Linguistic Inversion and Numerical Estimation

Sophie Savelkouls[1], Katherine Williams[2], & Hilary Barth*[2]

[1]Department of Psychology, Boston College, Boston, USA

[2]Department of Psychology, Wesleyan University, Middletown, USA

Corresponding Author:
Hilary Barth, Ph.D.
Department of Psychology
Judd Hall
Wesleyan University
207 High St.
Middletown, CT 06459
hbarth@wesleyan.edu

Number line estimation (NLE) performance is usually believed to depend on the magnitudes of presented numerals, rather than on the particular digits instantiating those magnitudes. Recent research, however, shows that NLE placements differ considerably for target numerals with nearly identical magnitudes, but instantiated with different leftmost digits (Lai, Zax, & Barth, 2018). Here we investigate whether this left digit effect may be due, in part, to the ordering of digits in number words. In English, the leftmost digit of an Arabic numeral is spoken first ("forty-one"), but Dutch number words are characterized by the inversion property: the rightmost digit of a two-digit number word is spoken first ("eenenveertig" - one and forty in Dutch). Participants (N = 40 Dutch-English bilinguals and N = 20 English-speaking monolinguals) completed a standard 0-100 NLE task. Target numerals were read aloud by an experimenter in either English or Dutch. Preregistered analyses revealed a strong left digit effect in monolingual English speakers' estimates: e.g., 41 was placed more than two units to the right of 39. No left digit effect was observed among Dutch-English bilingual participants tested in either language. These findings are consistent with the idea that the order in which digits are spoken might influence multi-digit number processing, and suggests linguistic influences on numerical estimation performance.

Keywords

numerical cognition, estimation, number line, left digit effect

Linguistic Inversion and Numerical Estimation

Number line estimation (NLE) tasks are widely used to inform theories of cognitive development and learning. These tasks have been shown to predict children's real-world math outcomes (Ramani & Siegler, 2008; Schneider et al., 2018, Siegler, 2016; Xing et al., 2020) and are useful educational tools, e.g. supporting fraction learning (Hamdan & Gunderson, 2017). In typical NLE tasks, participants estimate the location of target numerals (e.g., "36") on a horizontal line with labeled endpoints (e.g., 0 and 100). A nearly universal assumption underlying interpretation of these data is that performance is driven by the magnitudes of target numerals, independent of the specific digits instantiating them. However, recent studies demonstrate that digit level information, not just overall magnitude, strongly influences performance: people produce very different estimates for numbers with nearly identical magnitudes but different leftmost digits. For example, on a 0-1000 number line, numerals with different hundreds-place digits (e.g., 299/302) are placed in very different locations despite having magnitudes that should be indistinguishable on this scale, with large effect sizes for children and adults (Lai et al., 2018). In another study using a 0-100 NLE task, estimates were systematically different for two-digit numbers with different leftmost (tens place) digits; e.g., 41 was placed more than 2 units to the right of 39 (work that is currently under review; Williams, Zax, Patalano, & Barth, 2020). Both children and adults exhibit a left digit effect in NLE for two-digit and three-digit numbers.

Digit identity, not just overall magnitude, matters when people make numerical judgments such as NLE placements. This finding is consistent with price comparison studies in which participants judge $5.00 to be significantly more costly than $4.99, but judge $4.20 and $4.19 similarly, despite the equivalent numerical distances between the pairs (Thomas &

Morwitz, 2005). Importantly, left digit effects, as demonstrated in NLE and in price comparison studies, are potentially compatible with both decomposed and holistic theories of multidigit number processing and representation; we return to this issue in the General Discussion.

The current research investigates whether the order in which individual digits are spoken/read in two-digit numbers might play a role in multidigit number processing and the production of left digit effects in NLE. The English number word system parallels the place-value structure of Arabic numerals, with the leftmost place value spoken first (41 is "forty-one"). Some other languages do not follow this place-value structure. For example, German and Dutch number words are characterized by the inversion property with the tens and units inverted (e.g., 41 is "eenenveertig" – one-and-forty in Dutch). Some evidence consistent with an influence of inversion on numerical processing comes from performance patterns of speakers of inverted languages on numerical tasks involving magnitude comparison (Nuerk, Weger, & Willmes, 2005), transcoding (Pixner et al., 2011), and even complex arithmetic (Göbel, Moeller, Pixner, Kaufmann, & Nuerk, 2014). For example, transcoding errors (i.e., reading or writing digits in the wrong order) occur more frequently in inverted languages such as German (Zuber, Pixner, Moeller, & Nuerk, 2009). Children who speak a language consisting of two number-word systems (e.g., Czech) produce more errors for numbers presented in inverted form vs. when the same numbers are presented in their noninverted form (e.g., hearing "six-and-eighty" and writing '68' instead of '86'; Pixner et al., 2011).

