Summary accuracy feedback and the left digit effect in number line estimation

Kelsey Kayton[1] · Katherine Williams[1] · Claudia Stenbaek[1] · Gina Gwiazda[1] · Charles Bondhus[1] · Jordan Green[1] · Greg Fischer[1] · Hilary Barth[1] · Andrea L. Patalano[1]

Abstract

A robust left digit effect arises in number line estimation such that adults' estimates for numerals with different hundreds place digits but nearly identical magnitudes are systematically different from one another (e.g., 299 is placed too far to the left of 302). In two experiments, we investigate whether brief feedback interventions designed to increase task effort can reduce or eliminate the left digit effect in a self-paced 0–1,000 number line estimation task. Participants were assigned to complete three blocks of 120 trials each where the middle block contained feedback or no feedback. Feedback was in the form of summary accuracy scores (Experiment 1; $N = 153$) or competitive (summary) accuracy scores (Experiment 2; $N = 145$). In both experiments, planned analyses revealed large left digit effects in all blocks regardless of feedback condition. Feedback did not lead to a reduction in the left digit effect in either experiment, but improvements in overall accuracy were observed. We conclude that there are no changes in the left digit effect resulting from either summary accuracy feedback or competitive accuracy feedback. Also reported are exploratory analyses of trial characteristics (e.g., whether 299 is presented before or after 302) and the left digit effect.

Keywords Left digit effect · Number line estimation · Numerical cognition · Feedback

Introduction

Understanding numerical magnitudes is an important skill. Numerical magnitude estimation is often measured with the *number line estimation* task. This task is used as a skills assessment and education training tool, and is also important for understanding underlying cognitive processes (e.g., Barth & Paladino, 2011; Booth & Siegler, 2008; Brez et al., 2016; Schneider et al., 2018; Siegler & Opfer, 2003; Siegler & Ramani, 2009; Slusser et al., 2013; Xing et al., 2021; Zhu et al., 2017). In a typical number line estimation task, one is shown a blank horizontal line with a label at each end (e.g., 0 and 1,000) and is asked to estimate the location of target numerals on the line (e.g., 802; see Fig. 1).[1] A primary measure of task performance is overall accuracy error, which reflects the difference between one's placement and the correct location of each numeral. Overall accuracy error has

been linked to many measures of numerical competency, including counting and fraction skills in children (Hamdan & Gunderson, 2017; Hansen et al., 2015; Jordan et al., 2013; Östergren & Träff, 2013), math performance on standardized achievement tests (Booth & Siegler, 2008; Holloway & Ansari, 2009; Schneider et al., 2009; Tosto et al., 2017; see also Schneider et al., 2018, for review) and numeracy in adults (Patalano et al., 2020; Peters & Bjalkebring, 2015; Schley & Peters, 2014), even when controlling for potential confounding variables (e.g., Bailey et al., 2014; Geary, 2011; Hansen et al., 2015; Hornung et al., 2014; Östergren & Träff, 2013; Zhu et al., 2017).

Some error in number line estimation is known to be systematic. The bias that is the subject of most work in this area, in its simplest form, is a tendency to overestimate numerals on one half of the line and to underestimate numerals on the other half. This bias has been modeled as an S-shaped or an inverse-S-shaped curve, with the direction and degree

✉ Andrea L. Patalano
apatalano@wesleyan.edu

[1] Department of Psychology, Wesleyan University, 207 High Street, Middletown, CT 06459, USA

---

[1] We focus here on bounded number line estimation tasks, but note that there are also unbounded versions (e.g., in which one is given just the left bound plus an indication of the width of one unit) that elicit different error patterns (Cohen et al., 2018; Cohen & Blanc-Goldhammer, 2011).
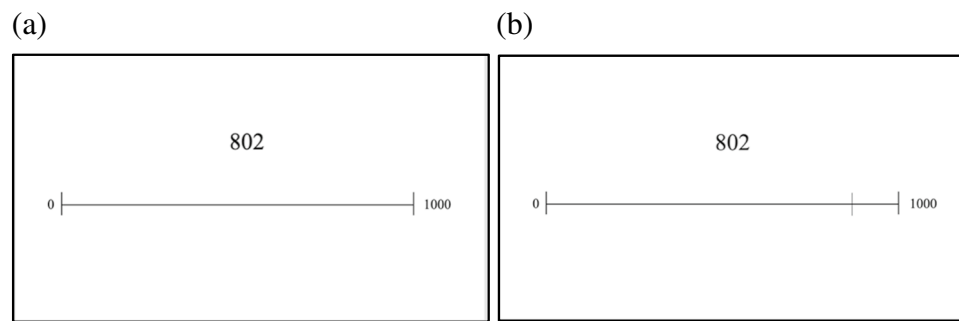
Fig. 1 Number line estimation display used here (**a**) before and (**b**) after response. *Note.* Participants clicked on the horizontal line to estimate the location of the target numeral. The vertical placement line (indicating the selected location) in the second image was red in color. A new target numeral appears in the same position above the line on each trial. The figures are scaled images of the center of the computer screen

of bias indicated by a parameter estimate β (e.g., Cohen & Blanc-Goldhammer, 2011; Slusser & Barth, 2017; but see Siegler et al., 2009). The shape of the curve is thought to be the result of imprecision in one's estimate of individual magnitudes (e.g., Dehaene et al., 2008; Siegler & Opfer, 2003) and in the relationship of the part to a whole (e.g., estimating 599 as a proportion of 1000; Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen et al., 2018; Slusser & Barth, 2013; see also Hollands & Dyre, 2000; Zax et al., 2019; Zhang & Maloney, 2012). The pattern of bias may also be multi-cyclical (e.g., two S-shapes in a row), with the number of cycles thought to depend on the number of reference points, besides the two endpoints, used to perform the task (e.g., using the line's midpoint of 500 as an additional reference point; Hollands & Dyre, 2000; Peeters et al., 2017; Slusser et al., 2013; Sullivan et al., 2011). All else being equal, placements are typically more accurate when more reference points are used.

In most work to date, it is the overall value of the target numeral that is used for predicting placements. For example, for targets 598 and 601, placements would be predicted to be similar because the targets have nearly the same overall magnitude. However, it has recently been demonstrated that the individual digits comprising a numeral also contribute to placement of a target on a line. Lai et al. (2018) observed that when asked to estimate the locations of three-digit numerals on a 0–1,000 number line, individuals exhibited *a left digit effect*, placing numerals with different leftmost (hundreds-place) digits but similar overall magnitudes farther apart than is warranted. In contrast, they placed numerals with different tens-place digits (e.g., 348 vs. 352) in the same location on the line, suggesting that the bias is driven by the leftmost digit. The effect is very large (ds $\approx$ 1 in adults), and appears in task variations such as a speeded version of the task (Lai et al., 2018; Williams et al., 2020), and various numerical ranges (e.g., 0–100; Vaidya et al., 2022; Savelkouls et al., 2020; Williams et al., 2021). There is also noticeable individual-level variation in the size of the left digit effect.

Although the left digit effect in number line estimation has been reported only relatively recently, related phenomena have been observed in *numerical comparison* tasks and in judgment tasks. Numerical comparison tasks involve deciding which of two numbers is larger (e.g., 59 vs. 61). In these tasks, a *distance effect* arises in that numerals that are farther apart produce faster response times (Dehaene et al., 1990; Moyer & Landauer, 1967). While these effects are most often attributed to comparisons of overall magnitudes, there is evidence that individual digits also matter. For example, longer response times are observed for comparable pairs with the same leftmost digit (47 vs. 49) than for pairs with a different leftmost digit (49 vs. 51; Moeller et al., 2009; Nuerk et al., 2011; Verguts & De Moor, 2005). Relatedly, in *price judgment* tasks, a product costing on or above a whole-dollar value (e.g., $5.00) is judged as significantly more costly than one priced just below the whole-dollar (e.g., $4.99), while products whose prices do not cross such a boundary are perceived as equally costly (e.g., $4.20 vs. $4.19; Beracha & Seiler, 2015; Lin & Wang, 2017; MacKillop et al., 2014; Manning & Sprott, 2009; Sokolova et al., 2020; Thomas & Morwitz, 2005, 2009). This finding has been extended to judgments about nutritional information (Choi et al., 2019), product evaluations (Thomas & Morwitz, 2005), medical records (Olenski et al., 2020), and hypothetical college portfolios (Patalano et al., 2022), revealing significant consequences of a left digit bias for real decisions.

