Received: 2 June 2020 | Revised: 5 March 2021 | Accepted: 5 March 2021
DOI: 10.1002/bdm.2247

RESEARCH ARTICLE

WILEY

The left digit effect in a complex judgment task: Evaluating hypothetical college applicants

Andrea L. Patalano[1] | Katherine Williams[1,2] | Gillian Weeks[1] | Kelsey Kayton[1] | Hilary Barth[1]

[1]Department of Psychology, Wesleyan University, Middletown, Connecticut, USA
[2]Department of Psychology, University of Texas at Austin, Austin, Texas, USA

Correspondence
Andrea L. Patalano, Department of Psychology, Wesleyan University, 207 High Street, Middletown, CT 06459, USA.
Email: apatalano@wesleyan.edu

Present address
Katherine Williams, Department of Psychology, University of Texas at Austin, Austin, Texas, USA

Funding information
National Science Foundation, Division of Research on Learning in Formal and Informal Settings, Grant/Award Numbers: DRL-1561214, DRL-1920445

Abstract

A left digit effect has been broadly observed across judgment and decision-making contexts ranging from product evaluation to medical treatment decisions to number line estimation. For example, $3.00 is judged to be a much greater cost than $2.99, and “801” is estimated strikingly too far to the right of “798” on a number line. Although the consequences of the effects for judgment and decision behavior have been documented, the sources of the effects are not well established. The goal of the current work is to extend investigations of the left digit effect to a new complex judgment activity and to assess whether the magnitude of the effect at the individual level can be predicted from performance on a simpler number skills task on which the left digit effect has also recently been observed. In three experiments (*N* = 434), adults completed a judgment task in which they rated the strength of hypothetical applicants for college admission and a self-paced number line estimation task. In all experiments, a small or medium left digit effect was found in the college admissions task, and a large effect was found in number line estimation. Individual-level variation was observed, but there was no relationship between the magnitudes of the effects in the two tasks. These findings provide evidence of a left digit effect in a novel multiattribute judgment task but offer no evidence that such performance can be predicted from a simple number skills task such as number line estimation.

KEYWORDS
judgment, left digit effect, number line estimation, numerical cognition

1 | INTRODUCTION

We frequently need to interpret numbers used in judgment and decision making (e.g., Is an 80% chance very likely? Is a cost of $3.20 a lot more than $2.75?). It is well documented that people often have difficulty using numerical information, as illustrated by a wide range of decision phenomena including denominator neglect (Reyna & Brainerd, 2008), base rate neglect (Bar-Hillel, 1980; Kahneman & Tversky, 1973), framing effects (Levin & Gaeth, 1988; Tversky & Kahneman, 1981), anchoring plus insufficient adjustment (Epley & Gilovich, 2001; Tversky & Kahneman, 1974), and probability distortion (Gonzalez & Wu, 1999; Tversky & Kahneman, 1992), to name just a few. There is growing reason to believe that such phenomena may have underpinnings in mental processes associated with numerical cognition. This prospect has led to recent research into the relationship between judgment biases and numerical cognition, including whether individual differences in biases might be predicted from foundational number skills (Patalano et al., 2020; Peters & Bjalkebring, 2015; Schley & Peters, 2014). The current work focuses on a phenomenon called the *left digit effect* or *left-digit bias*. The goals

This work includes number line estimation scores also used in Williams, Paul, et al. (2020). The second and third experiments reported here were preregistered (at https://aspredicted.org/sx7ws.pdf for Exp. 2, and https://aspredicted.org/ym86k.pdf for Exp. 3).

of the work are to extend the test of the left digit effect to a new complex judgment task and to assess whether the size of the effect at the individual level can be predicted from one's left digit effect on a more basic number skills task where the effect has also recently been demonstrated.

2 | LEFT DIGIT EFFECT IN JUDGMENT AND DECISION MAKING

The left digit effect in judgment and decision making refers to the phenomenon whereby decision-related numbers that cross a left digit boundary (e.g., 599 vs. 601, on either side of 600) are treated as more different than is warranted while those that do not cross a left-digit boundary (e.g., 603 and 605) are not. The term is attributed to Thomas and Morwitz (2005) who were motivated by the common belief that people underestimate the cost of products with prices ending in .99. They found that prices are judged as more different in magnitude when they have a different leftmost digit than when they do not. For example, a pen costing $3.00 is judged to be more expensive than one costing $2.99, while pens costing $3.60 and $3.59 are judged to cost about the same (Anderson & Simester, 2003; Lin & Wang, 2017; Manning & Sprott, 2009; Schindler & Kibarian, 1996; Thomas & Morwitz, 2009). The effect has been extended to smokers' self-reported higher likelihood of making a cessation attempt as increases in cigarette costs cross a left-digit boundary (e.g., from $5.80 to $6.00 per pack; MacKillop et al., 2014); to health-motivated consumers being more likely to buy and consume foods with nine-ending calorie counts than those with round number counts (Choi et al., 2019); and to evaluations of products based on consumer product ratings (Thomas & Morwitz, 2005).

In addition to laboratory studies, studies drawing on existing records have been used to demonstrate the left digit effect. In particular, home sales final transaction prices were found to be higher (relative to market value) for homes initially listed just below (e.g., $599,995) versus at a whole number (e.g., $600,000; Beracha & Seiler, 2015); discontinuous changes in selling prices were observed for used cars with odometer readings on either side of a left-digit boundary (Lacetera et al., 2012); and lower public assessments of school performance were given for schools with student grade averages just below a left-digit boundary (Olsen, 2013). Most recently, the left digit effect was strikingly demonstrated in physicians' treatment decisions based on evaluation of patients' ages (Olenski et al., 2020). In an analysis of several years of Medicare records, physicians were found more likely to recommend heart surgery for patients whose birthdates indicated they were 2 weeks younger than 80 years of age versus 2 weeks older. The physicians did not recommend surgery at different rates for patients of different ages that did not cross a decade boundary (e.g., 2 weeks younger vs. older than 78).

The source of the left digit effect has not been well established. For many years, the effect was attributed to domain specific heuristics such as the rounding of dollar values (Gabor & Granger, 1964). More recently, in the context of odometer readings, an attention-based heuristic was proposed whereby one devotes increasingly less attention as one moves rightward in a string of digits due to cognitive limits in information processing (Lacetera et al., 2012). Others have suggested that rather than being specific to decision making, the left digit bias might be related to the processing of numerical magnitudes more generally (Macizo & Ojedo, 2018; Thomas & Morwitz, 2005). This would mean that the bias, rather than arising during the translation of a decision-related numeral into a subjective rating (e.g., of costliness), arises even when the goal is simple magnitude estimation (e.g., place "570" on a number line). However, at the time this possibility was suggested, there was no known evidence of a left digit effect in magnitude estimation (but see Nuerk et al., 2004, for numerical comparison tasks).

3 | LEFT DIGIT EFFECT IN NUMBER LINE ESTIMATION

Skill in estimating numerical magnitude is central to many types of thinking. A numerical magnitude estimation task that has been used in several hundred studies as a skills assessment tool, an educational training tool, and a probe of underlying cognitive processes is the *number line estimation (NLE)* task. A standard version of the task involves estimating the location of a target numeral (e.g., "540") on a line labeled only at its endpoints (e.g., 0 and 1000), as in Figure 1. A common measure of an individual's performance on the task is overall accuracy, often reported as *percent absolute error (PAE)*, which is computed as $|actual\ placement - correct\ location|/range\ of\ target\ values * 100$, averaged over all trials. Extensive work reveals that overall accuracy in children and adolescents predicts diverse numerical skills at later ages, including standardized math achievement (Booth & Siegler, 2008; Hoffmann et al., 2013; Holloway & Ansari, 2009; Schneider et al., 2009, 2018). NLE accuracy in adults is also related to quantitative skills including more precise use of numbers in decision making (Patalano et al., 2020; Peters & Bjalkebring, 2015; Schley &

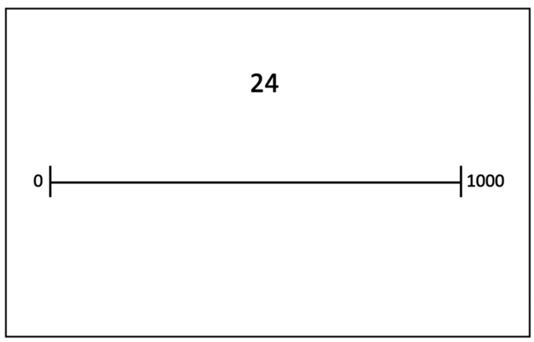

FIGURE 1 Example of a single trial of the number line estimation task. Note: Participants indicate the location of a target number ("24" here) on the line with a mouse click. A red line appears in the selected location to indicate that a response was recorded. A new target number appears in the same position above the line on each trial

Peters, 2014), and so the NLE task might become an important tool for predicting, understanding, and training decision-related skills.