Even adult speakers of inverted languages may be affected by the place-value structure of number words. Nuerk et al. (2005) tested adult German and English speakers on a number comparison task and observed a strong unit-decade compatibility effect in German but not in English (Nuerk et al., 2005). Participants comparing numbers in which decade and unit comparisons led to opposing responses (e.g., 28/61: 2<6 but 8>1) took longer to respond than participants comparing numbers in which decade and unit comparisons led to the same responses (e.g., 21/28: 2<6 and 1<8) and this effect was found to be greater in German than in English (see Nuerk et al., 2005 for a discussion; see also Nuerk, Weger, & Willmes, 2002). In another study, Italian children (speakers of a noninverted language) performed more accurately than Austrian children (speakers of an inverted language) on a 0-100 NLE task. Austrian children were particularly erroneous when estimating numbers with a large interdigit distance (82) and significantly underestimated numbers like ‘82’ where mixing up tens and units digits would result in a smaller number (i.e., ‘28’), while overestimating numbers like ‘27’ where mixing up tens and units would result in a larger number (i.e., ‘72’; Helmreich et al., 2011).

One remaining question is whether other types of estimation error may also differ for different number word systems. In particular, do speakers of an inverted language demonstrate a left digit effect in number line estimation, as speakers of English do? Potential inversion-related differences in left digit effects have not yet been investigated. Here we ask whether the left digit effect in 0-100 NLE occurs in the context of inverted number words. We hypothesized that a greater left digit effect might be observed for speakers of a non-inverted language (English) compared to speakers of an inverted language (Dutch) due to the differences in the place-value structure of number words across these languages. That is, if the left digit effect is due at least in part to the ordering of individual digits in spoken number words, we should find one or both of the following: (1) the left digit effect should be significantly greater for Dutch-English bilingual participants’ NLE performance in English, compared to their performance in Dutch, or (2) the left digit effect should be significantly greater for monolingual English participants’ performance compared to bilingual participants’ performance in English.

In Experiment 1, Dutch-English bilingual adults completed two standard 0-100 NLE tasks, one each in English and Dutch. Target numbers were read aloud by the experimenter and participants marked their locations on a response line. Differences between placements of numbers on either side of decade boundary values (e.g., 39/41) were computed; a left digit effect is indicated when placement differences are greater than the true difference between target numbers. In Experiment 2, to ensure that our findings were not due to non-linguistic characteristics of our bilingual sample and to compare bilingual and monolingual adults' performance, monolingual English-speaking adults completed the 0-100 NLE task in English.

Experiment 1

Method

Participants

Dutch-English bilingual adults (*M* = 39.42 years, range = 21-60 years, 23 female) were recruited in the Boston area through Dutch cultural events and social media (*n* = 39), or in the Netherlands (*n* = 3). Participants rated their language proficiency in production and understanding on a four-point scale (1=limited, 2=intermediate, 3=advanced and 4=native). Two participants were excluded for rating their Dutch language production *and* understanding to be below advanced levels. Participants in our final sample (*N* = 40) had advanced or native production (English: 3.43, Dutch: 3.83) and understanding (English: 3.53, Dutch: 3.95), learned English around age 9 (range = 0-13) and learned Dutch around age 1.[i] Participants had an average of 13.69 years of education in the Dutch language (range = 0-36) and knew an average of 1.85 additional languages or dialects (range = 0-6). Of the participants who completed a demographic questionnaire,[ii] 88% identified as Caucasian, 6% as Native American and 6% as

African American. Furthermore, 88% reported having completed a master's degree and 69% reported annual household incomes over $50,000.

Stimuli

For each trial of the paper-and-pencil NLE task, participants were shown a piece of paper (5.5" x 8.5") with a black number line measuring 24.5 cm across the middle of the page. The line's endpoints were labeled with 0 (left) and 100 (right). Participants completed two blocks of 30 trials each (60 total trials), one block in English and one block in Dutch. Block order was counterbalanced, with individuals randomly assigned to order conditions. The following target numbers were randomly presented once per block: 2, 4, 8, 12, 17, 18, 22, 23, 29, 31, 36, 38, 42, 47, 49, 51, 58, 59, 61, 62, 69, 71, 74, 78, 82, 86, 88, 92, 97, 99.