One important question that has emerged is the extent to which the left digit effect in number line estimation arises because people do not put sufficient effort into performing the task accurately. If it does, the effect might be reduced using simple motivational interventions designed to increase task effort (by encouraging participants to put more effort into making accurate placements). This question is important in that it speaks to the malleability of the left digit

effect, as well as the source of the effect and the contexts in which it is likely to emerge. If the left digit effect that has been observed in past work occurs under conditions in which task effort is not sufficiently high, the effect might be easily reduced or eliminated with a simple motivational intervention, and the effect should be unlikely to arise in everyday contexts of importance to an individual. Alternatively, if the effect is not easily reduced through such interventions, the effect should emerge even in everyday contexts in which one strives for accuracy, and elimination of the effect will likely demand more carefully developed interventions. This research question is also of value to researchers seeking to identify strong individual difference measures of magnitude estimation skills, as such measures should reflect as much as possible one's estimation skills, not variations in task effort.

There is some reason to believe that the left digit effect could be related to task effort. Although the specific cognitive processes that underlie the left digit effect are not well understood, it widely believed that the left digit effect reflects an overweighting of the leftmost digit of multidigit numerals (e.g., Lacetera et al., 2012; Thomas & Morwitz, 2005). It is possible that this overweighting arises, at least in part, because individuals give careful attention to the leftmost digit but devote less attention to rightward digits, leading the latter to be underweighted in magnitude estimates. If true, motivational interventions might reasonably lead to a reduction in the left digit effect. There are many possible reasons for how or why this might happen, but one possibility is that when people reduce attention to rightward digits, they do so strategically to reduce effort, aware that it is possible to give a "close enough" estimate without fully attending to rightward digits. When motivated to be as accurate as possible, they might increase attention to rightward digits in order to increase the precision of their estimate. By this description, intervention-based reductions in the left digit effect would arise from the use of general strategies associated with increasing task effort towards improving accuracy, rather than from trying to reduce the left digit effect per se. This is important in that we have no reason to believe that people are aware of the left digit effect or would be explicitly trying to reduce it in this context.

No work to date has considered the malleability of the left digit effect in response to motivational interventions intended to increase task effort; however, there are a few suggestive findings from related work in numerical cognition. Notably, Eyler et al. (2018) assessed whether a trial-by-trial feedback intervention would reduce overall accuracy error in number line estimation in adults. Although the intervention was brief (feedback was given on only five trials), researchers found that overall accuracy error decreased by about a third for a subset of participants (specifically, those with no college education). In more distantly related numerical cognition studies that use non-symbolic tasks (e.g., judging which set of dots is larger), accuracy feedback interventions have been found to improve performance in adults (De Wind & Brannon, 2012), with improvements attributed to increased motivation, rather than to any intervention-related changes in knowledge (Lindskog et al., 2013). In addition, variability of an individual's scores across time points on at least one non-symbolic task has been attributed to lapses of attention during task performance (Peters & Bjalkebring, 2015). All of these findings point to the possibility that a motivational intervention intended to increase task effort might serve to reduce the left digit effect in number line estimation.

Overview of experiments

In two experiments, three 120-trial blocks of a 0–1,000 self-paced number line estimation task were administered. There were two between-subjects conditions: no-feedback and feedback. In the no-feedback condition, all three blocks of the task were the same. In the feedback condition, the middle block was modified to serve as an intervention. In this block, a summary accuracy score (on a 0–100 scale where 100 is perfect accuracy across all 20 trials) was provided after each set of 20 trials along with instructions to try to improve one's score over time. Participants in this condition were also instructed that one reason people often give estimates that are not precisely correct is because they do not pay careful attention to all the digits, and were periodically reminded to pay attention to all digits. The purpose of the intervention was not to give detailed instructions or feedback that would indicate in what direction to adjust one's responses. Rather, the intervention served to test whether simply motivating efforts to perform the task as accurately as possible (including attending to all digits) would lead to a reduction in or elimination of the left digit effect (and reduction in accuracy error generally).

As in past work, we computed three dependent measures (Lai et al., 2018). The left digit effect was assessed using a dependent measure called a *hundreds difference score* (Lai et al., 2018), for which we focused on eight critical pairs of target values, called *hundreds pairs*, with similar magnitudes but different leftmost hundreds place digits (e.g., 199 and 201). Of interest was whether the larger value in each pair would be placed too far to the right of the smaller value on the line (assuming they should be placed in approximately the same location given the numerical range and physical line length used here). To use for comparison, we also computed *fifties difference scores* in the same manner except using nine pairs of target values, called *fifties pairs*, with similar magnitudes but surrounding fifties boundaries (e.g., 149 and 151). As a measure of overall accuracy error, we computed *percent absolute error* (PAE), a commonly used

measure of one's overall performance on the number line estimation task. PAE reflects the differences between one's placements and the correct locations of targets as a percentage of line length, using all trials except those used in computing hundreds and fifties difference scores.

If the motivational intervention leads to a reduction in the left digit effect, we should see an interaction effect in both experiments with regard to the hundreds difference score. Namely, there should be a greater reduction in the hundreds difference score across blocks in the feedback condition than in the no-feedback condition. In contrast, if the intervention does not reduce the left digit effect, we should not see any greater reduction of the hundreds difference score in the feedback condition relative to the no-feedback condition. Regarding PAE, given the suggestive findings of Eyler et al.'s (2018) feedback study, and the fact that there are likely many sources of the errors that contribute to PAE, we had reason to believe that PAE might decrease as a result of the feedback intervention whether or not the hundreds difference score also decreased.

In addition to asking the important question regarding the malleability of the left digit effect, we also used this opportunity to consider properties of trials that might be related to the size of the hundreds difference score. In exploratory analyses, we used combined data (from both experiments) from each participant's first block of trials to test whether each of the following properties of pairs of hundreds trials is related to the size of the hundreds difference score: (1) distance between paired targets (i.e., number of intervening trials), (2) order of paired targets (i.e., whether the larger or smaller target in a pair was presented first), and (3) pair boundary (i.e., which boundary the pair surrounded, as in 200, 300, etc.). One past study that looked at boundary pairs individually found a left digit effect for all pair boundaries except those at the 200 and 500 boundaries (Williams et al., in press), thereby suggesting differences across pair boundaries, but no research has yet considered pair distance or order. These exploratory analyses were conducted here to provide additional clues as to when and why the left digit effect emerges.

Experiment 1: Summary accuracy feedback

In this experiment, we compared performance across three blocks of a number line estimation task for participants who had a summary accuracy feedback intervention in the middle block versus those who did not (i.e., who had three identical blocks). If the left digit effect is reduced or eliminated as a result of summary accuracy feedback, we should see an interaction of condition by block; specifically, there should be greater reduction in the hundreds difference score across blocks in the feedback condition than in the no-feedback

condition. If the left digit effect also decreases as a result of task practice (regardless of whether the feedback intervention affects performance), we should also see a main effect of block, where scores decrease across blocks in both conditions. While our focus was on the left digit effect, we also conducted the analyses with our measure of overall accuracy error, PAE, to compare findings.

Method

Participants

Participants were 153 undergraduates (89 women, 63 men, one undisclosed) who received Introductory Psychology course credit for their participation. They were run individually in a lab setting in 1-h sessions. Participants were assigned in alternation to either a no-feedback condition ($n = 79$) or a feedback condition ($n = 74$).[2] They completed a number line estimation task (as well as several cognitive tasks and scales unrelated to this report). A power analysis ($1 - \beta = .80$, $\alpha = .05$) indicated samples of $n \approx 40$ per condition would be needed to detect a moderately small effect size (Cohen's $f = 0.15$) for the interaction between condition and block in an ANOVA. The study was approved by the Wesleyan Institutional Review Board; participants gave written consent to participate in the study and were debriefed at its conclusion.

Stimuli

Stimuli for a 0–1,000 number line estimation task were displayed using PsychoPy3 software onto a desktop computer monitor (47 cm wide × 27 cm high; screen resolution 2,560 × 1,440 pixels). On each trial, a horizontally centered target numeral (e.g., 47; 1.5 cm tall) located 8 cm above a black horizontal line (20 cm long) was presented, as shown in Fig. 1a. The horizontal line, which was in the center of the screen, had small vertical lines at each end (1 cm long). The endpoints of the horizontal line were labeled '0' on the left and '1,000' on the right (0.8 cm tall). After a participant selected a location on the line, a vertical red line (1 cm long) immediately appeared in the selected location, as shown in Fig. 1b.