NLE error is not unsystematic. Though not the focus of the present work, there is one pattern of bias in NLE that has received much attention. In its simplest form, it is that adults tend to overestimate target locations on one half of the line (responding too far to the right) and to underestimate target locations on the other half of the line (responding too far to the left). This pattern has been modeled as an inverse-S-shaped or an S-shaped curve, with a parameter estimate β indexing degree and direction of bias (e.g., Cohen & Blanc-Goldhammer, 2011; Slusser & Barth, 2017). Interestingly, rather than being specific to NLE, this bias has been found across a wide range of tasks that can be described as involving proportion judgment (see Hollands & Dyre, 2000; Zax et al., 2019; Zhang & Maloney, 2012, for reviews). Additionally, the pattern of bias is sometimes multicyclical, consistent with the use of intermediate reference points to divide a whole into smaller parts (Cohen & Blanc-Goldhammer, 2011; Slusser & Barth, 2017; Zax et al., 2019). This proportion judgment bias was, until recently, the primary known source of systematic error on the NLE task.

Recently, however, another type of systematic error has been identified: a robust left digit effect (Lai et al., 2018; Williams, Paul, et al., 2020) broadly similar to that seen in judgment and decision-making tasks. Lai et al. discovered that three-digit numbers of similar magnitude but different leftmost hundreds-place digits (e.g., 597 and 601) are placed in very different locations on a number line even though the numbers should be placed in approximately the same location. In contrast, numbers with different tens-place digits (falling on either side of a fifties boundary; e.g., 248 and 252; also four digits apart) are placed in the same location, suggesting that the phenomenon extends only to the leftmost digit (at least for three-digit numbers). This left digit effect in NLE is very large (*d*s > 1), occurs for both speeded and self-paced tasks, and occurs even when numbers with similar magnitudes are separated by many intervening trials (Williams, Paul, et al., 2020). There is also noticeable individual variation in that some individuals demonstrate a larger left digit effect than others. Collectively, these findings offer the first evidence of a left digit effect in numerical magnitude estimation and raise the likelihood that the left digit effect in decision making may arise from cognitive processes associated with more basic number skills given that it arises even in NLE tasks.

4 | POSSIBLE RELATIONSHIP BETWEEN LEFT DIGIT EFFECT MEASURES

The left digit effect in NLE is important in its own right as a novel bias in symbolic number processing and because the NLE task, given its simplicity, can be used to better understand why the effect arises and how to reduce it. A further question that arises in light of the discovery of the left digit effect in NLE is whether it is related to the left digit effect observed in judgment tasks of the type described earlier (pricing, medical decisions, product evaluations, etc.). If similar mental processes contribute to the left digit effect across tasks, and given that we know there is large individual-level variation in the magnitude of the left digit effect in NLE, we might expect that the size of one's left digit effect in NLE would predict the size of the same individual's left digit effect in judgment. A correlation in performance on the two tasks would provide suggestive evidence of a common underlying mechanism across tasks and would suggest that more complex judgments arise from more basic processes of numerical cognition. Such a correlation would also suggest that a NLE task could potentially be used to predict the size of one's left digit effect in a wide range of judgments or be used to train reductions in left-digit bias. For these reasons, it is important to begin to understand whether there is any relationship between measures of the left digit effect across judgment tasks.

The research approach used here is similar to that which led to a recent discovery of a different relationship between biases in NLE and decision making. It is well known that the interpretation of probabilities in decision making can be modeled as an inverse S-shaped curve: People overweight small probabilities and underweight large ones (Tversky & Kahneman, 1992). This bias had for many years been attributed to a range of decision-related cognitive processes from mental simulations of outcomes (Hogarth & Einhorn, 1990) to affective responses (Rottenstreich & Hsee, 2001) but not to more basic processes related to quantitative cognition. However, recently, the inverse S-shaped pattern of bias in interpreting probabilities was observed to be similar to that in tasks of proportion judgment such as NLE (as described earlier). When bias measures were collected on both tasks, the degree of one's bias in interpreting probabilities was in fact correlated with one's bias in NLE. The findings suggest proportion judgment as one explanation for both biases and NLE training as a possible way to reduce bias in decision making (Patalano et al., 2020). The present work uses a similar individual-variation, correlational approach (see also e.g., Peters & Bjalkebring, 2015; Schley & Peters, 2014, for other uses of this approach), but with a focus now on common biases in the interpretation of numerals on either side of left-digit boundaries.

5 | OVERVIEW OF EXPERIMENTS

One goal of the three experiments reported here is to test whether a left digit effect emerges in a new multiattribute judgment task: a laboratory-based “college admissions” task. We developed this task in part simply as a means of assessing theleft digit effect in judgment in the laboratory and also because past laboratory studies have primarily used single-attribute tasks involving prices, and a goal here was to generalize to a more complex task that could be conducted in the laboratory. The task we developed involves reviewing 20 hypothetical college applicants' portfolios and rating the strength of each candidate for admission. Each portfolio consists of five dimensions with numerical values: grade point average (GPA), quality of letters of reference, quality of extracurricular activities, SAT verbal score, and SAT math score. We considered the possibility that, given that participants would not have previous task experience and would likely use attributes inconsistently, any small differences in the interpretations of attribute values might have little observable influence on ratings. However, based on past work, we predicted that the left digit effect would emerge in this new judgment task.

In the college admissions task, four critical portfolios given to each participant were used to create a measure of one's left digit effect: a *boundary pair* and a *non-boundary* pair (as shown in Table 1). The two portfolios in the boundary pair (low and high boundary portfolios) have identical values on all dimensions except they diverge slightly across a left-digit boundary on SAT verbal scores (599 vs. 601; or on both SAT verbal and math scores in Experiments 2 and 3). Similarly, the two portfolios in the non-boundary pair (low and high non-boundary portfolios) have identical values on all dimensions except they diverge slightly on SAT verbal scores *not* around a left-digit boundary (621 vs. 623; or on both SAT verbal and math scores in Experiments 2 and 3). If ratings of paired applicants are more different when the values of a dimension are on either side of a hundreds boundary (as in the boundary pair) versus when they are not (as in the non-boundary pair), this would be evidence of a left digit effect in this task and would thus extend the bias to a novel, complex judgment task.

With this design, boundary portfolios cannot be directly compared with non-boundary portfolios because the SAT scores near a left-digit boundary are different in magnitude from those farther from such a boundary (and the two types of portfolios differ on other dimensions as well), so we do not conduct a repeated measures analysis of variance (ANOVA). Rather, after reporting the mean rating for each portfolio to show the pattern of means in the sample data, we compute for each participant a *judgment difference score = (rating of high boundary portfolio − rating of low boundary portfolio) − (rating of high non-boundary portfolio − rating of low non-boundary portfolio)*. This measure isolates the pattern of interest and is modeled after one used in past studies to compute the left digit effect in NLE, as will be described shortly. A positive judgment difference score (reliably > 0 across participants, on average, by a one-sample *t* test) would constitute evidence of a left digit effect because it means that the boundary ratings differ more than would be expected based on the difference in the non-boundary ratings.