Procedure

Target numbers were read aloud by the experimenter in the language corresponding to the block (English or Dutch). On each trial, the experimenter said, "If 0 goes here and 100 goes here, where does [X] go?" and participants indicated, with a pencil mark, the location of the number on the line. Participants were verbally debriefed and received a small thank-you gift. Researchers converted each mark to a number between 0-100 corresponding to its location on the number line by measuring its position from 0 using a ruler.

Analyses

Analyses were preregistered (https://aspredicted.org/85uf2.pdf) unless otherwise noted. Individual estimates that differed from the group mean for a given target number by more than 2 SDs were excluded from each block (3.80% of Dutch trials, 4.03% of English trials). To determine whether estimates exhibited a left digit effect, we calculated individual difference scores.[iii] For each participant and within each block, we calculated *placement for larger numeral - placement for smaller numeral* (e.g., placement for 31 minus placement for 29) for nine pairs of target numbers: 8/12, 18/22, 29/31, 38/42, 49/51, 59/61, 69/71, 78/82, 88/92. We then subtracted the true difference between pairs (e.g., we subtracted 2 from the difference between the estimates for 31 and 29). We averaged difference scores across nine pairs to calculate one mean decade difference score per participant. We planned to exclude participants if more than five pairs were missing due to outlier removal, but no participants required exclusion. Decade difference scores greater than zero, i.e., paired numbers placed too far apart with the larger number placed to the right, indicate a left digit effect.

For each block, we also calculated percent absolute error (PAE) as a measure of overall accuracy by dividing the absolute difference between the estimated location and the actual location of a number by the numerical range, then multiplying by 100 to give a percentage.

Results

Overall accuracy did not differ between blocks (English PAE: 3.09%, Dutch PAE: 3.17%; $t(39) = -.59$, $p = .56$) and there was no order effect ($F(1, 38) = .31$, $p = .58$). Figure 1 depicts individuals' mean decade difference scores and Figure 2 shows bias in the placements of individual target numerals. Decade difference scores did not differ from 0 in English ($M = 0.23$, $SD = 1.70$; $t(39) = 0.84$, $p = .20$, two-tailed) or in Dutch ($M = -0.21$, $SD = 1.66$; $t(39) = -0.79$, $p = .78$, two-tailed),[iv] providing no evidence of a left digit effect in this Dutch bilingual sample. Bayes factors provide positive support (Kass & Raftery, 1995) for the null finding that Dutch-English bilingual participants did not exhibit a left digit effect either in English ($BF_{01} = 4.22$) or in Dutch ($BF_{01} = 4.39$). To better assess the role of language, we compared decade difference scores from blocks with target numerals spoken in English vs. Dutch, finding no significant effect of language ($t(39) = -1.46$, $p = .15$), with a Bayes Factor of $BF_{01} = 2.20$ (weak evidence for the null hypothesis), or block order ($F(1, 38) = 1.18$, $p = .29$).

We were also interested in whether left digit effects emerged in some individuals, so we conducted exploratory analyses to determine whether the difference scores for each target pair were positive more often than predicted by chance at the individual level.[v] Two participants (*p*s < .002) showed a significant left digit effect in English; one additional participant ($p < .002$) showed a significant left digit effect in Dutch. Given the very small number of trials contributing to these analyses of individual participants' responses, though, results should be interpreted with caution.

Difference scores were not correlated (in either block, English or Dutch) with reported levels of English or Dutch language fluency, respectively (all *r*s >.100). This could be due to limited variability in reported levels of fluency.