Each block (of three) consisted of the same 120 target numerals between 0 and 1,000. Target numerals were selected to fall on either side of hundreds boundary values (eight pairs of values: 199/202, 298/302, 398/403, 499/502, 597/601, 699/703, 798/802, 899/901), fifties boundary

---

values (nine pairs of values, used as controls: 149/152, 248/252, 348/352, 449/451, 549/551, 648/653, 748/752, 849/853, 947/951),[3] and 82 non-boundary values (e.g., 235, 367, 411). Non-boundary values were used to compute percent absolute error (PAE), a standard measure of task accuracy (e.g., Lai et al., 2018; Petitto, 1990; Siegler & Booth, 2004; Siegler & Opfer, 2003). We use only non-boundary values in computing PAE in order to have an independent set of estimates for calculating this measure. Target numerals were presented one at a time in a different randomized order within each block for each participant. Numerals were "paired" for the purposes of analyses only; they were not paired during presentation.

Procedure

Each participant was seated in front of a computer and given written instructions followed by three blocks of 120 trials each. On each trial, participants responded with a mouse click and a vertical red line appeared in the selected location on the response line. The task was self-paced, but participants were instructed to respond as quickly and accurately as possible. After each response, a rectangular button icon (labeled "Next") appeared centered at the bottom of the screen to advance to the next trial. A 0.5-s blank screen was presented before each new trial. Coordinates of mouse clicks were recorded and converted to a number between 0 to 1,000, corresponding to the selected location on the response line.

In the no-feedback condition, all three blocks were identical and no feedback was given. In the feedback condition, the first and third blocks were identical but, in the middle block, participants received feedback after every 20 trials, six times in total. Feedback was presented as an overall accuracy score ranging from 0 to 100, where a higher score reflects greater accuracy. Participants were instructed that a score of 0 indicates their responses were as distant as they could be from the correct locations while a score of 100 indicates responses were precisely in the correct locations (see Online Supplemental Materials (OSM) for full instructions).

To calculate a feedback score, we first computed one's average PAE (*PAE = (|actual location – estimated location| / numerical range) * 100*) over the preceding 20 trials. We then calculated an accuracy score by subtracting the PAE from 100 (*accuracy score = 100 – PAE*). Finally, to obtain

---

Your last score was 30 out of 100.

Your score for this set of trials is 50 out of 100.

See if you can improve on this in the next set!

Remember to attend to *all* digits.

Fig. 2 Example of the summary feedback screen in the feedback condition. This feedback screen appeared after each set of 20 trials (a total of six times) in the middle block of the feedback condition only. The first line of text in the above was omitted on the first feedback screen; the last two lines were omitted on the sixth (last) feedback screen

the feedback score, we assigned all accuracy scores less than 90 to a feedback score of 0, while accuracy scores between 90 and 100 were rescaled to values between 0 and 100 (*feedback score = (accuracy score – 90) * 10*).[4] A feedback score of "50 out of 100," for example, would correspond to a PAE of 5% and an accuracy score of 95.

At the time summary feedback was provided, participants in the feedback condition were also instructed to do their best to improve their score and were reminded to carefully attend to all digits comprising each numeral (see Fig. 2). Participants could view the feedback screen for as long as they wished before continuing on to the next set of trials.

Results and discussion

Exclusions

All exclusion criteria and data analyses were planned unless otherwise indicated; exclusion criteria were the same as those in Lai et al. (2018) and Patalano et al. (in press). An individual's estimate for a target number was removed as an outlier from the computation of hundreds and fifties difference scores (but not from the computation of PAE) if it differed from the group mean for a given target by more than two standard deviations (3.26% of trials, on average, were removed within each block). In addition, we excluded from all analyses participants missing more than three hundreds pairs (as a result of outlier removal) from one or more blocks (*n* = 7 excluded). A total of 146 participants were in the final

---

[3] Additional target numerals around the boundaries of 50 (48/52) and 100 (99/101) were included in the stimulus set of 120 target numerals but were not included in any analyses (because the lower values were not three-digit numbers). The pattern of findings does not change with inclusion of these values in the computation of the average hundreds score or the average fifties score respectively in either Experiment 1 or Experiment 2.
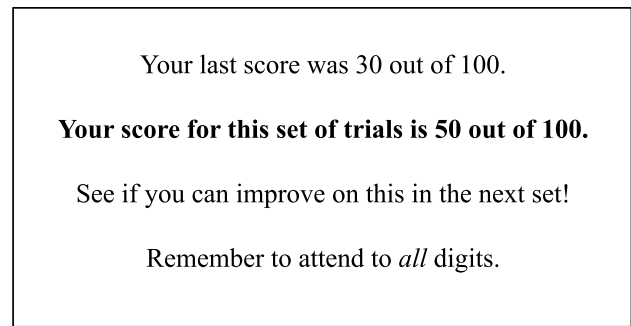
[4] We adjusted accuracy scores this way to create the feedback scores because adults nearly always have an accuracy score > 90 on this task. We wanted feedback scores to accurately reflect relative performance, but not to appear extremely high, to motivate participants in our sample.

Table 1 Descriptive statistics for Experiment 1

	Block 1	Block 2	Block 3
No-feedback condition ($n = 77$)			
Hundreds difference score	18.75 (20.42)	22.79 (18.15)	21.44 (20.15)
Fifties difference score	1.86 (16.46)	–2.42 (15.55)	1.07 (17.88)
Percent absolute error (PAE)	3.90 (1.32)	3.79 (1.32)	3.67 (1.32)
Feedback condition ($n = 69$)			
Hundreds difference score	20.34 (17.99)	15.76 (20.42)	19.61 (23.51)
Fifties difference score	–1.25 (13.59)	0.70 (11.36)	–2.78 (16.37)
Percent absolute error (PAE)	3.81 (1.17)	3.04 (1.15)	3.27 (1.52)

*SD*s are in parentheses. As predicted, all hundreds difference scores and PAEs (and no fifties difference scores) were reliably greater than 0 ($ps < .001$).

dataset (no-feedback condition $n = 77$, feedback condition $n = 69$). For descriptive and inferential statistics with outliers retained ($N = 153$, because participants no longer needed to be excluded for missing hundreds pairs) (see OSM).

Dependent measures

We computed three dependent measures for each participant: PAE, hundreds difference score, and fifties difference score. PAE was calculated as (|*actual location – estimated location*| */ numerical range) * 100*, using estimates of all non-boundary target values. Hundreds and fifties difference scores were computed as (*estimated location of larger numeral – estimated location of smaller numeral*), averaged over the eight hundreds pairs (e.g., the estimate for 703 minus the estimate for 699; see Lai et al., 2018) and, separately, over the nine fifties pairs (e.g., the estimate for 653 minus the estimate for 648). If specific left digits matter, hundreds pairs (target values with similar magnitudes but different leftmost hundreds place digits) will not be placed in the same location on the response line and the larger number in the pair will be placed to the right of the smaller number. Thus, evidence of a left digit effect comes from hundreds difference scores that are greater than zero. We included fifties pairs (target values with similar magnitudes and the same leftmost hundreds place digit) as controls; these numbers should be placed in the same location on the response line, and so we expected fifties difference scores to be no different from zero. We calculated hundreds and fifties difference scores for each block separately so that we could evaluate intervention-related changes in the left digit effect.

Planned analyses

See Table 1 for descriptive statistics for all dependent measures by condition and block. We first asked whether there was evidence of the left digit effect in each block and condition. To do this, we conducted *t*-tests (all two-tailed)

and found that hundreds difference scores were reliably greater than 0 in each of the three blocks in both the no-feedback ($ts > 8$, $ps < .001$, $ds = 0.92–1.26$) and the feedback condition ($ts > 6$, $ps < .001$, $ds = 0.77–1.13$). Fifties difference scores were not different from 0 in any block in either condition ($|t|s < 1.5$, $ps > .160$). Based on these findings, we conclude that there is a large left digit effect even when summary feedback is provided. There were no gender differences ($|t|s < 2$, $ps > .090$) except in block 3 of the no-feedback condition in which hundreds difference scores were larger for women than men ($M = 25.48$ vs. 15.43 respectively), $t(75) = 2.20$, $p = .031$; and in block 1 of the feedback condition, in which fifties difference scores were larger for women than men ($M = 1.69$ vs. –4.97 respectively), $t(65) = 2.03$, $p = .047$.