A second goal of the three experiments is to assess whether there is a relationship between the magnitude of one's left digit effect in the college admissions task and the corresponding measure in a NLE task. After completing the complex judgment task, participants completed a standard self-paced NLE task with numbers ranging from 0 to 1000. The number line task was administered as part of a separate study (unrelated to the present report) investigating the effects of accuracy feedback on performance, so only performance on the first block of the task (120 trials; prior to any feedback manipulation) is reported here. As in past work, we focus 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 is whether the larger value in each pair is 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). Following past work, we compute a measure of the left digit effect called the *hundreds difference score = (placement for larger numeral − placement for smaller numeral)* for each of the eight pairs. We then average the difference scores to create one mean hundreds difference score (which will just be referred to just as the hundreds difference score). A positive hundreds difference score (reliably > 0 across participants, on average, by a one-sample *t* test) would constitute evidence of a left digit effect because it means that the boundary ratings differ more than would be expected if individuals were responding accurately (i.e., if they were placing both values in the same location on the line and thus the difference were 0).

We note that although the judgment difference score in the college admissions task is modeled after the hundreds difference score in the NLE task, the measures differ in two ways. The first is in terms of the baseline used in each task. In the college admissions task, there is no objectively correct difference in how two portfolios that differ by a few SAT points should be rated, so the difference in ratings for non-boundary portfolios is used to establish a baseline. In the NLE task, there is an objective difference in how numbers that are only a few units apart should be placed on the number line. Namely, the numbers should be placed in nearly the same location, so the baseline is 0. The second way the two measures differ from one another is simply that there are fewer trials contributing to the computation of the effect in

TABLE 1 Target stimuli (college applicant portfolios) used in all experiments

						Experiment 1		Experiments 2 and 3	
Portfolio type	**Value type**	**Label**	**GPA**	**Letters**	**Essay**	**SAT-V**	**SAT-M**	**SAT-V**	**SAT-M**
Boundary	Low	DP	3.5	3	3	599	622	699	598
Boundary	High	LD[a]	3.5	3	3	601	622	701	602
Non-boundary	Low	DH	3.6[b]	3	3	621	589	712	611
Non-boundary	High	SW	3.6[b]	3	3	623	589	714	615

Note: Boundary stimuli were identical on all dimensions except SAT-V (Experiment 1) or both SAT-V and SAT-M (Experiments 2 and 3). For SATs, low and high boundary scores were a few points apart, below and above a hundreds boundary (e.g., 599 and 601). The same was true of non-boundary stimuli except that SATs were not near a hundreds boundary (e.g., 621 and 623).
Abbreviation: GPA, grade point average.
[a]Changed in Experiment 2 and Experiment 3 to LG so that there would be no overlapping initials with DP.
[b]Changed in Experiment 2 and Experiment 3 to 3.4.

the college admissions task than in the NLE task. This was intentional in that, because trials of the college admissions task require greater time and effort than those of the NLE task, a shorter task was considered important for sustaining participants' motivation and attention (and is consistent with the use of similar types of tasks in past work; e.g., Peters et al., 2006).

Of focal interest is whether measures of the left digit effect are correlated across tasks. We know of only one existing study on the left digit effect in NLE as a predictor of performance on more complex tasks. In this past study, no correlation was found between the left digit effect in NLE and SAT math score (Williams, Paul, et al., 2020). However, unlike the Williams, Paul, et al. study, the present work specifically uses measures of the left digit effect in both of the tasks rather than a general math score. In other work, Patalano et al. (2020) did find a correlation between curvilinear bias in NLE and in interpretation of probabilities, which speaks to NLE as a predictor of decision-making skills more generally. A power analysis ($1 - \beta = .80$, $\alpha = .05$) indicated samples of $N = 120$ would be needed to detect a moderately small correlation ($r = .25$) between measures here.

6 | EXPERIMENT 1: COLLEGE ADMISSIONS TASK

In Experiment 1, participants completed a college admissions task followed by a NLE task. In the college admissions task, SAT verbal scores were used to create high and low boundary portfolios (SAT verbal 599 vs. 601) and high and low non-boundary portfolios (SAT verbal 621 vs. 623). Using participants' ratings of the four critical portfolios, we computed a judgment difference score. A positive judgment difference score (reliably > 0 across participants, on average) would constitute evidence of a left digit effect. For the NLE task, we computed a hundreds difference score using participants' estimates of eight critical pairs of values falling around hundreds boundaries. We calculated the average difference score across eight pairs, yielding a single hundreds difference score per participant. A positive hundreds difference score (reliably > 0 across participants, on average) would constitute evidence of a left digit effect. A reliable correlation between individual-level judgment difference scores and hundreds differences scores would demonstrate that the NLE task can be used to predict the size of one's left digit effect in decision making and would provide evidence of common cognitive contributors to the effect across tasks. We predicted that we would observe a left digit effect in both tasks and that the magnitude of the left digit effect would be correlated across tasks.

6.1 | Method

6.1.1 | Participants

Adults ($N = 134$ undergraduates; 77 women and 57 men) completed the study in exchange for introductory psychology course credit.[1] Participants came to the lab individually for 1-h sessions, where they completed the college admissions task followed by the NLE task.[2] They were sequentially assigned to a counterbalancing order for the admissions rating task (Order 1 $n = 68$; Order 2 $n = 66$). All studies were approved by the Wesleyan University Institutional Review Board; written informed consent was obtained from all participants.

6.1.2 | Admissions rating task

The task was to review hypothetical portfolios of college applicants and to rate the strength of each applicant for admission. It was programmed using Psychopy software (www.psychopy.org; Peirce et al., 2019). Each portfolio was labeled with the applicant's initials (e.g., "AK") and values for five quantitative dimensions: GPA (1.0–4.0 scale), rated quality of letters of reference (1–3 scale, where 1 is *low*) rated quality of essay (1–3 scale, where 1 is *low*), SAT verbal score[3] (200–800 point scale), and SAT math score (200–800 point scale). See Figure 2 for a sample portfolio (and supporting information for all stimuli). Participants were asked to imagine they were admissions officers at a school that treats all five dimensions as equally important. They were also informed of the range of possible values for each dimension. Ratings of applicants were elicited with a sliding scale (a horizontal bar 8 in. long with an

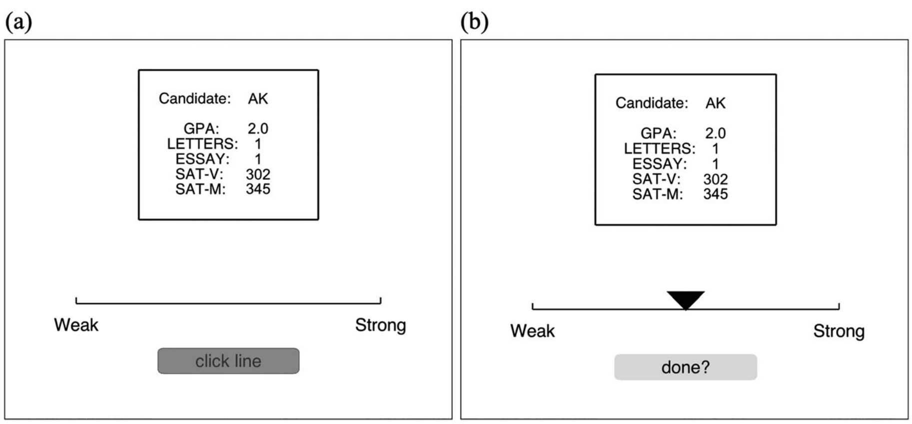

FIGURE 2 Example of a single trial of the college admissions task. Note: Screen images represent (a) the start of a trial and (b) after a participant clicks on the line (the triangle appears wherever the line is first clicked and then can be slid along the line). GPA, grade point average

adjustable slider as in Figure 2) labeled with endpoints "Weak" and "Strong." The triangle slider shown in Figure 2 appeared after the first click on each trial, in the location the line was clicked, and then could be slid to adjust one's response. For analyses, individual responses were translated to a number between 0 and 100.

There were four critical portfolios: a low and a high boundary portfolio and a low and a high non-boundary portfolio (see Table 1). The two boundary portfolios were identical to one another except their SAT verbal scores differed by two points that crossed a hundreds boundary (599 vs. 601). The two non-boundary portfolios were also identical to one another except their SAT verbal scores differed by two points that did *not* cross a hundreds boundary (621 vs. 623). The boundary portfolios had lower SAT verbal scores than the non-boundary ones but higher GPAs and SAT math scores, resulting in four similarly strong portfolios. The remaining 16 portfolios were fillers, created to reflect a varied pool of applicants and to disguise study goals.