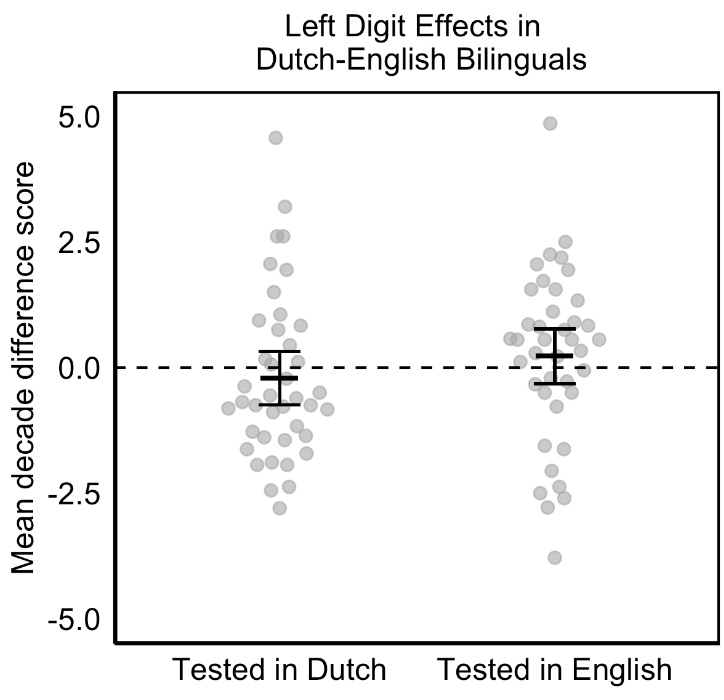

Figure 1. Dutch-English bilinguals' individual mean decade difference scores (see Analyses), are shown for 0-100 number line placements when tested in Dutch (left) and in English (right). Mean

decade difference scores for each participant (grey dots) and group means and 95% confidence intervals (black lines) are shown. Difference scores of zero result when the larger numeral in the pair is placed the veridical number of units to the right of the smaller (e.g., if 31 is placed 2 units to the right of 29). Difference scores are positive when the larger numeral is placed more than the veridical number of units to the right of the smaller (e.g., if 31 is placed more than 2 units to the right of 29). Difference scores greater than zero indicate a left digit effect.

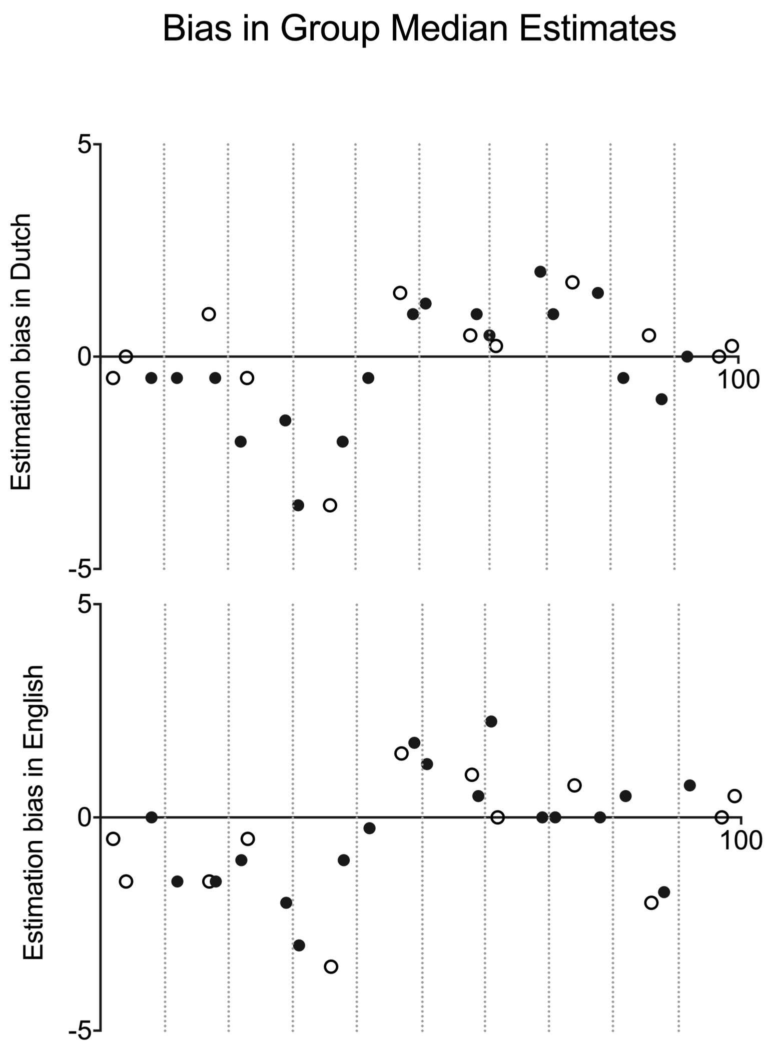

Figure 2. Bias in group median estimates for paired target numerals located on either side of decade boundaries (filled circles, e.g., 38 and 42) and for additional target numerals (open circles) on the 0-100 number line when tested in Dutch (top) and in English (bottom). Dotted lines indicate a decade boundary (e.g., 40). Accurate estimates fall on $y = 0$, positive y values indicate placement too far to the right, and negative y values indicate placement too far to the left.

Experiment 2

Bilingual adults in Experiment 1 did not show a left digit effect when tested either in Dutch or in English on the 0-100 NLE task. Experiment 2 was designed to allow us to compare bilingual and monolingual adults' performance, and to ensure that the Experiment 1 findings did not arise from non-linguistic characteristics of our bilingual sample. Monolingual English-speaking adults, similar to our bilingual sample in SES, education, and age, completed one block of the 0-100 NLE task in English.