To assess whether summary feedback leads to a reduced left digit effect (even if it does not fully extinguish it), we conducted a two-way mixed ANOVA with one between-subjects independent variable (condition: no-feedback, feedback) and one within-subjects variable (block: 1, 2, 3). We predicted that if summary feedback leads to a reduction in the left digit effect, a condition by block interaction should arise for hundreds difference scores; we found none, $F(2, 288) = 1.85$, $MSE = 370.27$, $p = .159$. There was also no effect of condition, $F(1, 144) = 1.34$, $MSE = 478.25$, $p = .249$, or block, $F(2, 288) = 0.17$, $MSE = 370.27$, $p = .844$ (see Fig. 3). If summary feedback leads to increased overall accuracy, there should be a condition by block interaction for PAE; in this case, an interaction was found, $F(2, 288) = 11.58$, $MSE < 0.01$, $p < .001$, $\eta_p^2 = .07$; and there were main effects of block, $F(2, 288) = 24.00$, $MSE < 0.01$, $p < .001$, $\eta_p^2 = .14$, and condition ($F(1, 144) = 4.25$, $MSE < 0.01$, $p = .041$, $\eta_p^2 = .03$ (see Fig. 4). PAE generally decreased across blocks, and especially from block 1 to block 2 in the feedback condition, as shown in Table 1. Overall, we found that summary accuracy feedback led to modest improvements in overall accuracy but did not reduce the left digit effect.
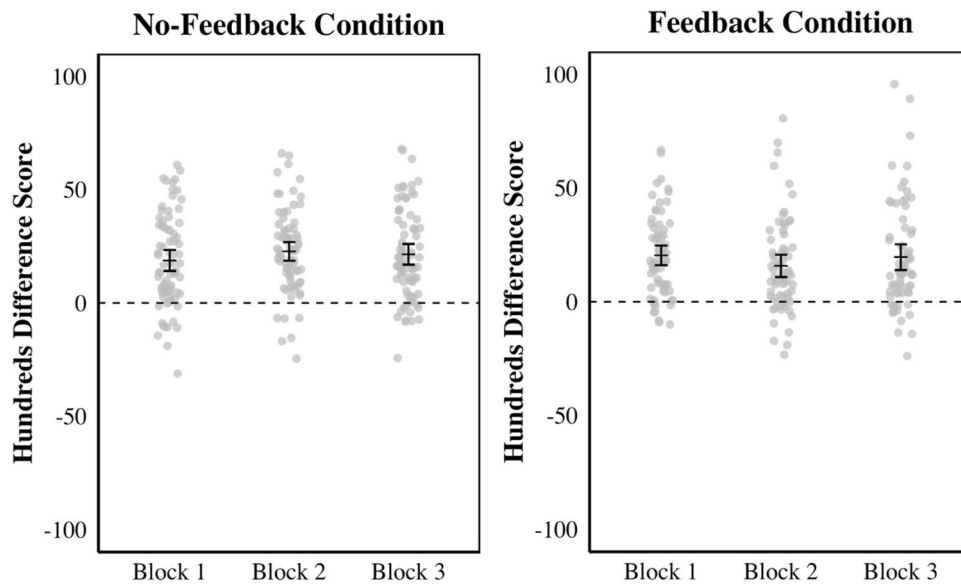
No-Feedback Condition

Feedback Condition

Fig. 3 Hundreds difference score (by condition and block) in Experiment 1. The error bars reflect ±1 SE from the mean. Hundreds difference scores greater than 0 reflect a left digit effect. The distance between points within a block along the x-axis is not meaningful; this spread of scores (here and in similar graphs) was produced to clearly show individual scores
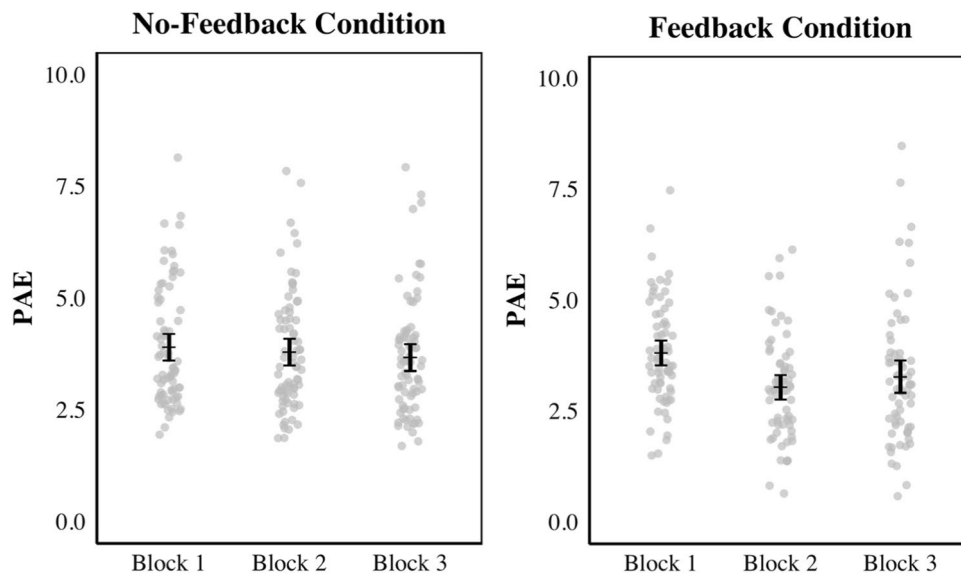
No-Feedback Condition

Feedback Condition

Fig. 4 Percent absolute error (PAE) (by condition and block) in Experiment 1. PAE values are percentages. The error bars reflect ±1 SE from the mean. Larger PAE indicates greater accuracy error on the task

In sum, in Experiment 1 we replicated the left digit effect and the large effect size. The findings provide no evidence, however, that summary accuracy feedback leads to a reduction in the left digit effect, and thus do not suggest that the effect observed in many studies could be reduced if greater effort were devoted to the task. In contrast, summary accuracy feedback did lead to reductions in overall accuracy error (specifically, a 13% reduction in accuracy error in the

feedback condition), suggesting that increased task effort can lead to improvements in performance more generally. These findings notwithstanding, we did notice that, descriptively, the pattern of means for the hundreds difference score was consistent with a feedback effect. In particular, the hundreds difference scores were smallest in block 2 of the feedback condition. It is possible that our intervention was not strong enough to motivate a change in performance that would be

detectable in our study, so we conducted a second experiment with a stronger intervention. Specifically, we added a competitive game context to further motivate accurate number line estimation performance.

Experiment 2: Competitive (summary) accuracy feedback

This experiment largely followed the design and procedure of Experiment 1 except that the summary accuracy feedback was enhanced with the addition of a scoreboard that was described as ranking the top ten highest scoring games of players that semester. Participants were instructed to try to get one's own screen name onto (or to move one's name up) the scoreboard. This design is supported by studies showing that use of competitive games can enhance motivation (e.g., Burguillo, 2010; Cagiltay et al., 2015; Chen et al., 2018). As in Experiment 1, we asked whether the feedback intervention would lead to a reduction in the hundreds difference score. If the previous findings were the result of the intervention being insufficiently motivating to lead to a reduction in the left digit effect, we would expect to see a condition by block interaction emerge here. That is, we would expect participants to show greater reduction in the hundreds difference score across blocks in the feedback condition relative to the no-feedback condition. However, if the left digit effect is not reduced by increasing one's task effort, the pattern of findings should instead be similar to Experiment 1.

Method

Participants

Participants were 145 undergraduates (85 women, 60 men) who received introductory psychology course credit for their participation. The first 60 participants were run individually in a lab setting in 1-h sessions, and the remaining 85 participants completed the study in a remote session with the guidance of an experimenter via phone (due to health and safety measures for a coronavirus outbreak).[5] The participants were assigned to either a no-feedback ($n = 81$) or a feedback condition ($n = 64$) in alternation (for participants in the lab) or by random assignment (for remote participants).[6] They completed a number line estimation task (as well as several cognitive tasks and scales unrelated to this report). The study

---

[5] The 2020 coronavirus pandemic began as this study was being conducted.

[6] An additional ten participants were not able to complete the study (e.g., due to computer malfunctions) and were not included in the final number.

Your last score was 60 out of 100.
Your game score is 87 out of 100.
Good job, Cardinal1! This score made it on the scoreboard.

Ranking	Score	Player
1	95	Blue72
2	92	Psyc1
3	89	AlliG
4	87	Wes22
5	87	Cardinal1
6	86	JCW
7	86	Psyc1
8	83	Macgyver
9	82	JohnK
10	81	JohnK

See if you can beat your score next game!
Be sure to attend to all digits.

Fig. 5. Competitive "scoreboard" feedback display presented in Experiment 2. This feedback screen appeared after every 20 trials (six times in total) during the middle block of the feedback condition. The first time it was displayed, the first line of text was omitted; the last time it was displayed, the last two lines were omitted. The display was in black and white except that the line of text starting with "Your game score is…" was written in red, and all of the text in the scoreboard was written in green

was approved by the Wesleyan Institutional Review Board; participants gave written consent to participate in the study and were debriefed at its conclusion.

Stimuli

The stimuli were the same as in Experiment 1.