Participants saw and rated portfolios one at a time. In Trials 1–6, the same filler portfolios were presented in the same order to all participants. These trials were treated as practice to orient participants to the range of candidates and the task (though this purpose was not revealed to participants). The remaining 14 portfolios (presented in Trials 7–20) were shown in one of two orders. In Order 1, critical portfolios were presented at Trial 7 (low boundary), 10 (low non-boundary), 13 (high boundary), and 19 (high non-boundary). In Order 2, they were reversed (e.g., Trial 7 was a filler, and Trial 8 was the high non-boundary).

6.1.3 | NLE task

This task (Lai et al., 2018) assesses one's ability to identify the approximate locations of numbers on a response line. The task was programmed using Psychopy software. On each trial, participants were presented with a target numeral (e.g., "47") centered above a horizontal line (20 cm wide) with endpoints labeled "0" on the left and "1000" on the right. A total of 120 unique target numerals were used including eight pairs of numerals around hundreds boundaries (16 targets: 199/202, 298/302, 398/403, 499/502, 597/601, 699/703, 798/802, and 899/901); the remaining target numerals were evenly sampled throughout the full 0–1000 range (e.g., 235, 367, 411; see supporting information for full list). Target numerals were presented individually in a different computer-generated random order to each participant.

Participants were seated in front of a computer and given instructions to select a position on the line (with a mouse click) to estimate the location of the given target numeral (as quickly and accurately as possible), a 500-ms pause separated trials. The location of each mouse click was recorded and converted to a number between 0 to 1000, corresponding to the selected location on the line. While participants completed three blocks of 120 trials each, we report only the first block here (later blocks were administered as part of a separate study).[4]

6.2 | Results

6.2.1 | College admissions task

There were no order effects ($t(132) = 0.57$, standard error (SE) = 2.49, $p = .572$). We collapsed over order for further analyses. The mean rating for each of the four types of critical stimuli are shown in Figure 3. The figure reveals that, for the pair of boundary portfolios, participants gave higher ratings to the portfolio with the higher SAT verbal score ($M = 71.40$, standard deviation (SD) = 13.10) relative to the portfolio with the lower score ($M = 70.14$, $SD = 12.30$). For non-boundary portfolios, higher ratings were given by participants to the portfolio with the lower SAT verbal score ($M = 70.87$, $SD = 11.63$) relative to the portfolio with the higher score ($M = 69.43$, $SD = 12.36$). Although neither difference was statistically significant ($|t|$s < 1.60, ps > .100), the focal question of interest is whether the difference in ratings for the boundary portfolios (which are identical except that SAT verbal scores differ by two points across a left-digit boundary) is reliably greater than that for the non-boundary portfolios (which are identical except that SAT verbal scores differ by two points across a non-boundary and thus provide a baseline for comparison). This question cannot be answered with the paired t tests.

To answer this question, we created a *judgment difference score = (rating of high boundary portfolio – rating of low boundary portfolio) – (rating of high non-boundary portfolio – rating of low non-boundary portfolio)* for each participant. On this measure, a value greater than 0 indicates a left digit effect because it means that the difference in ratings for the boundary portfolios is greater than the difference for the non-boundary portfolios. The judgment difference score across participants was normally distributed with a $M = 2.69$ ($SD = 14.38$,

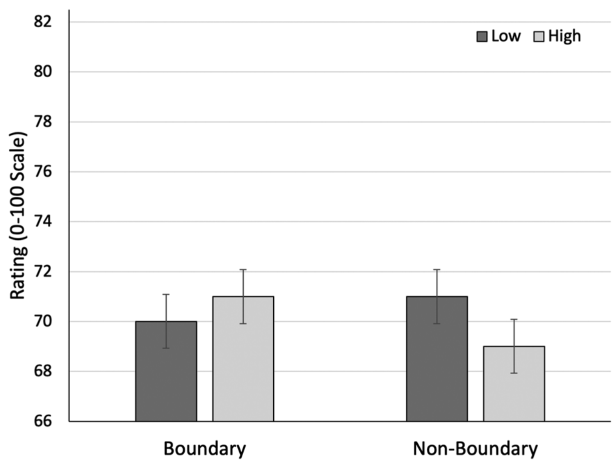

FIGURE 3 Effects of boundary type and relative magnitude of SAT verbal score on ratings of strength of admissions applicant portfolios using a sliding scale (Experiment 1). Note: Analysis of the judgment difference score (see Section 6.2) reveals that the high boundary minus the low boundary rating is greater than the high non-boundary minus the low non-boundary rating ($p < .05$). Error bars indicate 1 unit of SE in each direction. In all studies, boundary and non-boundary ratings are not directly comparable (see Section 6.1)

range = −41–37) and was found to be significantly different from 0 ($t(133) = 2.16$, $SE = 1.24$, $p = .032$, $d = 0.19$, CI[0.23, 5.14]), evidence of a small left digit effect (57% of participants had scores > 0). There were no gender differences in scores ($t(132) = 0.83$, $SE = 2.52$, $p = .410$).

6.2.2 | NLE task

Following Lai et al. (2018), individual estimates were excluded if they differed from the group mean for a target numeral by more than 2 SDs ($M = 3.66\%$ of trials). All 120 trials were used to compute a measure of overall accuracy called *PAE* = |*actual placement – correct location*|*/1000 * 100*. Consistent with past work, the mean error was 3.98% ($SD = 1.20$; range = 1.95–9.01). We calculated a *hundreds difference score = placement for larger numeral – placement for smaller numeral* for each critical pair, then calculated the average across eight pairs yielding one average hundreds difference score for each participant. Hundreds difference scores greater than 0 indicate a left digit effect. Across participants, the hundreds difference score had a $M = 25.89$ ($SD = 21.53$, range = −37.63–85.43) and was significantly greater than 0 ($t(133) = 13.92$, $SE = 1.86$, $p < .001$, $d = 1.20$, CI[22.21, 29.57]), evidence of a large left digit effect (89% of participants had scores > 0). There were no gender differences ($t(132) = -0.18$, $SE = 3.78$, $p = .854$).

6.2.3 | Relationship between measures

We assessed the relationship between the judgment difference score and hundreds difference score, the two measures of the left digit effect, using a Pearson correlation. Although the correlation was in the predicted direction, no statistically significant relationship was found, $r(132) = -.11$, $p = .224$. There was also no relationship between the judgment difference score and the measure of overall accuracy on the NLE task called PAE, $r(132) = -.04$, $p = .664$.

6.3 | Discussion

We found a small left digit effect in the college admissions task and a large left digit effect in the NLE task (the latter similar in magnitude to past work; Lai et al., 2018). Experiment 1 provides evidence that the left digit effect, while smaller in this context, does extend to the novel judgment task. There was also large variation in measures of the left digit effect on both tasks for individual participants. However, there was no reliable relationship between the magnitudes of one's left digit effect across the two tasks. The latter finding offers no evidence of common cognitive contributors to individual differences in the left digit effect across tasks and no evidence that the NLE task can be used to predict bias in more complex judgment tasks such as the college admissions task.

One unexpected pattern in the college admissions task was that the difference in the ratings for the high and low boundary portfolios was not large, and the direction of the difference reversed for the non-boundary portfolios. This pattern, while not inconsistent with a left digit effect, is still unexpected. We considered the possibility that higher value portfolios were systematically underrated or that the lower value ones were overrated, either of which could produce the observed pattern. While we did not identify any particular elements of Experiment 1 that would give rise to such a pattern, there were some procedural decisions that might have influenced ratings more generally that we address in Experiment 2.

One such procedural decision of Experiment 1 was that we spaced boundary portfolios farther apart (eight intervening trials) than non-boundary portfolios (five intervening trials). The unmatched spacing was intended to obscure the study's goals. However, greater experience with the task suggested to us that, even with matched spacing, the goals of the study would not be transparent. A second limitation of Experiment 1 was that the placement of critical portfolios among fillers was not counterbalanced; rather, each critical portfolio was preceded by different fillers. It is possible that the immediately preceding portfolios influenced the evaluation of the critical portfolio. In Experiment 2, we modify the procedure to remove both of these confounds. We also add a manipulation of SAT math score (similar to the manipulation of SAT verbal score) in an attempt to amplify any left digit effect that might exist.