Method

Participants

Twenty monolingual English-speaking adults (M = 39.43 years, range = 22 – 71 years, 10 female) were recruited from Boston College. Because our goal was to recruit participants matched in age, education, and income to participants in Experiment 1, we restricted our sample to graduate students, professors, and staff in the Psychology Department. All identified as Caucasian, 75% reported having a master's or doctoral degree, and 75% reported an annual household income over $50,000. All reported learning English from birth and having native English language proficiency.

Stimuli, Procedure, Analyses

The procedure was identical to that of Experiment 1, except that participants only completed one block of the English NLE task. Individual estimates that differed from the group mean by more than 2 SDs were excluded (5.5% of trials).

Results

Overall accuracy was similar to previous adult studies (PAE: 2.76%; Williams, et al., 2020). Consistent with previous findings with English-speaking adults, decade difference scores were significantly greater than zero (M = 1.01, SD = 1.60; $t(19)$ = 2.82, p = .005, d = 0.63).[vi] Participants placed target numbers around decade boundaries (like 39 and 41) farther apart than their magnitudes would predict, exhibiting a left digit effect (see Figure 3 for mean deacde difference scores, and Figure 4 for bias in the placements of individual target numerals). Bayes Factors provide positive support for this finding, BF_{10} = 4.77. To ensure that these findings are driven by left digit effects rather than potentially reflecting an overestimation of higher numbers in pairs of nearby target numerals in general, we conducted post-hoc analyses using average difference scores for eight pairs of numbers falling within decades (4/8, 36/38, 47/49, 71/74, 74/78, 82/86, 86/88, and 97/99).[vii] Difference scores for these within-decade pairs did not differ from 0 (M = -0.56, SD = 1.21; $t(19)$ = -2.08, p = .052, two-tailed). Further, a paired sample t-test comparing difference scores for critical between-decade pairs (like 39 and 41) and within-decade pairs (like 36 and 38) revealed that participants differentiated pairs of numbers between decades significantly more than pairs of numbers within a decade (like 36 and 38), $t(19)$ = -2.66, p = .016, d = 0.59.

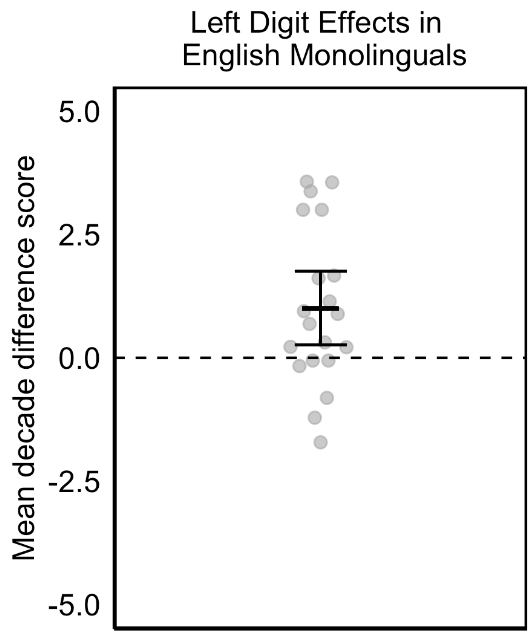

Figure 3. English monolinguals' individual mean decade difference scores, a measure of left digit effects (see Analyses), are shown for 0-100 number line placements. Mean decade difference scores for each participant (grey dots) and group means and 95% confidence intervals (black lines) are shown.

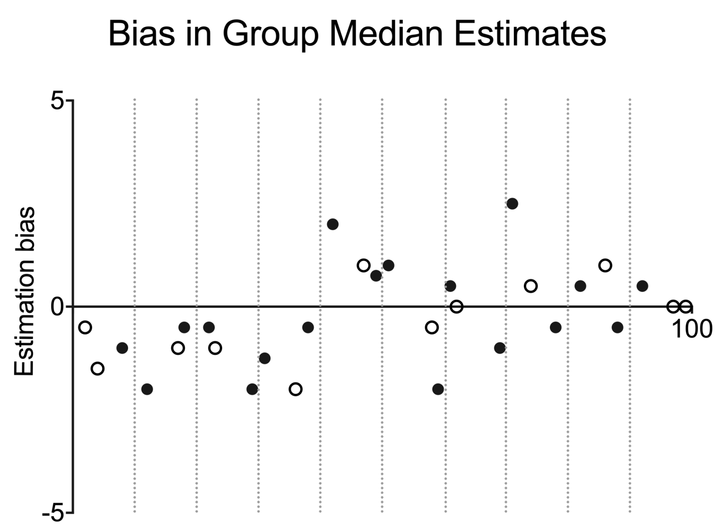

Figure 4. Bias in group median estimates for paired target numerals located on either side of decade boundaries (filled circles) and for additional target numerals (open circles) on the 0-100 number line. Dotted lines indicate a decade boundary. Accurate estimates fall on $y = 0$, positive y values indicate placement too far to the right, and negative y values indicate placement too far to the left.