Procedure

For lab participants, the number line estimation task was programmed as in Experiment 1. For remote participants, it was programmed using lab.js (lab.js. org; Henninger et al., 2019) and administered via the Open Lab platform (open-lab.online). The task procedure was the same as in Experiment 1 except that the sets of 20 trials in the feedback block were framed as computer games. At the start of the study, participants in the feedback condition were asked to provide a screen name (e.g., a made-up name, nickname, etc.) that would be used during the games. Feedback scores, although computed the same way as in Experiment 1, were now framed as competitive game scores. At the end of each game (that is, each set of 20 trials in the middle block in the feedback condition only), the feedback screen included (in addition to the summary accuracy feedback from Experiment 1)

Table 2 Descriptive statistics for Experiment 2

	Block 1	Block 2	Block 3
No-feedback condition ($n = 71$)			
Hundreds difference score	17.91 (19.72)	15.39 (18.20)	14.87 (15.76)
Fifties difference score	−1.74 (13.46)	0.54 (16.22)	−2.50 (14.80)
Percent absolute error (PAE)	3.51 (1.07)	3.45 (1.12)	3.46 (1.18)
Feedback condition ($n = 60$)			
Hundreds difference score	19.80 (20.53)	17.01 (17.66)	14.99 (16.44)
Fifties difference score	−2.50 (16.47)	0.13 (13.93)	3.45 (17.60)
Percent absolute error (PAE)	3.58 (1.18)	2.88 (1.07)	3.02 (1.21)

*SD*s are in parentheses. As predicted, all hundreds difference scores and PAEs reported above (but not fifties difference scores) were reliably greater than 0 (ps < .001).

a scoreboard containing the top ten best game scores to date in ranked order and the screen names of the players who earned them. The screen also included a statement indicating whether the participant's score for the last game played earned a position on the scoreboard (e.g., shown in Fig. 5 as "Good job, Cardinal1…"; see OSM for full instructions and feedback text). The individualized feedback and the potential to get one's own screen name onto the scoreboard were intended to provide additional motivation to participants.

The scoreboard initially contained the same ten top scores (ranging from 80 to 95 on a 0–100 scale; see OSM for the list of top scores) at the start of the feedback block. However, similar to a real video game, whenever a participant scored higher than any one of the ten scores on the board, the participant's score was added to the board in its correct ranked location. Displaced scores would then be shifted down or off the board accordingly (as shown in Fig. 2 where "Cardinal1" has had one game score that has earned a position on the board). The starting list contained representative high scores from players of the prior semester except that the lowest score was adjusted downward to be more easily attainable (so that it was a motivating goal) and the highest score was adjusted upward to be nearly unattainable (thereby ensuring no participant could exhaust all goals). At the end of the study, participants were debriefed and the scoreboard was returned to its original state to provide the same experience for each participant.

Results and discussion

Exclusions

All exclusion criteria and data analyses were preregistered unless otherwise indicated. Individual estimates that were more than two standard deviations away from the group mean for a given target numeral were excluded as outliers (3.3% of trials, on average, were removed within each block). As in Experiment 1, participants were excluded from final analyses if more than three hundred pairs were missing (i.e., were removed as outliers; $n = 14$). A total of 131 participants were included in the final dataset for analysis (no-feedback $n = 71$, feedback $n = 60$). As in Experiment 1, the dependent measures were hundreds difference score, fifties difference score, and PAE. If competitive summary accuracy feedback reduces the left digit effect, hundreds difference scores should decrease more across blocks for participants who receive feedback than for those who receive no feedback. However, if feedback does not reduce the left digit effect, any decrease in hundreds difference score across blocks should be consistent across conditions. For descriptive and inferential statistics with outliers retained ($N = 145$), see OSM.

Preregistered analyses

See Table 2 for descriptive statistics for all dependent measures by condition and block. We first asked if there was evidence of a left digit effect. To do this, we conducted one-sample t-tests (two-tailed), which revealed that hundreds difference scores were reliably greater than 0 in all blocks for both the no-feedback (ts > 7, ps < .001, ds = 0.85–0.94) and feedback condition (ts > 7, ps < .001, ds = 0.91 − 0.96). As in Experiment 1, this is evidence of a left digit effect. As predicted, fifties difference scores were not reliably greater than 0 in any block for the no-feedback (|t|s < 1.5, ps > .157) and feedback condition (|t|s < 2, ps > .134). There were no gender differences in any block in either condition (|t|s < 1.5, ps > .177). Based on these findings, and building on Experiment 1, we conclude that there is a robust left digit effect even when competitive summary accuracy feedback is provided.

The next analyses answer the main research question of whether competitive summary accuracy feedback reduces the left digit effect. We conducted a mixed ANOVA with one between-subjects independent variable (condition: no-feedback, feedback) and one within-subjects variable
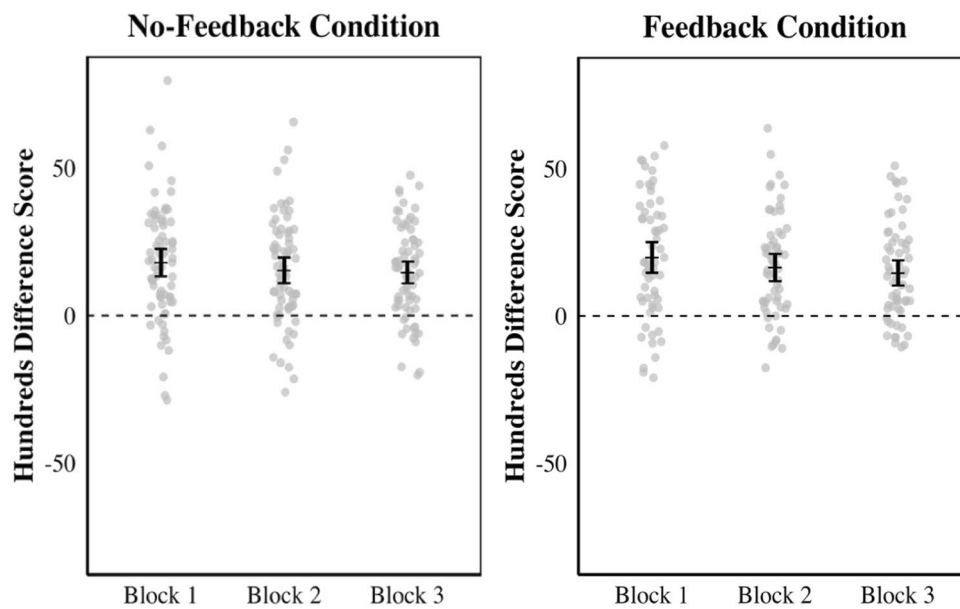
Fig. 6. Hundreds difference score (by condition and block) in Experiment 2. The error bars reflect ±1 SE from the mean hundreds difference score. Hundreds difference scores greater than 0 reflect a left digit effect
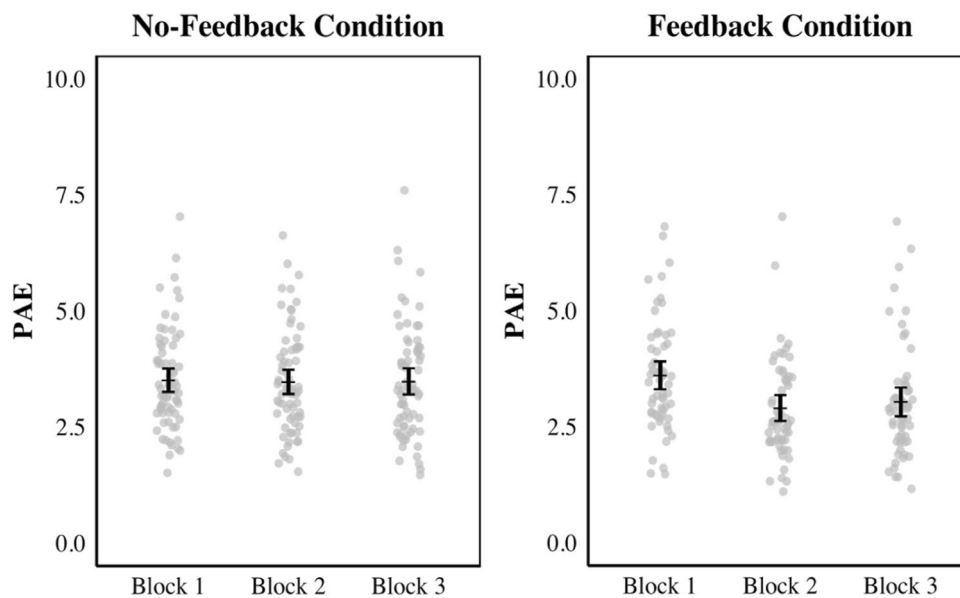