7 | EXPERIMENT 2: MODIFIED COLLEGE ADMISSIONS TASK

Experiment 2 was the same as Experiment 1 except for modifications to the counterbalancing scheme and the four critical portfolios for the college admissions task. Critical portfolios were presented in trials 10, 13, 16, and 19; the portfolios appeared in these positions in one of four counterbalancing orders. The 16 filler items from Experiment 1 were presented in a single order across the remaining trials. The aims of this design were to ensure that the two boundary portfolios and the two non-boundary portfolios were separated by the same number of intervening fillers and that each critical portfolio appeared an equal number of times in each of the four locations. To amplify the size of any left digit effect, we also manipulated SAT math score (along with SAT verbal score; see Table 1). For example, if an SAT verbal score fell just above a hundreds boundary for one of the critical portfolios, the SAT math score also fell just above a different hundreds boundary in the same portfolio (e.g., SAT verbal score of 701 and SAT math score of 602). (Additional changes to critical portfolios are reported in Section 7.1.) No changes were made to the NLE task. As in Experiment 1, we expected the judgment difference score and hundreds difference score across participants to be reliably greater than 0, indicating left-digit effects in both tasks. We also expected the effect size for the judgment difference score to be greater than it was in Experiment 1. Finally, if the lack of correlation between measures in Experiment 1 was due to the left digit effect in the college admissions task having been weakly elicited, we would expect to see a correlation here.

7.1 | Method

7.1.1 | Participants

Adults (N = 157 undergraduates; 92 women, 64 men, and 1 unidentified) completed the study in exchange for introductory psychology course credit.[5] Participants came to the lab individually for 1-h sessions during which time they completed the college admissions task, the NLE task, and several unrelated tasks. Participants were sequentially assigned to one of four counterbalancing orders of the college admissions task ($n \approx 39$ per order).

7.1.2 | Materials and procedure

The college admissions task was the same as Experiment 1 with the following exceptions. First, the filler stimuli appeared in a single order for all participants (the sequencing matched Experiment 1, Order 1 in supporting information); the four critical stimuli were presented on trials 10, 13, 16, and 19. Second, participants saw the critical stimuli in one of four counterbalancing orders (LB = low boundary, LN = low non-boundary, HB = high boundary, and HN = high non-boundary): HB/HN/LB/LN (Order 1), HN/HB/LN/LB (Order 2), LB/LN/HB/HN (Order 3), and LN/LB/HN/HB (Order 4). Third, the four target stimuli were modified from Experiment 1, as shown in Table 1. Most notably, SAT math scores, rather than being the same across paired portfolios, differed by four points that crossed a hundreds boundary (boundary portfolio) or by four points that did not cross a hundreds boundary (non-boundary portfolio). Two GPAs were also changed in order to better equate overall strength of boundary and non-boundary portfolios (though the design does not depend on these being equated).

7.2 | Results

7.2.1 | College admissions task

There were no order effects (F(3,153) = 0.60, SE = 196.45, p = .617). We collapsed over order for further analyses. The mean rating for each of the four critical stimuli is shown in Figure 4. The figure reveals that, for the pair of boundary portfolios, higher ratings were given by participants on average to the portfolio with the higher SAT scores (M = 76.54, SD = 11.69) relative to the lower scores (M = 71.43, SD = 11.59), a larger positive difference than in Experiment 1. For non-boundary portfolios, higher ratings were also given to the portfolio with the higher SAT scores (M = 74.59, SD = 12.27) over the lower scores (M = 73.65, SD = 12.41), but the difference was much smaller, as predicted. The difference in ratings for the high and low boundary portfolios was statistically significant (t(156) = 5.78, SE = 0.88,

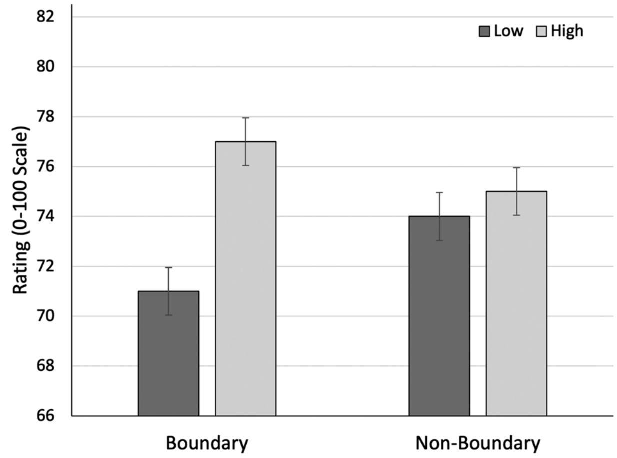

FIGURE 4 Effects of boundary type and relative magnitude of SAT verbal and math scores on ratings of strength of admissions applicant portfolios using a sliding scale (Experiment 2). Note: Analysis of the judgment difference score (see Section 7.2) reveals that the high boundary minus the low boundary rating is greater than the high non-boundary minus the low non-boundary rating (p < .05). Error bars indicate 1 unit of SE in each direction

p < .001), while the difference for the non-boundary portfolios was not (t(156) = 1.17, SE = 0.81, p = .244).

As in Experiment 1, we created a judgment difference score for each participant. Recall that a value greater than 0 indicates that the difference in ratings of the two boundary portfolios is greater than that of the two non-boundary portfolios and thus that there is a left digit effect. The judgment difference score across participants was normally distributed with a M = 4.17 (SD = 13.96, range = −41–33) and was significantly different from 0 (t(156) = 3.74, SE = 1.11, p < .001, d = 0.30, CI[1.96, 6.37]), evidence of a moderate left digit effect (60% of participants had scores > 0). There were no gender differences in scores (t(154) = 1.13, SE = 2.28, p = .262).

7.2.2 | NLE task and relationship between tasks

After outliers were removed (M = 4.45% of trials), PAE was 3.89% (SD = 1.28; range = 1.50–8.14), similar to Experiment 1. If leftmost digits influence performance, estimates should be greater for target numerals just above a hundreds boundary than for those just below the boundary. Across participants, the hundreds difference score had a M = 19.98 (SD = 19.51, range = −31.14–69.43) and was significantly greater than 0 (t(156) = 12.83, SE = 1.56, p < .001, d = 1.02, CI[16.90, 23.05]), evidence of a large left digit effect (86% of participants had scores > 0). There were no gender differences (t(154) = 0.28, SE = 3.19, p = .430). To test for a relationship between the judgment difference score and the hundreds difference score, we ran a Pearson correlation. No statistically significant relationship was found, r(155) = .02, p = .777 (further, the correlation was positive, while it was negative in the first study). There was also no relationship between the judgment difference score and the PAE (r(155) = −.10, p = .228). Both findings are similar to those of Experiment 1.

7.3 | Discussion

The study replicates the finding of a left digit effect in the college admissions task. In addition, the ratings of the four critical portfolios followed the anticipated pattern: Portfolios with the high boundary values were rated as much stronger than those with the low boundary values, while portfolios with high non-boundary values were rated as only slightly stronger than those with low non-boundary values. We do not know whether any specific methodological modifications contributed to the pattern of findings but suspect that the modified counterbalancing scheme contributed to high values now being rated consistently higher than low values. It is also likely that the manipulation of both SAT verbal and SAT math scores amplified the left digit effect in Experiment 2 (as only verbal scores were manipulated in Experiment 1). Consistent with Experiment 1, there was no relationship between one's left digit effect in the college admissions task and the NLE task, suggesting that the null finding in Experiment 1 was not due to the weaker elicitation of the effect there.

The findings thus far provide evidence of a left digit effect in the college admissions task and the NLE task and no relationship between the size of the left digit effect in the two tasks. In Experiment 3, we sought to replicate and extend the finding from Experiment 2. We repeated the procedure of Experiment 2 except that, for the college admissions task, we used a different type of response scale. Specifically, rather than the sliding response scale (which used verbal boundaries of "weak" and "strong"), we used a numerical response scale (i.e., rate portfolio strength on a scale from 1 to 100). The purpose of this change is to ensure that the left digit effect emerges across commonly used response modalities (and thus that the effect is not driven by the conversation of a magnitude to a particular response format). As in Experiments 1 and 2, we also considered the relationship between measures of the left digit effect across tasks. However, with no reason to predict that a relationship might exist only with a numerical response scale, we expected to replicate the finding of no relationship between task measures here.