This finding of a left digit effect replicates past work (for 0-100 NLE, Williams, et al., 2020; see also Lai et al., 2018 for 0-1000 NLE). As in Experiment 1, we conducted exploratory analyses to investigate whether left digit effects emerge at the individual level. One participant ($p = .002$) showed an effect at the individual level. Given the small number of trials this should be interpreted with caution.

Combined Results

The following analyses were not preregistered. We next compared performance across experiments between bilingual (total $N = 40$) and monolingual adults ($N = 20$), restricting analyses to the first block only (resulting in three independent groups of 20 participants each). A One-Way ANOVA suggested that difference scores differed by Language Group (1=bilingual: English, 2=bilingual: Dutch, 3=monolingual: English) ($F(2,57) = 5.44$, $p = .007$, $\eta^2 = .160$). Post-hoc comparisons using Tukey's HSD tests suggested that bilingual participants tested in Dutch ($M = -0.48$, $SD = 1.53$) had smaller difference scores than monolingual English participants ($M = 1.01$, $SD = 1.60$, $p < .01$); neither group differed from bilingual participants tested in English ($M = 0.43$, $SD = 1.57$, ps > .17). We next compared overall accuracy across language groups. A

One-Way ANOVA suggested that PAE did not differ by Language Group ($F(2,57) = 0.91$, $p = .41$). Monolingual and bilingual adults produced similar overall error, but only monolinguals demonstrated a left digit effect.

General Discussion

In previous studies, English-speaking adults and children placed numbers with different leftmost digits but very similar magnitudes (e.g., 38/42 or 398/402) farther apart than would be predicted by their magnitudes alone (Lai et al., 2018; Williams, et al., 2020). Here we investigated a potential influence of language on this left digit effect, using a 0-100 number line task to assess whether the inversion property of Dutch number words might lead to a reduced or eliminated effect. If the left digit effect in English is in part due to the structure of English number words (in which "41" is "forty-one"), then a left digit effect might not emerge in languages with number word systems characterized by an inverted place-value structure of tens and units such as Dutch (in which "41" is "eenenveertig" – one and forty). We also considered the possibility that Dutch-English bilingual adults might show a greater left digit effect when tested in English (using noninverted number words) compared to their performance in Dutch (using inverted number words). Consistent with previous reports, monolingual English-speaking adults (Experiment 2) demonstrated a left digit effect: they placed numbers with different leftmost digits too far apart on a 0-100 scale (e.g., 41 was placed more than 2 units to the right of 39). Bilingual adults (Experiment 1) did not display a left digit effect in either English or Dutch. The findings here do not converge neatly with past work attributing more erroneous performance to inversion of number words. Rather, they suggest that in some cases, inversion could prove advantageous in NLE tasks by reducing or eliminating the left digit effect.

Combined analyses revealed no differences in overall accuracy across participant groups and confirmed that participants were influenced by leftmost digits to varying degrees: monolingual English speakers, but not Dutch-English bilinguals, showed a significant left digit effect, and bilinguals performed more like English-speaking monolinguals when tested in English compared to their performance in Dutch. Within-participant comparisons of Dutch-English bilingual adults revealed comparable performance in Dutch and English, suggesting that knowing an inverted language could potentially influence performance regardless of the presented number word structure. This is remarkable given that the majority of bilingual participants resided in the United States and thus, we assume, had greater exposure to noninverted English number words on a daily basis. An interesting avenue for future work would be to explore the relationship of language exposure, fluency, and degree of left digit effects in NLE. Follow-up work might investigate whether the effect is limited to individuals whose native language consists of noninverted number words.

One limitation of the current work is that we could not control for language exposure. There was likely large variability in how frequently our Dutch-English bilingual participants spoke Dutch. These participants had, on average, 13.69 years of education in the Dutch language, suggesting that most completed the majority of their schooling in Dutch and likely also received mathematical instruction in Dutch. We did not collect data on the frequency with which participants spoke Dutch, the language in which they spontaneously counted, or their language of instruction for mathematical learning, and so we cannot rule out the possibility that these characteristics of language use and exposure play an important role. The age of acquisition of English in our bilingual sample was also variable (0-13 years) and so we cannot speak to any differential impact of age of acquisition or fluency on performance. Future research could explore the influence of these factors.