Fig. 7. Percent absolute error (PAE) (by condition and block) in Experiment 2. PAE values are percentages. The error bars reflect ±1 SE from the mean. Larger PAE indicates greater accuracy error on the task

(block: 1, 2, 3). If competitive feedback leads to a sustained reduction in the left digit effect, a condition by block interaction should arise for hundreds difference scores, but we found none, $F(2, 258) = 0.12$, $MSE = 249.23$, $p = .888$. There was also no main effect of block, $F(2, 258) = 2.09$, $MSE = 249.23$, $p = .126$, or condition, $F(1, 129) = 0.30$, $MSE = 486.14$, $p = .586$ (see Fig. 6), consistent with Experiment 1. If competitive feedback

leads to better overall accuracy on the task, there should be a condition by block interaction for PAE. In this case, as in Experiment 1, an interaction was found, $F(2, 258) = 13.89$, $MSE < 0.01$, $p < .001$, $\eta_p^2 = .10$ (see Fig. 7). As shown in Table 2, PAE generally decreased across blocks in the feedback condition but did not change in the no-feedback condition. There was also a main effect of block, $F(2, 258) = 19.49$, $MSE < 0.01$, $p < .001$, $\eta_p^2 =$

Table 3 Hundreds difference score by pair boundary

Pair boundary							
200	300	400	500	600	700	800	900
8.12 (46.27)	6.44 (61.04)	40.57 (58.00)	9.02 (19.63)	27.63 (46.91)	25.62 (61.44)	17.91 (50.00)	16.03 (33.83)

Values reflect average hundreds difference scores; *SD*s are in parentheses.

.13. Unlike Experiment 1, there was no main effect of condition on PAE, $F(1, 129) = 2.80$, $MSE < 0.01$, $p = .097$.

In sum, in Experiment 2 we again found a large left digit effect in all blocks of both conditions. This finding replicates the findings of Experiment 1. Specifically, we found that summary accuracy feedback (this time within a competitive game context) led to improvements in overall accuracy but did not lead to reductions in the left digit effect. We conclude that while increased task effort can produce improvements in overall accuracy, it does not itself reduce the left digit effect. We discuss these findings further in the *General discussion.* Before doing so, we present exploratory analyses of trial-related predictors of the left digit effect.

Combined trial-based exploratory analyses

All participants whose data were analyzed in Experiments 1 and 2 ($N = 277$) were included in the present analyses. The purpose of these analyses was to evaluate possible task-related predictors of the hundreds difference score at the individual-trial level. We considered three predictors: pair boundary, pair distance, and pair order. Pair boundary refers to the boundary for each hundreds pair (e.g., coded as 1 = 200 boundary, 2 = 300 boundary, etc.); each participant was shown one pair of each type (eight total). For each hundreds pair for each participant, we also computed a pair distance and pair order. We did this by subtracting the trial number of the smaller target numeral from the trial number of the larger target numeral. The absolute value of the difference is the pair distance, and the direction (positive or negative) of the difference is the pair order. If pair order is positive (coded as 1), the smaller numeral was presented before the larger numeral; if it is negative (coded as -1), the larger numeral was presented before the smaller numeral. The pair distance and pair order of each pair differed for each participant because trials were presented in a different randomized order to each participant. In analyses, pair boundary and order are treated as categorical variables, and pair distance as a scale variable. We mean-centered pair distance for use in modeling (but use untransformed values for descriptive statistics).

To analyze the data, we used a linear mixed effects model in SPSS, with a maximum likelihood (ML) method and Satterthwaite approximation of degrees of freedom.

The hundreds difference score was entered as the dependent variable. Pair boundary, pair distance, and pair order were entered as fixed effects, and participant was entered as a random effect. The primary model tested was a random intercept model (i.e., a different y-intercept for each participant but the same slope coefficients). We used fixed effects *F*-values to identify statistically significant predictors. Pair boundary was found to be a significant predictor of hundreds difference score, $F(7, 1830) = 15.55$, $p < .001$, as was pair order, $F(1, 2069) = 7.65$, $p = .006$, but not pair distance, $F(1, 2066) < 0.01$, $p = .914$. The percentage of variance explained by participant was 3.02%. Descriptively, pairs surrounding 200, 300, and 500 boundaries were associated with the smallest hundreds difference scores, while those surrounding 400, 600, and 700 boundaries were associated with the largest scores (consistent with Williams et al., in press; see Table 3). For pair order, the hundreds difference score was greater when the smaller numeral was presented before the larger numeral ($M = 21.59$, $SD = 51.49$) rather than the reverse ($M = 16.21$, $SD = 48.59$).

General discussion

In two experiments, we tested whether receiving motivational feedback might reduce or eliminate the left digit effect in number line estimation. In Experiment 1, we asked whether receiving summary accuracy feedback (every 20 trials during the middle block), combined with instructions to attend equally to all three digits, would reduce or eliminate the left digit effect. In Experiment 2, we expanded the intervention to include a competitive goal, namely, one of trying to surpass the summary accuracy feedback scores of past high scorers. In both experiments, feedback did not reduce the left digit effect but did lead to modest improvements in overall accuracy. Additionally, in exploratory analyses, we found that the size of the left digit effect depended on the hundreds pair (it was larger for some pairs over others), and the order of presentation of targets numerals in the pair (it was larger when the lower target was presented first) but did not depend on the number of intervening trials. These experiments support the conclusion that the left digit effect is a robust phenomenon that emerges even when motivation to perform the task accurately is high, and that the left digit

effect observed in past studies is unlikely to simply be the result of insufficient task effort.

Although they did not impact the left digit effect, the feedback interventions did lead to reliable reductions in overall accuracy error. Previously, Eyler et al. (2018) found that trial-by-trial feedback in a brief number line estimation task led to reduced accuracy error, but this was only for a subset of participants who had no college education. The present findings build on this work in showing an intervention's effectiveness even for individuals with some college education, and even when no direct feedback about target locations is given. The findings are also consistent with past studies using other types of magnitude estimation tasks showing that low motivation and task inattention contribute to error (Lindskog et al., 2013; Peters & Bjalkebring, 2015). Because number line estimation tasks are frequently used to measure an individual's estimation skill and to predict future math achievement (especially PAE; see Schneider et al., 2018), it is important to understand the extent to which task effort contributes to performance. The present findings indicate that overall accuracy in number line estimation may vary with task effort (PAE changed ~15% pre- vs. post-intervention), and may be more strongly affected than the left digit effect (which did not change here).

One might ask whether the different patterns of findings for the two dependent measures in the present work might be due in part to the nature of the feedback given. The feedback score was a variant of PAE, so it might on the surface not seem surprising that only PAE decreased following feedback. The reasons we did not use a measure of feedback more directly related to the left digit effect were twofold. The first is that the goal was to enhance task effort, including the allocation of attention to each digit in multi-digit targets, not to give specific direction on how to adjust responses. Summary accuracy feedback served this purpose well in that the feedback itself could not be readily used to determine when or how to adjust one's number line placements. Second, it would not have been possible to create feedback tied to the hundreds difference score since there would have been only one or two targets of this type (and likely not paired ones, e.g., 699/703) contributing to each feedback score. This said, it is plausible one could construct a very different study in which only hundreds trials are presented, and in which feedback is derived from hundreds difference scores. Such a study could be used to test whether motivated individuals could learn to reduce the left digit effect in response to feedback specifically about the left digit effect.

What we infer from these studies is that simply trying harder to perform the number line estimation task is insufficient for reducing the left digit effect (even when one is reminded to attend to all digits as a means of improving performance). Some cognitive accounts of the left digit effect suggest that the overweighting of leftmost digit may be subject to strategic control, such as when people use rounding and truncating heuristics (e.g., ignoring cents in pricing; Gabor & Granger, 1964), or when working memory limitations demand narrowing one's focus of attention (as with odometer readings; Lacetera et al., 2012). These accounts do not easily explain the present findings. However, one account that does provide an explanation for the present findings is that the overweighting of the leftmost digit occurs automatically during the process of converting a symbol to a magnitude (perhaps as a result of left to right reading; Thomas & Morwitz, 2005). It would be expected that such an early, automatic cognitive process of symbol-to-magnitude conversion might be inaccessible to strategic efforts to alter the weighting of each digit. This account would explain why motivational interventions intended to increase task effort did not lead to a reduction in the left digit effect.