8 | EXPERIMENT 3: MODIFIED COLLEGE ADMISSIONS TASK WITH NUMERICAL RESPONSE SCALE

Experiment 3 was similar to Experiment 2 except for a change in the scale used to elicit ratings in the college admissions task (as well as one procedural change described in Section 8.1). Participants provided ratings on a numerical scale from 0 = *weak* to 100 = *strong* by typing a numeral between 0 and 100 into a text box. If the left digit effect in the college admissions task generalizes across response modalities, it should replicate here: The difference between high and low boundary ratings should be greater than the difference between high and low non-boundary ratings, with the latter close to 0, as in Experiment 2. Alternatively, we did consider that the numerical scale might be less intuitive and that responses might be more difficult to recall across trials. As a result, responses might be less consistent across trials, which could obscure any left-digit bias. However, despite this possibility, based on the findings of Experiments 1 and 2, we predicted that we would replicate the left digit effect in both tasks but that the two measures of the effect would again not be correlated with one another.

8.1 | Method

8.1.1 | Participants

Adults (N = 143 undergraduates; 83 women and 60 men) completed the study in exchange for introductory psychology course credit.[6] The first 57 participants came to the lab individually for 1-h sessions during which time they completed the college admissions task and the NLE task; the remainder participated remotely with tasks administered through a web browser (for health safety during a virus outbreak).[7] Participants were sequentially assigned to one of four counterbalancing orders of the college admissions task ($n \approx 36$ per order).

8.1.2 | Materials and procedure

The college admissions task was the same as Experiment 2 with the exception that the sliding response scale was replaced with the text label "Rating (from 0 = weak to 100 = strong)" followed by a text box for one's response. The NLE task was the same as that used in Experiments 1 and 2. The overall procedure was also the same as in Experiments 1 and 2 with the exception that some participants engaged in tasks remotely. For the latter group, both the college admissions task and the NLE task were reprogrammed using lab.js software (lab.js.org; Henninger et al., 2019) and administered using the Open Lab platform (open-lab.online). Remote participants were instructed to sit in a quiet room with no distractions for the duration of the study. They spoke to the experimenter by phone at the start and end of their scheduled session and the experimenter was available by phone throughout the session to answer questions, a procedure intended to match the in-person experience as closely as possible.

8.2 | Results

8.2.1 | College admissions task

There was no procedural (in-person vs. remote) effect ($t(141) = -0.46$, $SE = 2.29$, $p = .645$). Unlike Experiments 1 and 2, there was an order effect ($F(3, 139) = 5.09$, $MSE = 163.91$, $p = .002$, in that the judgment difference score was largest for Order 2 and smallest for Order 1 (all means were > 0 except Order 1). As in Experiments 1 and 2, we collapsed over order for further analyses. The mean rating for each of the four critical stimuli is shown in Figure 5. The figure reveals that for the pair of boundary portfolios, higher ratings were given by

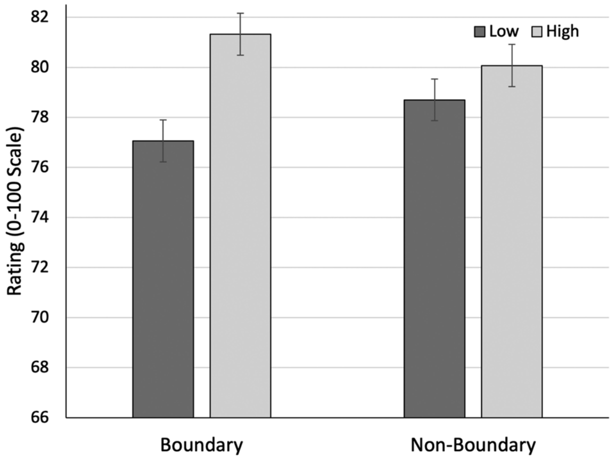

FIGURE 5 Effects of boundary type and relative magnitude of SAT verbal and math scores on ratings of strength of admissions applicant portfolios using a numerical scale (Experiment 3). Note: Analysis of the judgment difference score (see Section 8.2) reveals that the high boundary minus the low boundary rating is greater than the high non-boundary minus the low non-boundary rating ($p < .05$). Error bars indicate 1 unit of *SE* in each direction

participants to the portfolio with the higher SAT scores ($M = 81.32$, $SD = 9.60$) relative to the lower scores ($M = 77.07$, $SD = 10.90$), similar to (but smaller than) the difference reported in Experiment 2. For non-boundary portfolios, higher ratings were also given by participants to the portfolio with the higher SAT scores ($M = 80.07$, $SD = 9.43$) over the lower scores ($M = 78.70$, $SD = 9.87$), but the difference was very small, as in Experiment 2. The difference in ratings for the high and low boundary portfolios was statistically significant ($t(142) = 5.15$, $SE = 0.83$, $p < .001$) as was that for the non-boundary portfolios ($t(142) = 2.22$, $SE = 0.62$, $p = .028$).

As in Experiments 1 and 2, of focal interest was the judgment difference score. The judgment difference score across participants was normally distributed with $M = 2.88$ ($SD = 13.35$, range = –35–47) and was significantly different from 0 ($t(142) = 2.58$, $SE = 1.12$, $p = .011$, $d = 0.22$, CI[0.68, 5.09]), evidence of a small left digit effect (57% of participants had scores >0). The effect size was larger than in Experiment 1 ($d = 0.19$) but smaller than Experiment 2 ($d = 0.30$). There were no gender differences in scores ($t(141) = -1.10$, $SE = 2.26$, $p = .272$).

8.2.2 | NLE task and relationship between tasks

After outliers were removed ($M = 3.84\%$ of trials), PAE was 3.63% ($SD = 1.17$; range = 1.46–7.29), similar to Experiments 1 and 2. Across participants, the hundreds difference score had a $M = 31.50$ ($SD = 16.18$, range = –27.13–72.50) and was significantly greater than 0 ($t(142) = 23.28$, $SE = 1.35$, $p < .001$, $d = 1.95$, CI[28.83, 34.18]), evidence of a large left digit effect (93% of participants had scores > 0). There were no gender differences ($t(141) = -0.18$, $SE = 2.75$, $p = .647$). As in Experiments 1 and 2, no reliable relationship was found between left digit effect measures, $r(141) = .04$, $p = .605$, or between the judgment difference score and PAE, $r(141) = .16$, $p = .063$, although the latter approached statistical significance.

8.3 | Discussion

Experiment 3 replicates the pattern of college admissions ratings found in Experiment 2 that contributed to a left digit effect: Portfolios with the high boundary values were given much higher ratings than those with the low boundary values, while portfolios with high non-boundary values were rated as only slightly higher than those with low non-boundary values. Further, Experiment 3 provides evidence that the left digit effect is not specific to a particular response format: the pattern emerged here with the use of a numerical response scale instead of the previously used sliding response scale. The effect size was smaller here than in Experiment 2, perhaps due to the change in response format (and participants gave higher ratings overall). However, given unexpected changes across studies in the in-person versus remote procedure for some participants, we do not draw conclusions about the smaller effect size. Consistent with Experiments 1 and 2, this study provides no evidence of a relationship between the magnitude of one's left digit effect in the college admissions and NLE tasks.

9 | COMBINED DATA: MODEL OF CANDIDATE RATINGS

In exploratory analyses using the data across all three experiments ($N = 434$), we considered whether participants, as instructed, weighed the portfolio dimensions equally in assigning ratings to candidates, as emergence of a left digit effect at the individual level depends on use of SAT scores to form one's judgments. Toward addressing this question, we ran a linear regression to predict candidate ratings using the five portfolio dimensions. For each participant, we obtained a standardized beta (β) value for each dimension. Averaged across the three studies, the β values were as follows: GPA ($M = 0.42$, $SD = 0.17$), letters of recommendation ($M = 0.13$, $SD = 0.14$), essay ($M = 0.13$, $SD = 0.16$), SAT verbal ($M = 0.22$, $SD = 0.20$), and SAT math ($M = 0.22$, $SD = 0.16$). The greatest weight was given to GPA, followed by SAT verbal and SAT math scores. The model (including a constant) explained an average of 91% of variance in ratings (range = 43%–100%, except for one participant with R^2 of 1%). There was no relationship between the judgment difference score and SAT verbal β value, SAT math β value, or R^2 ($|r|$s < .07, ps > .199), evidence that differences in weighting of SAT scores in the college admissions task did not modulate the size of the observed left digit effect. Further, controlling for these variables did not change the relationship between the judgment difference score and the hundreds difference score ($r \approx -.03$ across all three experiments).

