about estimation placement when the midpoint was not provided. Analyses of mean number of fixations and dwell time around the benchmarks are provided in the Supplementary Material.

In summary, patterns of fixations and dwell times showed that strategy selection aligned with target size, with participants using the low benchmarks for low target numbers and the high benchmarks for high target numbers. Furthermore, participants had longer dwell times and made more fixations across the line and around the benchmarks in the implicit- than in the explicit-midpoint condition. Thus, although both groups used benchmark-based strategies, in the absence of an explicit midpoint, participants made more fixations and spent longer fixating on the line, including around the benchmarks, in order to make accurate estimates.

Model Fit for Benchmark Estimation Patterns

Following Luwel et al. (2018) and Di Lonardo et al. (2020), one- and two-cycle power models were fit for participants across conditions and number lines. Cyclic power models have been used to assess strategy use, but these statistical models rely only on final estimates; strategy use is inferred from the model that best fits the participant's estimates. A one-cycle model is assumed to reflect benchmark strategies that only use the endpoints, whereas a two-cycle model is assumed to reflect the use of both the endpoints and midpoint. The percentage of participants best fit by each of these models (based on the Akaike Information Criterion [AIC]) is presented in Table 2.

(Insert Table 2 approximately here)

Consistent with the findings of Di Lonardo et al. (2020), cyclic power models did not accurately reflect participants' strategy use. For instance, although fixation and dwell time data provide strong evidence for the use of the midpoint for participants in the implicit-midpoint condition, most people in this condition were fit by one-cycle models for both the typical and atypical number line. Furthermore, for the typical number line, people in the explicit-midpoint condition made the greatest proportion of first fixations around the midpoint and fixations and dwell times were equally high around the midpoint and the endpoints, yet the majority were best fit with a one-cycle model. Only for the atypical explicit-midpoint number line did the data for a majority of participants fit a two-cycle model, emphasizing that estimates were most accurate around the midpoint and endpoints. Thus, model fits provide information about accuracy, not about strategy (Di Lonardo et al., 2020).

Discussion

In this study, we examined how the presence of a labelled midpoint influenced number line estimation accuracy and strategy choice for adults' performance on typical and atypical number lines. Participants estimated on both a typical (i.e., 0-10,000) and an atypical (i.e., 0-7000) number line, with half of the participants estimating on lines with the midpoint value and location provided (i.e., explicit-midpoint condition) and the other half estimating on lines without a midpoint (i.e., implicit-midpoint condition). Luwel et al. (2018) and Di Lonardo et al. (2020) both found that people made less accurate estimates on atypical number lines. Although both Luwel et al. and Di Lonardo et al. suggested that participants' less accurate estimates on atypical number lines resulted from difficulties in calculating the numerical value of the midpoint, Luwel et al. concluded that participants were less likely to use benchmark strategies for atypical number lines because of the calculation difficulty. In contrast, Di Lonardo et al. concluded that despite calculation difficulties, adults used benchmark strategies on both typical and atypical number lines.

Effects on Accuracy

In the present study, using eye-tracking data, we compared accuracy and strategy use in the presence or absence of an explicit midpoint. We found support for the hypothesis that accuracy would be better on the atypical number line when an explicit midpoint was provided. On the atypical 7,000 number line, participants in the explicit-midpoint condition made significantly more accurate estimates than participants in the implicit-midpoint condition. In contrast, for the typical 10,000 number line, estimation accuracy did not significantly differ between the two conditions. We further explored accuracy by examining the pattern of error around potential benchmarks. As hypothesized, accuracy around the benchmarks for the typical number line formed an M-shaped pattern of error, regardless of whether the midpoint was implicit or explicit. In contrast, for the atypical number line, the error pattern was tent-shaped in the implicit-midpoint condition, similar to that observed in other research with atypical endpoints (e.g., Di Lonardo et al., 2020; Luwel et al., 2018), whereas it was M-shaped in the explicit-midpoint condition (e.g., Ashcraft & Moore, 2012; Di Lonardo et al., 2020; Luwel et al., 2019). Thus, in support of our hypotheses, providing the numerical value of the midpoint appears to be beneficial when the midpoint value is difficult to calculate (Peeters, Sekeris, et al., 2017).