As an aside, it is not surprising that overall accuracy error was reduced here even though the left digit effect was not, as there are likely many more potential cognitive contributors to overall accuracy error, including mental magnitude representations (e.g., Dehaene et al., 2008; Siegler & Opfer, 2003), proportion judgment skills (e.g., Cohen et al., 2018; Cohen & Blanc-Goldhammer, 2011; Slusser & Barth, 2017), and use of benchmark strategies (e.g., using the midpoint of the number line to guide placements; Sullivan et al., 2011; Peeters et al., 2017). Notably, a highly motivated individual might flexibly use a larger number of reference points to do the task, perhaps identifying reference points at quartiles rather than only at the midpoint of the line. This would be one way people might reduce overall accuracy error without needing to develop more precise magnitude representations or qualitatively different strategies.

One interesting question the present work does not address is whether providing more direct feedback or instruction about the left digit effect itself could lead to a reduction in or elimination of the effect. We have ongoing studies that take a more direct approach to reducing the effect including through trial-by-trial accuracy feedback (Williams et al., in press), and by simply telling people about the effect (Gwiazda et al., 2021). With these studies, we hope to address the question of whether more direct instruction is effective. Note that even if the effect arises from an automatic process, an individual knowledgeable about the effect could develop strategies for reducing the effect. One might imagine that an individual could adjust their initial estimate or magnitude upwards or downwards as needed (i.e., add a correction for assumed error), even if there is no change in one's initial estimates. However, past research has shown that many similar types of judgment biases are surprisingly resistant even to direct interventions, such as in the case of anchoring and adjustment, a number-related bias in which estimates are influenced by irrelevant primes

(Hoch & Loewenstein, 1989; Pulford & Colman, 1997). We suspect that the left digit effect may turn out to be similarly resistant to direct interventions, especially if it does arise from an automatic symbol-to-magnitude conversion process.

Exploratory analyses were used here to try to learn more about contributors to the left digit effect. Two novel findings were that the left digit effect was greater when the smaller target in a pair preceded the larger one, and that the number of intervening trials did not matter. To make sense of the order effect, we considered that the left digit effect might be driven to a greater extent by the misplacement of smaller targets than by that of larger ones (see also Rubaltelli et al., 2021). Using data from the first block of both studies, we found that this was the case: the smaller (below-boundary) targets were underestimated by 19.94 units while the larger (above-boundary) targets were underestimated by only 3.74 units. Returning to the question of order effects, we speculate that smaller targets might be underestimated more when they are presented first because these placements cannot be guided by those of the larger targets (which are generally more accurate and might essentially serve as reference points). It is also possible that a minimal task practice (i.e., being early in the trial sequence in general) has a larger impact on the placements of smaller relative to larger targets in pairs. Further research is needed in order to replicate these exploratory findings, perhaps in an experimental context.

As mentioned earlier, overall accuracy in number line estimation has been used to predict math achievement in children (see Schneider et al., 2018) and number-based decision making in adults (Patalano et al., 2020; Schley & Peters, 2014). Only one study to date has addressed whether the left digit effect is also related to mathematical competence (Williams et al., 2020). Individuals with a smaller left digit effect had a higher verbal SAT score (but not math score), and this was found only for the subset of participants doing a speeded version of the task. In the exploratory analyses here, we found that only 3% of variation in the left digit effect on individual pairs of trials could be attributed to the individual, with some variance attributed to other variables such as pair boundary and pair order. The findings suggest care in future studies when comparing the size of the left digit effect across individuals who have not performed the same version of the task (e.g., who are given different targets, target order, etc.). They also suggest that it is important to devote future attention to whether it makes sense to treat the size of the left digit effect as an individual difference measure (e.g., by assessing test-retest reliability).

In conclusion, the present studies offer strong evidence that the left digit effect is robust and that previous findings of a left digit effect are not due to a lack of motivation or insufficient task effort alone. This work builds on past findings showing a strong left digit effect in number line estimation in adults and children (Lai et al., 2018; Williams et al.,

2020). Beyond number line estimation, the left digit bias has been shown to be of consequence for decisions ranging from whether or not a surgeon sends a patient to surgery (Olenski et al., 2020) to whether a student chooses to retake a standardized test (e.g., SAT test; Goodman et al., 2020), to purchases of cars (Lacetera et al., 2012) and stocks (Bhattacharya et al., 2012), to driving speed (Rubaltelli et al., 2021). To reduce bias-related errors in judgment, some have even restricted the use of misleading numbers, such as in consumer pricing in Israel (Davidovich-Weisberg, 2013). Given that the left digit effect has significant individual and societal consequences, it remains important to work to understand when and why the effect arises, and to strive to reduce bias in the interpretation of numerals.

Open Practices Statement

Experiment 2 was preregistered at https://www.aspredicted. org/kx2be.pdf. The preregistration document has one error: "A fifties difference score will be calculated by taking the mean of the individual difference scores for 10 [it should read '9'] pairs of numerals…" This is a typographical error and does not reflect a change in planned procedure. Data collected in Experiments 1 and 2 for use in planned analyses are available at https://www.osf.io/qn5hb/.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13421-022-01278-2.

References

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science*, *17*(5), 775–785. https://doi.org/10.1111/desc.12155

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

Beracha, E., & Seiler, M. J. (2015). The effect of pricing strategy on home selection and transaction prices: An investigation of the left-most digit effect. *Journal of Housing Research*, *24*(2), 147–161.

Bhattacharya, U., Holden, C. W., & Jacobsen, S. (2012). Penny wise, dollar foolish: Buy–sell imbalances on and around round numbers. *Management Science,* *58*(2), 413–431. https://doi.org/10.1287/mnsc.1110.1364

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

Brez, C. C., Miller, A. D., & Ramirez, E. M. (2016). Numerical estimation in children for both positive and negative numbers. *Journal of Cognition and Development*, *17*(2), 341–358. https://doi.org/10.1080/15248372.2015.1033525

Burguillo, J. C. (2010). Using game theory and competition-based learning to stimulate student motivation and performance. *Computers & Education*, *55*(2), 566–575. https://doi.org/10.1016/j.compedu.2010.02.018

Cagiltay, N. E., Ozcelik, E., & Ozcelik, N. S. (2015). The effect of competition on learning in games. *Computers & Education*, *87*, 35–41. https://doi.org/10.1016/j.compedu.2015.04.001

Chen, C.-H., Law, V., & Chen, W.-Y. (2018). The effects of peer competition-based science learning game on secondary students' performance, achievement goals, and perceived ability. *Interactive Learning Environments*, *26*(2), 235–244. https://doi.org/10.1080/10494820.2017.1300776

Choi, J., Jessica Li, Y., & Samper, A. (2019). The influence of health motivation and calorie ending on preferences for indulgent foods. *Journal of Consumer Research*, *46*(3), 606–619. https://doi.org/10.1093/jcr/ucz002

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

Cohen, D. J., Blanc-Goldhammer, D., & Quinlan, P. T. (2018). A mathematical model of how people solve most variants of the number-line task. *Cognitive Science*, *42*(8), 2621-2647. https://doi.org/10.1111/cogs.12698

Davidovich-Weisberg, G. (2013). Israel to abolish deceptive pricing ending in .99 shekels. *Haaretz* (October 20), https://www.haaretz.com/israel-news/business/.premium-israeli-shoppers-no-more-nis-9-99-1.5276194

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 626–641. https://doi.org/10.1037/0096-1523.16.3.626

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, *320*(5880), 1217–1220. https://doi.org/10.1126/science.1156540

De Wind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience, 6*, 68. https://doi.org/10.3389/fnhum.2012.00068

Eyler, R. F., Cordes, S., Szymanski, B. R., & Fraenkel, L. (2018). Use of feedback to improve mental number line representations in primary care clinics. *BMC Medical Informatics and Decision Making*, *18*(1), 40. https://doi.org/10.1186/s12911-018-0618-6

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*(6), 1539–1552. https://doi.org/10.1037/a0025510

Goodman, J., Gurantz, O., & Smith, J. (2020). Take two! SAT retaking and college enrollment gaps. *American Economic Journal: Economic Policy*, *12*(2), 115-58. https://doi.org/10.1257/pol.20170503

Gwiazda, G., Bondhus, C., Kayton, K., Barth, H., & Patalano, A.L. (2021). Telling people about the left digit effect in number line estimation does not reduce the effect [Unpublished manuscript]. Department of Psychology, Wesleyan University.