In Experiment 1, our sample size was limited by available resources for recruitment of Dutch-English bilinguals; Experiment 2's recruitment of English monolinguals was restricted to achieve a comparable sample. While larger sample sizes would be ideal for future studies, our sample sizes are unlikely to have prevented us from detecting true left digit effects given the robust and large effects previously observed in NLE (Lai et al., 2018, Williams, et al., 2020). However, the present findings may be specific to two-digit numbers in the 0-100 range, and may not generalize to larger or less familiar ranges. Do left digit effects emerge for Dutch-English bilinguals in NLE involving three-digit number words (e.g., for numbers falling around hundreds boundary values like 299 and 301), as they do for English-speaking monolingual adults? Unlike two-digit number words, three-digit number words are not inverted in Dutch and thus may be one interesting avenue for future investigations of the potential role of language in non-verbal number line task performance. Another important consideration is that differences in performance were only observed across groups that could not be randomly assigned (monolinguals and bilinguals). We think it is unlikely that our findings were driven by differences in age, math experience, or overall intelligence because our samples were of comparable age, education and income levels. Nevertheless, it is also possible that bilingual participants in Experiment 1 had greater math skills and perhaps relied less heavily on leftmost digits than monolingual participants in Experiment 2. Some initial evidence against this idea comes from a recent study investigating the relationship between the left digit effect and performance on the SAT (a standardized test published by the College Board; see College Board, 2018). In particular, researchers found that the left digit effect in a speeded 0-1000 NLE task with English speaking adults was not predicted by SAT math scores, but was related to SAT verbal scores under some conditions (Williams, Paul, Barth, & Patalano, in press).

Multiple existing models of multi-digit number processing could potentially account for the observed left digit effects in numerical judgments. One possibility, consistent with holistic models of multi-digit number processing (Dehaene et al., 1990), is that when converting from numerical symbols to magnitudes, a left digit anchoring effect arises such that the encoded magnitude becomes anchored on the leftmost digit, resulting in estimates that are disproportionally influenced by that digit (Thomas & Morwitz, 2005; 2009). Another possibility, consistent with decomposed place-value models of multi-digit number processing (McCloskey, 1992), is that when multi-digit numbers are encoded, representations of the individual digits are weighted in order of the place-value structure of Arabic numerals, such that greater weight is placed on the leftmost digit (e.g., the tens place digit in two-digit numbers and the hundreds place digits in three-digit numbers). Thus, differences in the perceived magnitude of numbers with a different leftmost digit are exaggerated (Huber et al., 2016). In our view, findings to date of left digit effects in NLE do not distinguish between theories of the cognitive processing of multidigit numbers, but it is possible that further elucidation of the sources of left digit effects may do so. These findings are, however, directly relevant to ongoing debates about the cognitive processes that underlie NLE performance, further supporting the idea NLE placements are dependent not only upon target numerals' overall magnitudes but on the specific digits that comprise them.

This work contributes to the growing body of evidence that in multiple subdomains of numerical cognition, children and adults use digit level information to inform their numerical estimates. Both greater estimation error (e.g., mixing up the tens and units digits for numbers with a large interdigit distance like ’28,’ and instead locating ‘82’ on a number line; Helmreich et al., 2011) and trouble with unit-decade integration into the place-value structure of the Arabic number system have been observed considerably more in inverted languages (Nuerk, et al., 2005; Helmreich et al., 2011). The present study suggests that inversion may actually prove advantageous for some elements of NLE performance. More research is needed to determine the extent to which NLE performance, and in particular the left digit effect, is influenced by linguistic information.The influence of language properties on NLE performance, with specific consideration of language with a tens-to-units place-value structure or inverted units-to-tens structure, is a worthwhile area for future work to explore.

Author Note

We thank the participants who contributed to this study and the Dutch cultural organizations that facilitated. This work was supported by NSF DRL-1920445.

Open Science Statement

This work was formally preregistered. Data are also available on the Open Science Framework here: https://osf.io/b3nzy/?view_only=d409846e1aee4709998192a610a8f242.