How Does a Labelled Midpoint Influence Number Line Strategies?

Do participants engage in different strategies in the presence of an explicit midpoint? In the present study, strategy use was examined through first fixations, overall fixations, and dwell times across the number line and around potential benchmarks across number line ranges and midpoint conditions. Luwel et al. (2018) concluded that adults are less likely to use benchmark-based strategies on atypical number lines because they have difficulty calculating benchmark values. In contrast, Di Lonardo et al. (2020) found that although participants use benchmark-based strategies on atypical number lines, but they assign the incorrect value to the benchmarks. For example, on a 0-7,000 number line, participants may divide the line into two segments, but incorrectly assign a midpoint value of 3,000 or 4,000 instead of 3,500. In the present experiment, we found that participants relied on benchmark-based strategies and target size for both typical and atypical number lines, regardless of whether an explicit midpoint was provided. However, people made more accurate estimates around the midpoint and made more first fixations around the midpoint when the midpoint was explicit. This pattern suggests that people were drawn to the visual cue provided by the labelled reference point.

Although first fixations provided the most useful information about strategy selection, mean number of fixations and dwell times provided some additional evidence for the use of benchmark-based strategies. Overall, participants in the implicit-midpoint condition had longer dwell times around the benchmarks than participants in the explicit-midpoint condition. However, in general, participants in the implicit-midpoint condition made more fixations and had longer fixation durations, regardless of whether the target number was near a benchmark. This pattern suggests participants in the implicit-midpoint condition required more adjustments to narrow down the final location of their estimates (Sullivan et al., 2011). Nonetheless, participants in both the implicit- and explicit-midpoint conditions continued to use the benchmarks after their first fixations were made to guide their estimates on both typical and atypical number lines. Therefore, providing the midpoint helped participants choose a final location, whereas accuracy only improved around the midpoint. These findings suggest that calculation difficulties rather than strategy differences explain the lower accuracy on the atypical number line for the implicit- versus explicit-midpoint conditions.

Power Models to Infer Strategy Use

In the present study, *n*-cycle power functions were used to determine the best model fit for final estimates to reflect approaches used by other researchers (e.g., Ashcraft & Moore, 2012; Barth & Paladino, 2011; Luwel et al., 2018; Slusser et al., 2013). We found that model fits did not accurately reflect strategy selection (see also Di Lonardo et al., 2020). The eye-tracking data provided evidence for the use of benchmark-based strategies, especially the origin and midpoint for typical and atypical number lines both in the presence and absence of a labelled midpoint. In contrast, the *n*-cycle model fits suggested that only participants in the explicit-midpoint condition used strategies that involved the midpoint and endpoints, and that they only used these strategies for the atypical number line. Thus, the *n*-cycle models reflect the accuracy of the estimates but not strategy selection. Accordingly, we suggest that multiple measures are necessary to understand how people approach number line estimation under various conditions.

Limitations and Future Research

The present study has several limitations. First, the use of the eye-tracking paradigm resulted in approximately 5% of trials being lost because there were no final fixations. In general, in eye-tracking studies it is important to have many trials to offset expected data loss. The benefits of using eye tracking outweighed this limitation because this method provided important information about strategy use.

Second, we did not include an unlabelled midpoint condition, in which participants would have seen the spatial location of the midpoint but not its numerical value (Peeters, Sekeris, et al., 2017). Including this condition in future work would help to disambiguate the role of spatial bias versus numerical bias in the implicit midpoint condition. Specifically, when people bisect horizontal lines they show a spatial bias, referred to as pseudo neglect, in that they place the bisection to the left of the midpoint and underestimate the midpoint relative to its true location (Bowers & Heilman, 1980; Di Lonardo et al., 2020; Rotondaro et al., 2015). Thus, providing unlabelled midpoints may result in different patterns of estimation and strategy selection than those that people show in either the labelled or implicit midpoint conditions. In future studies, comparisons of the usefulness of a labelled vs. unlabelled midpoint for typical and atypical number lines would help distinguish between the benefits associated with having a marked midpoint (e.g., improved strategy selection) and those specifically associated with knowing the value of the midpoint.