With the larger data set, we also asked whether an individual's R^2 for the college admissions task could be predicted from either of the NLE measures. The R^2 is not a measure of the left digit effect, but it does reflect consistency in one's use of attributes to inform one's ratings across trials and, as such, it may be higher for those with stronger number skills. In fact, we found that the R^2 value for the college admissions task was correlated with hundreds difference score (r(432) = −.12, p = .017) and PAE (r (432) = −.12, p = .013). Although the correlations are small, they do suggest that individuals with better numerical magnitude estimation might also use attributes more consistently in their judgments.

10 | GENERAL DISCUSSION

There are two key findings. The first finding is that we found a small but statistically significant left digit effect in the college admissions task: Matched candidates were rated as more different from one another when SAT scores had different leftmost digits. That the size of the effect increased when we manipulated two instead of just one SAT score is further evidence that an overreliance on the leftmost digits of SAT scores was driving the portfolio ratings. The effect was also found using two different types of response scales and thus does not appear to be specific to a particular response format. These findings add to existing evidence of a left digit effect in judgment and decision-making contexts (e.g., Thomas & Morwitz, 2005) and extend the effect—which is most often studied using consumer prices—to the use of standardized test scores. This is important given that many past studies have used values for which there might also be domain-specific strategies, such as the truncating of cents in the context of prices (MacKillop et al., 2014), or the days or weeks beyond one's most recent birthday (i.e., 78 and 50 weeks is still commonly "78 years old"; Olenski et al., 2020). The present studies also provide laboratory evidence of bias not just during single-value judgments but also when one must integrate multiple relevant values into an overall assessment, consistent with other types of studies that use existing records to infer a left-digit bias in decision making.

The second key finding is that even though there were large individual-level variations in the left digit effect in the judgment and decision-making task and the basic numerical magnitude estimation task used here, the size of one's left digit effect was not reliably correlated across tasks. Such a correlation would have provided evidence that the NLE task might be used both to predict the size of one's left digit effect across a range of judgment tasks and might be used as a debiasing technique to train reductions in the left digit bias across tasks. It also would have provided evidence of a common underlying mechanism across tasks, supporting the possibility that more complex judgments arise from more basic processes of numerical cognition. While the present findings do not rule out these possibilities, they raise several questions that should be considered in future work.

One question raised is whether variation in the size of the left digit effect in the NLE task in fact reflects a stable individual difference. It is possible that variations in performance do not reflect stable individual differences. We do know that test–retest reliability is about r = .45 for number line accuracy (e.g., PAE; Schneider et al., 2018), but such reliability has not yet been established for the hundreds difference score in this task. An important next step will be to study the test–retest reliability for the hundreds difference score toward providing evidence of consistency in scores over time. Because the judgment difference score used here for the college admissions task was based on a small number of trials, it might also be valuable to develop a modified judgment task that includes a greater number of critical trials, to ensure a robust measure of left-digit bias in this context (see Aczel et al., 2015; McElroy & Dowd, 2007; Teovanović, 2019, for individual differences in judgment heuristics). Despite these limitations, we do know that measures of the left digit effect have in the past been related to individual difference measures such as thinking style (using a single-item price comparison task; Tu & Pullig, 2018) and verbal standardized test achievement (Williams, Paul, et al., 2020), and even to consistency in use of attributes in the college admissions task here, suggesting that they do likely capture some stable variation in task performance related to number cognition.

A second question raised is whether the NLE task and the college admissions tasks draw on a large number of skills that are not shared across tasks, perhaps obscuring any common source of variance in performance. For example, in the NLE task, overall accuracy has been shown to depend not only on numerical magnitude estimation but also on spatial skills and strategies including proportional reasoning (Cohen & Blanc-Goldhammer, 2011; Slusser & Barth, 2017; Sullivan et al., 2011). Where a target is placed on the number line depends in part on whether one uses the midpoint and other reference points on the number line to guide responses (Peeters et al., 2017). In judgment tasks such as the college admissions task, individuals likely differ in how they approach the task, specifically, in whether they adopt a more intuitive or deliberative strategy for integrating dimensional information into a single assessment of each portfolio (Payne et al., 2008; Usher et al., 2011). Individuals may also engage higher order types of reasoning about numbers, such as associating .99 on a consumer price with the presence of a discount or sale (Gabor, 1977; Gabor & Granger, 1964). Although we used a similar response scale in both the numerical estimation and college admissions tasks (in Experiments 1 and 2), future work might involve use of tasks that are even more closely aligned, while still contrasting the activity of estimating the location of a target number on a response line with judging its cost, quality, strength, and so forth. This approach could suggest what degree of similarity is needed for left-digit measures to converge.

Recall that several researchers have proposed that the left digit effect in judgment and decision making might be explained by cognitive processes associated with numerical magnitude estimation (Huber et al., 2016; Macizo & Ojedo, 2018; Thomas & Morwitz, 2005, 2009). The present findings do not rule out this possibility, and the recent finding of the left digit effect in NLE (Lai et al., 2018) supports it. It remains likely that theories of numerical magnitude estimation will ultimately at least partially explain the left digit effect in judgment and decision making. According to one compelling recent model of number to quantity conversion (Dotan & Dehaene, 2020), each digit in a numeral (that is read) is first bound to a syntactic role (e.g., hundreds place and tens place) and then is weighted according to its role. Digit-based quantities are then integrated into a whole-number quantity. According to this model, one has access to weighted digit-based magnitudes that might be used when task relevant, but the model also assumes rapid construction of whole-number estimates. This model may be a promising candidate for understanding the left digit effect in NLE, in that the effect could easily be explained as an overweighting of leftmost digits during the integration process.

What still would need to be explained is, to the extent there is variation in the overweighting of the leftmost digit during the integration process, *why* might this be the case? The starting point of the present work was an assumption that appropriate weighting and integration of digit magnitudes may reflect a learned skill and thus that individuals with stronger magnitude estimation skills might, for example, have better calibration in their weighting of digits across tasks. This is still possible, but the present findings do not lend new support to this possibility. An alternative possibility is that overweighting of the leftmost digit might reflect across-trial fluctuations in attention, motivation, or response speed. Overweighting might also vary as a function of what magnitudes have been recently activated (e.g., in prior trials) or are being evaluated concurrently. Such possibilities would offer another explanation as to why the left digit effect was not related across tasks here, even if the underlying mechanism of overweighting the leftmost digit during integration is shared across tasks.

NLE accuracy has been used to predict performance on standardized math achievement tests (Booth & Siegler, 2008; Hoffmann et al., 2013; Holloway & Ansari, 2009; Schneider et al., 2009, 2018) and adults' numeracy (Peters & Bjalkebring, 2015; Schley & Peters, 2014), and the NLE task has been used as a training tool to improve numerical skills in children (Schneider et al., 2018). Recent research has been aimed at understanding whether the left-digit bias in NLE might similarly predict other types of performance and whether it might also be improved through NLE training. Ongoing work in our lab, for example, explores the developmental trajectory of the left digit effect (Williams, Zax, et al., 2020) and whether it is possible to reduce the left digit effect through feedback interventions with adults (Williams, Xing, et al., 2020). Although the present findings suggest some caution when considering some of these possibilities with regard to the left digit effect, it will be important to explore the issues further, as they have implications for identifying and reducing judgment bias in decision making in contexts ranging from consumer purchases to health behavior and medicine.

ACKNOWLEDGMENTS

This work benefited from National Science Foundation (Division of Research on Learning in Formal and Informal Settings) grants DRL-1561214 and DRL-1920445 to HB and ALP. We thank undergraduates in the Reasoning and Decision Making Lab at Wesleyan University for their assistance.

CONFLICT OF INTERESTS

The authors have no conflicts of interest to declare.