References

College Board. (2018). *Inside the SAT Test*. SAT Suite of Assessments. https://collegereadiness.collegeboard.org/sat/inside-the-test

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

Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. C. (2014). Language affects symbolic arithmetic in children: the case of number word inversion. *Journal of Experimental Child Psychology*, *119*, 17-25. https://doi.org/10.1016/j.jecp.2013.10.001

Hamdan, N., & Gunderson, E. (2017). The number line is a critical spatial-numerical representation: Evidence from a fraction intervention. *Developmental Psychology, 53*, 587-596.

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H. C., & Moeller, K. (2011). Language effects on children's nonverbal number line estimations. *Journal of Cross-Cultural Psychology*, *42*, 598-613. https://doi.org/10.1177/0022022111406026

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

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

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*, 107–157. https://doi.org/10.1016/0010-0277(92)90052-J

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

Nuerk, H.-C., Moeller, K., & Willmes, K. (2015). Multi-digit number processing. In *The Oxford Handbook of Numerical Cognition* (pp. 106–139).

Nuerk, H. C., Weger, U., & Willmes, K. (2002). A unit-decade compatibility effect in German number words. *Current Psychology Letters: Behaviour, Brain & Cognition*.

Nuerk, H.-C., Weger, U., & Willmes, K. (2005). Language effects in magnitude comparison: Small, but not irrelevant. *Brain and Language*, *92*, 262–277. https:/doi.org/10.107/j.bandl.2004.06.107

Pixner, S., Zuber, J., Heřmanová, V., Kaufmann, L., Nuerk, H. C., & Moeller, K. (2011). One language, two number-word systems and many problems: Numerical cognition in the Czech language. *Research in Developmental Disabilities*, *32*, 2683–2689. https:/doi.org/10.1016/j.ridd.2011.06.004

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773-795.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*, 375-394.

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*, 1467-1484.

Siegler, R. S. (2016). Magnitude knowledge: the common core of numerical development. *Developmental Science*, *19*: 341–361.

Thomas, M., & Morwitz, V. (2005). Penny Wise and Pound Foolish: The Left- Digit Effect in Price Cognition. *Journal of Consumer Research*, *32*, 54-64. https://doi.org/10.1086/429600

Thomas, M., & Morwitz, V. (2009). Heuristics in numerical cognition: Implications for pricing. *Handbook of pricing research in marketing*, 132-149.

Williams, K., Paul, J., Barth, H., & Patalano, A.L. (in press). Individual differences in numerical estimation performance: Standardized test scores and the left digit effect. *Brain and Behavior.*

Williams, K., Zax, A., Patalano, A.L., & Barth, H. (2020). Left digit effect in numerical estimation across development. [Manuscript submitted for publication]. Department of Psychology, Wesleyan University.

Xing, C., Zax, A., George, E., Taggart, Xing, C., Zax., A., Taggart, J., Bass, I., George, E., & Barth, H. (2020). Numerical estimation strategies are correlated with math ability in school-age children. [Manuscript submitted for publication]. Department of Psychology, Wesleyan University.

Xing, C., Paul, J., Zax, A., Cordes, S., Barth, H., & Patalano, A.L. (2019). Probability range and probability distortion in a gambling task. *Acta Psychologica*, 197, 39-51.

Zuber, J., Pixner, S., Moeller, K., & Nuerk, H.-C. (2009). On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity. *Journal of Experimental Child Psychology*, *102*, 60–77. https://doi.org/10.1016/j.jecp.2008.04.003

[i] All participants included in our final sample learned Dutch from birth, with the exception of one participant who learned Dutch at age 21.

[ii] Because the questionnaire was distributed after participants completed the study, this information was only obtained for 16/40 of participants.

[iii] We used a slightly different method from that of Lai et al. (2018), who assumed that numerals like 399 and 401 should be, on average, placed in the same location on a 0-1000 scale if there is no left digit effect. We made no such assumption for the 0-100 task because it is possible that numbers like 29 and 31 are distinguishable on a 0-100 scale; this method is consistent with that of Williams et al, 2020.

[iv] Excluding the pair 8/12 to restrict our analyses to 2-digit numerals did not change the results: differences scores were not greater than 0 in either block (English: $p = .183$; Dutch: $p = .273$).

[v] The pair 8/12 was excluded from these analyses because it does not consist of two numerals with leftmost digits.

[vi] Findings remained the same with the pair 8/12 excluded, with difference scores > 0 ($M = 1.15$, $SD = 1.60$; $t(19) = 3.21$, $p = .005$, $d = 0.72$).

[vii] For this analysis we only included pairs of numbers within a decade if their true difference was 2, 3, or 4 units apart (analogous to our analysis of critical pairs of numbers around decade boundaries, e.g., 29/31).