Third, this study did not address the issue of why adults fail to correctly calculate the value of the midpoint for atypical number lines. Are they unable to calculate the value in their head? Do they estimate the value rather than calculating the exact value? The calculation of the midpoint is only required once; the value can then be stored and retrieved for use on remaining trials. In future studies, researchers could ask participants to report the value of the midpoint for both typical and atypical number lines.

Conclusions

This study provided evidence about the effects of a labelled midpoint for adults' performance on typical and atypical number lines. Eye-tracking data provided useful insights into strategy selection whereas behavioural data provided information about estimation accuracy. First, the tent-shaped pattern of error often observed for atypical number lines occurs when solvers miscalculate the numerical value of the midpoint (Di Lonardo et al., 2020; Link et al., 2014; Luwel et al., 2018). Second, we replicated the finding that participants rely on benchmark-based strategies for both typical and atypical number lines (Di Lonardo et al., 2020). Notably, we found that adults use midpoints but not quartiles as benchmarks. Providing the midpoint for atypical number lines helps participants to make more accurate estimates only for target numbers around the midpoint. Third, when participants were provided with an explicit midpoint, they appeared to be more confident in their estimates in that they made fewer fixations and had shorter dwell times per trial on both typical and atypical number lines. In summary, these findings corroborate previous findings and provide new insights into strategy selection in the presence and absence of labelled benchmarks.

References

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology*, *111*(2), 246–267. https://doi.org/10.1016/j.jecp.2011.08.005

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

Bowers, D., & Heilman, K. M. (1980). Pseudoneglect: Effects of hemispace on a tactile line bisection task. *Neuropsychologia*, *18*(4–5), 491–498. https://doi.org/10.1016/0028-3932(80)90151-7

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review*, *18*(2), 331–338. https://doi.org/10.3758/s13423-011-0059-z

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. https://doi.org/10.1177/1747021819881631

Experiment Builder. (n.d.). SR Research.

Heine, A., Thaler, V., Tamm, S., Hawelka, S., Schneider, M., Torbeyns, J., De Smedt, B., Verschaffel, L., Stern, E., & Jacobs, A. M. (2009). 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. https://doi.org/10.1002/icd.640

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

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

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *The Journal of Problem Solving*, *7*(1). https://doi.org/10.7771/1932-6246.1167

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

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, and Cognition*, *27*(1), 157–175. https://doi.org/10.1037/0278-7393.27.1.157

Link, T., Nuerk, H.-C., & Moeller, K. (2014). On the Relation between the Mental Number Line and Arithmetic Competencies. *Quarterly Journal of Experimental Psychology*, *67*(8), 1597–1613. https://doi.org/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/Revue Canadienne de Psychologie Expérimentale*, *72*(4), 253–263. https://doi.org/10.1037/cep0000153

Luwel, K., Peeters, D., & Verschaffel, L. (2019). Developmental change in number line estimation: A strategy-based perspective. *Canadian Journal of Experimental Psychology*, *73*(3), 144–156. https://doi.org/10.1037/cep0000172

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. https://doi.org/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. https://doi.org/10.3389/fpsyg.2017.01082

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

Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction*, *7*(1), 55–78. https://doi.org/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*(S1), 359–363. https://doi.org/10.1007/s10339-015-0675-z

Rotondaro, F., Merola, S., Aiello, M., Pinto, M., & Doricchi, F. (2015). Dissociation between line bisection and mental-number-line bisection in healthy adults. *Neuropsychologia*, *75*, 565–576. https://doi.org/10.1016/j.neuropsychologia.2015.07.016

Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smedt, B., Verschaffel, L., Jacobs, A. M., & Stern, E. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development*, *23*(3), 409–422. https://doi.org/10.1016/j.cogdev.2008.07.002

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

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. https://doi.org/10.1111/1467-9280.02438

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

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology. General*, *142*(1), 193–208. https://doi.org/10.1037/a0028560

Sowinski, C., Dunbar, K., & LeFevre, J.-A. (2014). *Calculation Fluency Test*. Carleton University. Available on the Open Science Framework at https://osf.io/hrgka/

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

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. https://doi.org/10.1037/dev0000124