ORCID

Andrea L. Patalano https://orcid.org/0000-0002-5829-7703
Katherine Williams https://orcid.org/0000-0001-8963-7574
Hilary Barth https://orcid.org/0000-0003-2937-2716

ENDNOTES

[1] Four additional participants completed the study but, after outlier removal, were missing > 3 pairs of hundreds difference scores. Per preregistered exclusion criteria, they are not included in this report.

[2] Tasks were administered in this order because the NLE task was a much longer task than the admissions rating task (120 trials vs. 20 trials, respectively). Our assessment was that the NLE task would be more likely to influence performance on the rating task (as the latter was less transparently about magnitude estimation) than the reverse.

[3] The formal name of the verbal section of the SAT is "Evidence Based Reading and Writing."

[4] That the first block of NLE trials was part of another study had no impact on the present study's design. The first block of NLE trials was a standard 120 trials administered to all participants. It was not until a second block of trials (not reported here) that a feedback manipulation was introduced in which half of participants were given feedback on their performance after each set of 20 trials while the other half of participants were not.

[5] Four additional participants completed the study but, after outlier removal, were missing > 3 pairs of hundreds difference scores, and so are not included in this report.

[6] Four additional participants completed the study but, after outlier removal, were missing > 3 pairs of hundreds difference scores, and so are not included in this report; two further participants were excluded for misunderstanding the college admissions task instructions (one gave 0–10 ratings, and the other gave 0–1000 ratings).

[7] The 2020 coronavirus pandemic began during the running of the study.

REFERENCES

Aczel, B., Bago, B., Szollosi, A., Foldes, A., & Lukacs, B. (2015). Measuring individual differences in decision biases: Methodological considerations. *Frontiers in Psychology*, *6*, 1770. https://doi.org/10.3389/fpsyg.2015.01770

Anderson, E., & Simester, D. (2003). Effects of $9 price endings on retail sales: Evidence from field experiments. *Quantitative Marketing and Economics*, *1*(1), 93–110. https://doi.org/10.1023/A:1023581927405

Bar-Hillel, M. (1980). What features make samples seem representative? *Journal of Experimental Psychology: Human Perception and Performance*, *6*(3), 578–589. https://doi.org/10.1037/0096-1523.6.3.578

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), 147–162. https://doi.org/10.1080/10835547.2015.12092101

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

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

Dotan, D., & Dehaene, S. (2020). Parallel and serial processes in number-to-quantity conversion. *Cognition*, *204*, 104387. https://doi.org/10.1016/j.cognition.2020.104387

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, *12*(5), 391–396. https://doi.org/10.1111/1467-9280.00372

Gabor, A. (1977). *Pricing: Principles and practices*. London: Heinemann.

Gabor, A., & Granger, C. W. J. (1964). Price sensitivity of the consumer. *Journal of Advertising Research*, *4*(4), 40–44.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, *38*(1), 129–166. https://doi.org/10.1006/cogp.1998.0710

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

Hoffmann, D., Hornung, C., Martin, R., & Schiltz, C. (2013). Developing number–space associations: SNARC effects using a color discrimination task in 5-year-olds. *Journal of Experimental Child Psychology*, *116*(4), 775–791. https://doi.org/10.1016/J.JECP.2013.07.013

Hogarth, R. M., & Einhorn, H. J. (1990). Venture theory: A model of decision weights. *Management Science*, *36*(7), 780–803. https://doi.org/10.1287/mnsc.36.7.780

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

Huber, S., Nuerk, H.-C., Willmes, K., & Moeller, K. (2016). A general model framework for multisymbol number comparison. *Psychological Review*, *123*(6), 667–695. https://doi.org/10.1037/rev0000040

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251. https://doi.org/10.1037/h0034747

Lacetera, N., Pope, D. G., & Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review*, *102*(5), 2206–2236. 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

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, *15*(3), 374–378. https://doi.org/10.1086/209174

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

Macizo, P., & Ojedo, F. (2018). The processing of price comparison is not holistic but componential. *Psychological Research*, *82*(6), 1144–1157. https://doi.org/10.1007/s00426-017-0884-4

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

McElroy, T., & Dowd, K. (2007). Susceptibility to anchoring effects: How openness-to-experience influences responses to anchoring cues. *Judgment and Decision Making*, *2*(1), 48–53.

Nuerk, H.-C., Weger, U., & Willmes, K. (2004). On the perceptual generality of the unit-decade compatibility effect. *Experimental Psychology*, *51*(1), 72–79. https://doi.org/10.1027//1618-3169.51.1.72

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

Olsen, A. L. (2013). Leftmost-digit-bias in an enumerated public sector? An experiment on citizens' judgment of performance information. *Judgment and Decision Making*, *8*(3), 365–371.

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

Payne, J. W., Samper, A., Bettman, J. R., & Luce, M. F. (2008). Boundary conditions on unconscious thought in complex decision making. *Psychological Science*, *19*(11), 1118–1123. https://doi.org/10.1111/j.1467-9280.2008.02212.x

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

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

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

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*(5), 407–413. https://doi.org/10.1111/j.1467-9280.2006.01720.x

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*(1), 89–107. https://doi.org/10.1016/j.lindif.2007.03.011

Rottenstreich, Y., & Hsee, C. S. (2001). Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological Science*, *12*(3), 185–190. https://doi.org/10.1111/1467-9280.00334

Schindler, R. M., & Kibarian, T. M. (1996). Increased consumer sales response through use of 99-ending prices. *Journal of Retailing*, *72*(2), 187–199. https://doi.org/10.1016/S0022-4359(96)90013-5

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., 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

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

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

Teovanović, P. (2019). Individual differences in anchoring effect: Evidence for the role of insufficient adjustment. *Europe's Journal of Psychology*, *15*(1), 8–24. https://doi.org/10.5964/ejop.v15i1.1691

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. R. Rao (Ed.), *Handbook of pricing research in marketing* (pp. 132–149). Northampton, MA: Edward Elgar Publishing. https://doi.org/10.4337/9781848447448

Tu, L. L., & Pullig, C. (2018). Penny wise and pound foolish? How thinking style affects price cognition. *Marketing Letters*, *29*(2), 261–273. https://doi.org/10.1007/s11002-018-9460-y

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Usher, M., Russo, Z., Weyers, M., Brauner, R., & Zakay, D. (2011). The impact of the mode of thought in complex decisions: Intuitive decisions are better. *Frontiers in Psychology*, *2*, 37. https://doi.org/10.3389/fpsyg.2011.00037

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 and Behavior*, *10*, e01877. https://doi.org/10.1002/brb3.1877

Williams, K., Xing, C., Bradley, K., Patalano, A. L., & Barth, H. (2020). Potential moderators of the left digit effect in numerical estimation. [Manuscript in preparation] Department of Psychology, Wesleyan University.

Williams, K., Zax, A., Patalano, A.L., & Barth, H. (2020). Left digit effects in numerical estimation across development. [Manuscript under revision] 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. T. (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

AUTHOR BIOGRAPHIES

Andrea L. Patalano is a professor in the Department of Psychology and the Program in Neuroscience and Behavior at Wesleyan University, where she runs the Reasoning and Decision Making Lab. Her research interests include number-related judgment biases, indecisiveness and decisional delay, and neural underpinnings of decision making.

Katherine Williams, formerly a joint lab coordinator at Wesleyan University, is now a graduate student in Developmental Psychology at the University of Texas at Austin. She is broadly interested in how social, linguistic, and cognitive factors interact to support learning.

Gillian Weeks is an undergraduate at Wesleyan University majoring in Psychology and a member of the Reasoning and Decision Making Lab. She conducts research on the left digit effect in complex judgment and is particularly interested in clinical applications.

Kelsey Kayton is the lab coordinator for the Reasoning and Decision Making Lab at Wesleyan University. She is interested in biases in judgment and decision making that impact how people evaluate themselves and others.

Hilary Barth is a professor in the Department of Psychology at Wesleyan University and director of the Cognitive Development Lab at Wesleyan. Research in her lab investigates mathematical cognition; thinking and learning about number, space, time, and probability; social cognition; and decision making.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Patalano AL, Williams K, Weeks G, Kayton K, Barth H. The left digit effect in a complex judgment task: Evaluating hypothetical college applicants. *J Behav Dec Making*. 2021;1–14. https://doi.org/10.1002/bdm.2247