Hamdan, N., & Gunderson, E. A. (2017). The number line is a critical spatial-numerical representation: Evidence from a fraction intervention. *Developmental Psychology, 53*(3), 587–596. https://doi.org/10.1037/dev0000252

Hansen, N., Jordan, N. C., Fernandez, E., Siegler, R. S., Fuchs, L., Gersten, R., & Micklos, D. (2015). General and math-specific predictors of sixth-graders' knowledge of fractions. *Cognitive Development*, *35*, 34–49. https://doi.org/10.1016/j.cogdev.2015.02.001

Henninger, F., Shevchenko, Y., Mertens, U., Kieslich, P. J., & Hilbig, B. E. (2019). lab.js: A free, open, online study builder. PsyArXiv. https://doi.org/10.31234/osf.io/fqr49

Hoch, S. J., & Loewenstein, G. F. (1989). Outcome feedback: hindsight and information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 605–619. https://doi.org/10.1037/0278-7393.15.4.605

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

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

Hornung, C., Schiltz, C., Brunner, M., & Martin, R. (2014). Predicting first-grade mathematics achievement: The contributions of domain-general cognitive abilities, nonverbal number sense, and early number competence. *Frontiers in Psychology*, *5*, 272. https://doi.org/10.3389/fpsyg.2014.00272

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology*, *116*(1), 45–58. https://doi.org/10.1016/j.jecp.2013.02.001

Lacetera, N., Pope, D. G., & Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review*, *102*(5), 2206–36. https://doi.org/10.1257/aer.102.5.2206

Lai, M., Zax, A., & Barth, H. (2018). Digit identity influences numerical estimation in children and adults. *Developmental Science*, *21*(5), e12657. https://doi.org/10.1111/desc.12657

Lin, C. H., & Wang, J. W. (2017). Distortion of price discount perceptions through the left-digit effect. *Marketing Letters*, *28*(1), 99-112. https://doi.org/10.1007/s11002-015-9387-5

Lindskog, M., Winman, A., & Juslin, P. (2013). Are there rapid feedback effects on approximate number system acuity? *Frontiers in Human Neuroscience*, *7*, 270. https://doi.org/10.3389/fnhum.2013.00270

MacKillop, J., Amlung, M. T., Blackburn, A., Murphy, J. G., Carrigan, M., Carpenter, M. J., & Chaloupka, F. (2014). Left-digit price effects on smoking cessation motivation. *Tobacco Control*, *23*(6), 501–506. https://doi.org/10.1136/tobaccocontrol-2012-050943

Manning, K. C., & Sprott, D. E. (2009). Price endings, left-digit effects, and choice. *Journal of Consumer Research*, *36*(2), 328-335. https://doi.org/10.1086/597215

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology*, *103*(4), 503–515. https://doi.org/10.1016/j.jecp.2009.02.006

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520. https://doi.org/10.1038/2151519a0

Nuerk, H.-C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2011). Extending the mental number line: A review of multi-digit number processing. *Zeitschrift für Psychologie/Journal of Psychology, 219*(1), 3–22. https://doi.org/10.1027/2151-2604/a000041

Olenski, A. R., Zimerman, A., Coussens, S., & Jena, A. B. (2020). Behavioral heuristics in coronary-artery bypass graft surgery. *New England Journal of Medicine*, *382*(8), 778–779. https://doi.org/10.1056/NEJMc1911289

Östergren, R., & Träff, U. (2013). Early number knowledge and cognitive ability affect early arithmetic ability. *Journal of Experimental Child Psychology*, *115*(3), 405–421. https://doi.org/10.1016/j.jecp.2013.03.007

Patalano, A. L., Zax, A., Williams, K., Mathias, L., Cordes, S., & Barth, H. (2020). Intuitive symbolic magnitude judgments and decision making under risk in adults. *Cognitive Psychology*, *118*, 101273. https://doi.org/10.1016/j.cogpsych.2020.101273

Patalano, A. L., Williams, K., Weeks, G., Kayton, K., & Barth, H. (2022). The left digit effect in a complex judgment task: Evaluating hypothetical college applicants. *Journal of Behavioral Decision Making, 5*(1), e2247. https://doi.org/10.1002/bdm.2247

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

Peters, E., & Bjalkebring, P. (2015). Multiple numeric competencies: When a number is not just a number. *Journal of Personality and Social Psychology*, *108*(5), 802–822. https://doi.org/10.1037/pspp0000019

Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction*, *7*(1), 55–78. https://doi.org/10.1207/s1532690xci0701_3

Pulford, B. D., & Colman, A. M. (1997). Overconfidence: Feedback and item difficulty effects. *Personality and Individual Differences*, *23*(1), 125–133. https://doi.org/10.1016/S0191-8869(97)00028-7

Rubaltelli, E., Manicardi, D., Orsini, F., Mulatti, C., Rossi, R., & Lotto, L. (2021). How to nudge drivers to reduce speed: the case of the left-digit effect. *Transportation Research Part F: Traffic Psychology and Behaviour*, *78*, 259–266. https://doi.org/10.1016/j.trf.2021.02.018

Savelkouls, S., Williams, K., & Barth, H. (2020). Linguistic inversion and numerical estimation. *Journal of Numerical Cognition*, *6*(3), 263-274. https://doi.org/10.5964/jnc.v6i3.273

Schley, D. R., & Peters, E. (2014). Assessing "economic value" symbolic-number mappings predict risky and riskless valuations. *Psychological Science*, *25*(3), 753–761. https://doi.org/10.1177/0956797613515485

Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: their interrelations in grades 5 and 6. *Journal of Educational Psychology*, *101*(2), 359–372. https://doi.org/10.1037/a0013840

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. https://doi.org/10.1111/cdev.13068

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

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

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*(3), 545–560. https://doi.org/10.1037/a0014239

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education*, *3*(3), 143-150. https://doi.org/10.1111/j.1751-228X.2009.01064.x

Slusser, E., & Barth, H. (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. https://doi.org/10.1037/a0028560

Sokolova, T., Seenivasan, S., & Thomas, M. (2020). The left-digit bias When and why are consumers penny-wise and pound foolish? *Journal of Marketing Research, 57*(4), 771-788. https://doi.org/10.1177/0022243720932532

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. https://doi.org/10.3758/s13423-011-0081-1

Thomas, M., & Morwitz, V. (2005). Penny wise and pound foolish: The left-digit effect in price cognition. *Journal of Consumer Research*, *32*(1), 54–64. https://doi.org/10.1086/429600

Thomas, M., & Morwitz, V. (2009). Heuristics in numerical cognition: Implications for pricing. In V. Rao (Ed.), *Handbook of Pricing Research in Marketing* (pp. 132–149), Edward Elgar Publishing. https://doi.org/10.4337/9781848447448

Tosto, M. G., Petrill, S. A., Malykh, S., Malki, K., Haworth, C., Mazzocco, M. M., ... & Kovas, Y. (2017). Number sense and mathematics: Which, when and how? *Developmental Psychology*, *53*(10), 1924–1939. https://doi.org/10.1037/dev0000331

Vaidya, L., Eisen, S., Zax, A., Williams, K., Patalano, A.L., & Barth, H. (2022). Left digit effects in number line estimation are robust across task format [Unpublished manuscript]. Department of Psychology, Wesleyan University.

Verguts, T., & De Moor, W. (2005). Two-digit comparison: Decomposed, holistic, or hybrid? *Experimental Psychology*, *52*(3), 195–200. https://doi.org/10.1027/1618-3169.52.3.195

Williams, K., Paul, J., Zax, A., Barth, H., & Patalano, A.L. (2020). Number line estimation and standardized test performance: The left digit effect does not predict SAT math score. *Brain & Behavior, 10*(12):*e01877*. https://doi.org/10.1002/brb3.1877

Williams, K., Zax, A., Patalano, A.L., Barth, H. (in press). Left digit effects in numerical cognition across development. *Journal of Cognition and Development*, 1-22.

Williams, K., Bradley, K., Xing, C., Barth, H., Patalano, A.L. (2021). *Potential moderators of the left digit effect in numerical estimation* [Unpublished manuscript]. Department of Psychology, Wesleyan University.

Xing, C., Zax, A., George, E., Taggart, J., Bass, I., George, E., & Barth, H. (2021). *Numerical estimation strategies are correlated with math ability in school-age children* [Manuscript submitted for publication]. Department of Psychology, Wesleyan University.

Zax, A., Williams, K., Patalano, A. L., Slusser, E., Cordes, S., & Barth, H. (2019). What do biased estimates tell us about cognitive processing? Spatial judgments as proportion estimation. *Journal of Cognition and Development*, *20*(5), 702–728. https://doi.org/10.1080/15248372.2019.1653297

Zhang, H., & Maloney, L.Y. (2012). Ubiquitous log odds: A common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience*, *6*, 1. https://doi.org/10.3389/fnins.2012.00001

Zhu, M., Cai, D., & Leung, A. W. S. (2017). Number line estimation predicts mathematical skills: Difference in grades 2 and 4. *Frontiers in Psychology*, *8*, 1576. https://doi.org/10.3389/fpsyg.2017.01576