Table 1

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

Source	*df*	MS	*F*	*p*	ηp^2	BF_{inc}
Range	1	2.35e-4	0.46	.51	.02	0.021
Benchmark	2.65	0.086	19.60	<.001***	.42	∞
Target Size	1	0.001	1.91	.18	.07	∞
Condition	1	0.002	2.83	.10	.10	1387.08
Target Size x Benchmark	2.16	0.087	18.92	<.001***	.41	∞
Target Size x Range	1	4.97e-6	0.01	.92	<.001	0.016
Target Size x Condition	1	5.21e-4	0.80	.38	.03	0.166
Benchmark x Range	2.10	.007	2.32	.11	.08	0.034
Benchmark x Condition	2.65	.023	5.20	.004***	.16	8873.78
Range x Condition	1	4.91e-5	0.10	.76	.004	0.016
Target Size x Benchmark x Range	3.13	.003	3.79	.01**	.12	0.005
Target Size x Benchmark x Condition	2.16	.004	0.89	.42	.03	0.157
Target Size x Range x Condition	1	7.60e-5	0.15	.70	.006	<.001
Benchmark x Range x Condition	2.10	.003	0.91	.41	.03	0.002
Target Size x Benchmark x Range x Condition	3.13	9.98e-4	1.27	.29	.05	<.001
Total	84.41					

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

Table 2

Percentage of participants in each condition whose data are best fit by one- and two-cycle models

Number Line	One-Cycle	Two-Cycle	Neither*
Implicit-Midpoint: Typical	64	21	14
Implicit-Midpoint: Atypical	54	38	8
Explicit-Midpoint: Typical	53	33	13
Explicit-Midpoint: Atypical	7	80	13

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

Figure 1

The four number line conditions: (a) 10,000 implicit-midpoint, (b) 7,000 implicit-midpoint, (c) 10,000 explicit-midpoint, (d) 7,000 explicit-midpoint.

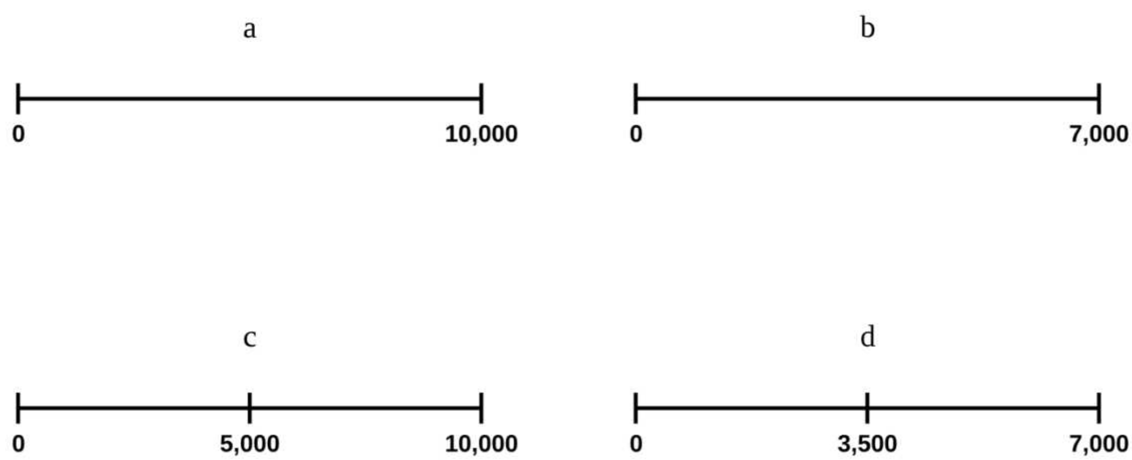

Figure 2

Mean percentage of error (PAE) for typical (i.e., 10,0000) and atypical (i.e., 7,000) number lines as a function of benchmark label condition. Error bars represent 95% confidence intervals.

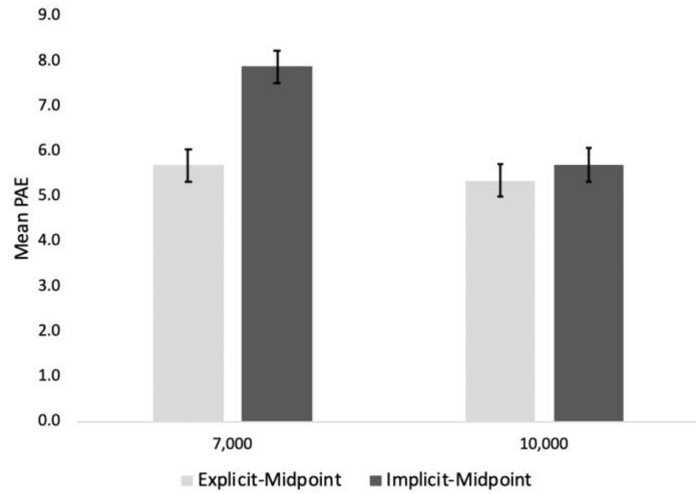

Figure 3

Mean percentage of error (PAE) around the benchmarks for the atypical (left) and typical (right) number lines as a function of benchmark label condition. Error bars represent 95% confidence intervals.

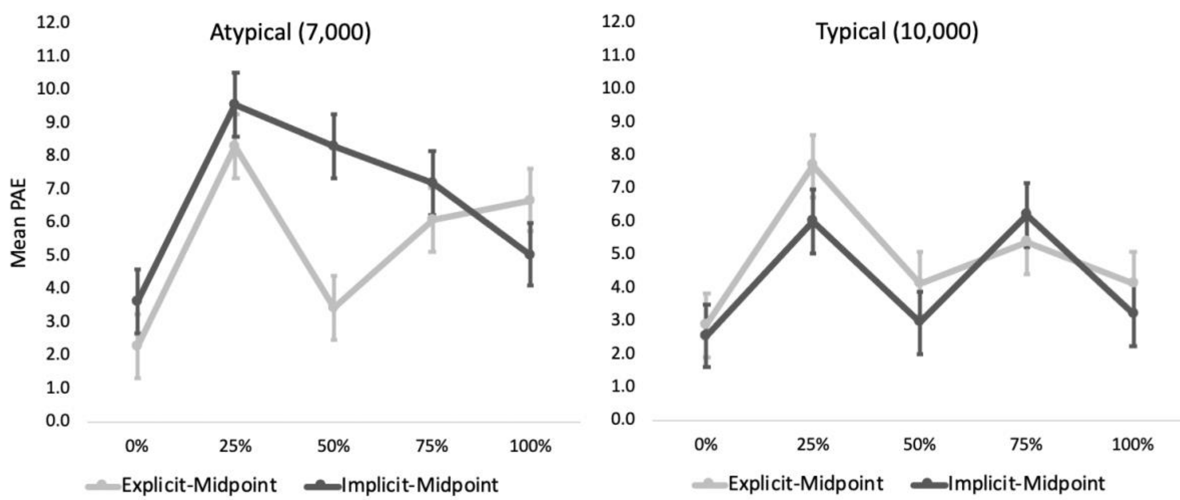

Figure 4

Proportion of first fixations observed in each region for low (light grey) and high (dark grey) target numbers: (a) 7,000 implicit-midpoint, (b) 10,000 implicit-midpoint, (c) 7,000 explicit-midpoint, (d) 10,000 explicit-midpoint.

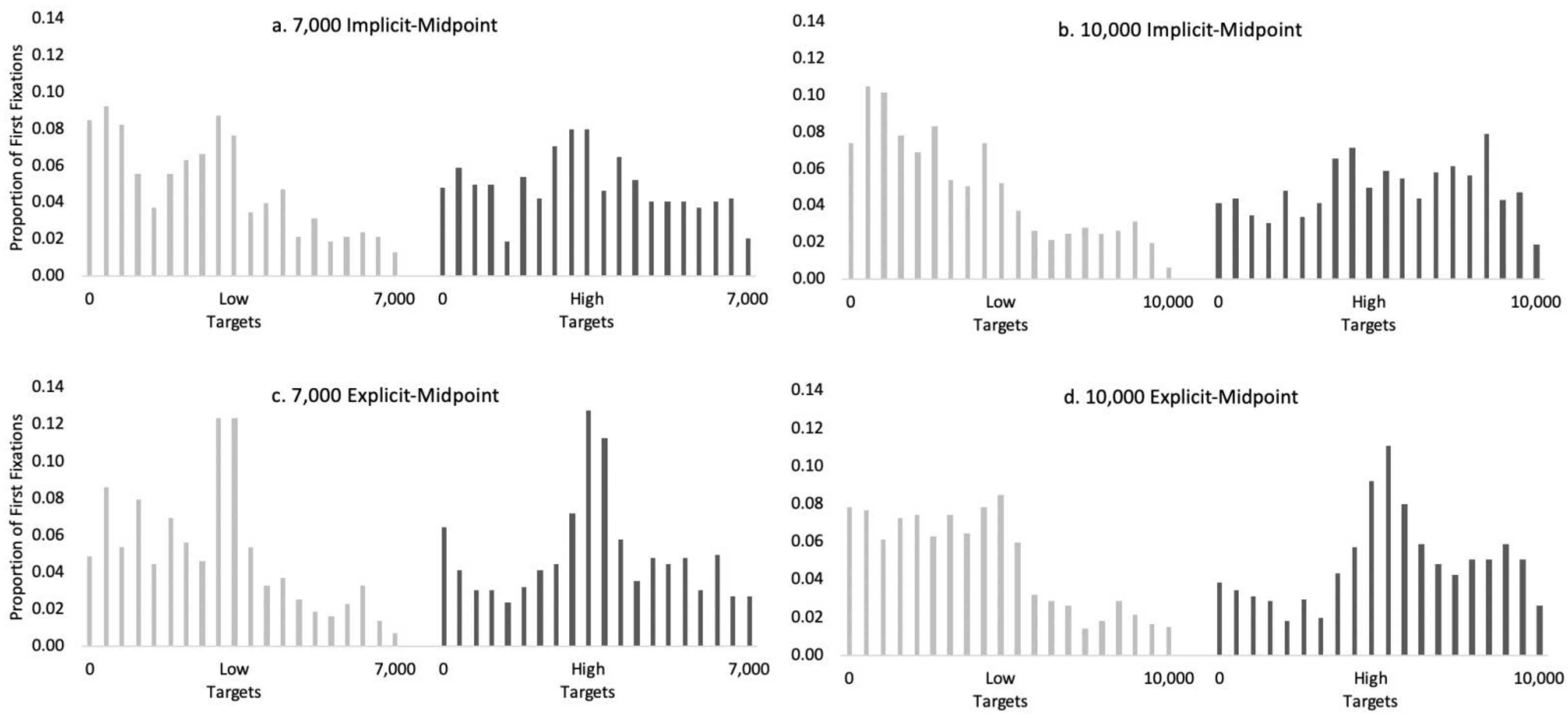

Figure 5

Proportion of first fixations around the benchmarks as a function of benchmark label condition. Error bars represent 95% confidence intervals.

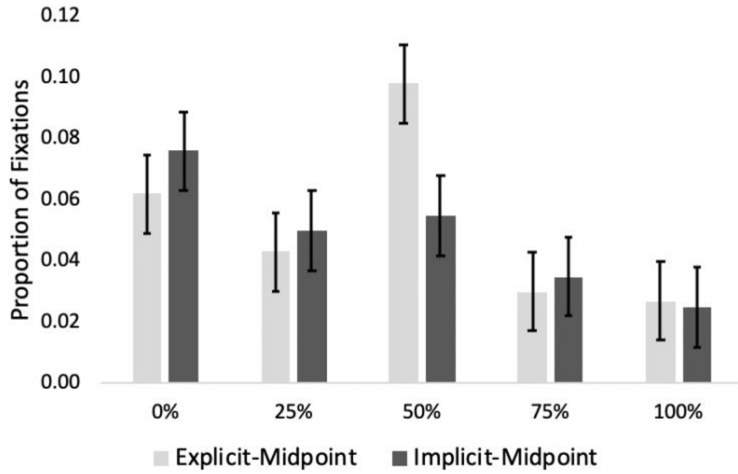

Figure 6

Proportion of first fixations for (a) low target numbers and (b) high target numbers for typical and atypical number lines. Error bars represent 95% confidence intervals.