Beyond numerical quantity:

A study of the sub-base-five effect in single-digit comparison

Samuel Lepoittevin[1,2], Laurie Geers[1,2], & Michael Andres[1,2]

[1] Psychological Sciences Research Institute, Université catholique de Louvain, Belgium.

[2] Institute of Neuroscience, Université catholique de Louvain, Belgium.

Corresponding author:

Samuel Lepoittevin

Institut de Recherche en Sciences Psychologiques

Université catholique de Louvain

Place Cardinal Mercier, 10

1348 Louvain-la-Neuve

Belgium

Email: samuel.lepoittevin@uclouvain.be

Author information:

Laurie Geers : laurie.geers@uclouvain.be

Michael Andres: michael.andres@uclouvain.be

Credit authorship contribution statement

Samuel Lepoittevin: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. Laurie Geers: Formal Analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. Michael Andres: Conceptualization, Methodology, Formal Analysis, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgements

The authors thank Alain Guillet for his help with the data analyses.

Declaration of interest

None.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: SL is a research fellow, LG is a post-doctoral researcher, and MA is a research associate at the Fonds National de la Recherche Scientifique (FRS-FNRS, Belgium). This research was supported by grant T.0006.20 (FRS-FNRS, Belgium) and grant ARC21-26-112 from the Fédération Wallonie Bruxelles (FWB, Belgium).

Abstract

Previous findings suggest that the mental representations involved in symbolic number comparison carry information not only about the numerical quantity they denote, but also about higher-order properties. While the time to compare pairs of digits increases continuously with numerical size, former studies also revealed a categorical effect characterised by a step increase in response times for pairs with at least one digit > 5 compared to those including only digits ≤ 5. In line with the embodied cognition view, this step increase was interpreted as an inheritance from finger counting practices. This "sub-base-five" effect would find its origin in the need to generate the mental representation of two hands to process numbers > 5, whereas one hand is sufficient to represent numbers ≤ 5. We provide evidence against this interpretation using new analyses designed to uncouple the continuous effect of numerical size and the categorical sub-base-five effect. We demonstrate that the time to compare two Arabic numerals is primarily determined by numerical size, with no steeper increase between pairs with or without number > 5. We explain that the sub-base-five effect in digit comparison was likely an artefact of the modelling method previously used to control for numerical size. We nevertheless identified a robust response time increase when comparing 5 and 7, which could not be accounted for by numerical size. These findings question the existence of a sub-base-five effect, which was considered as a marker of embodied cognition, but prompt interest for a previously unnoticed effect emphasizing the special status of particular prime numbers 5 and 7 in symbolic number comparison.

Keywords: Sub-base-five effect; Symbolic number comparison; Single-digit; Numerical size effect; Embodied cognition.

Introduction

Humans are equipped with an approximate number system, available from birth and shared with other species, which enables them to process numerosities (Dehaene, 1997; Nieder, 2016; Nieder & Dehaene, 2009). The dominant view assumes that the same representation underlies the processing of the numerical magnitude derived from symbolic numbers (Dehaene, 2001, 2007; Piazza et al., 2007). Behavioural studies indeed showed that both non-symbolic comparison and symbolic comparison give rise to similar effects, which have been described as the signature of this number system representing quantities as overlapping distributions of activation along an internal continuum (Dehaene, 2007; Dehaene & Changeux, 1993). The numerical distance effect refers to the observation that the comparison of two numbers takes less time as their numerical distance increases (Moyer & Landauer, 1967), presumably due to less overlap between activation distributions located further apart on the mental number line (Dehaene & Changeux, 1993). The numerical size effect refers to the observation that, when the numerical distance is fixed, the response time increases with numerical size (Parkman, 1971), presumably due to an increasing overlap of activation distributions on the side of large numbers (Dehaene & Changeux, 1993). However, this dominant view is still debated (Núñez, 2017; Reynvoet & Sasanguie, 2016; Rips et al., 2008). The effect of numerical size and distance have received alternative interpretations in frameworks that do not imply a common representation for symbolic and non-symbolic numbers (Krajcsi, 2017; Marinova et al., 2021; Verguts & Van Opstal, 2014). Aside from the debate surrounding the origin of the distance and size effects, other experimental results suggest that the processing of symbolic number magnitude involves representations whose content would not be reducible to mere numerical quantity. Indeed, the study of response times in digit pair comparison has revealed another effect, indicating the influence of a sub-base-five organisation as participants process symbolic number magnitude (Domahs et al., 2010).

The study of Domahs and colleagues (2010) showed that participants take more time to identify the largest or the smallest of two digits when at least one of them is larger than 5. The authors argued that this “sub-base-five effect” could not be explained by numerical distance, as it was kept fixed within the pairs (i.e. two units), nor by numerical size, as the analyses specifically focused on the residuals from the regression of response times on the mean value of the numerals to be compared. They claimed that this effect is inherited from finger counting practices that create a distinction between numbers smaller or equal to 5 (which can be represented on one hand) and numbers larger than 5 (whose finger-based representation requires two hands). Supportive evidence comes from the comparison of the numerical performance of Germans, who typically use both hands to represent numbers larger than 5, and Chinese, who learned to count up to 9 using one single hand. A sub-base-five effect was found in Germans but not in Chinese, suggesting a specific influence of finger counting practices on the response times. The authors proposed that digit comparison would involve finger-based representations, and that the sub-base-five effect would reflect the cost of mentally generating the representations of two hands for numbers > 5 vs. one hand for numbers ≤ 5. While the nature of this process was not further specified, the authors evoked the possibility that it could be based on motor or visual imagery. Hence, the sub-base-five effect was considered as evidence that finger counting experience shapes the processing of the magnitude of symbolic numbers, and as a key argument for the idea of embodied numerical cognition (e.g., Berteletti & Booth, 2016; Fischer & Brugger, 2011; Roesch & Moeller, 2015; Sixtus et al., 2023; Soylu et al., 2018).

The present study aims to provide a stringent test of the hypothesis that the processing of symbolic number magnitude involves a representation organised according to a sub-base-five structure. Indeed, a detailed examination of the procedure used by Domahs and colleagues (2010) to reveal the sub-base-five effect raises questions. The critical issue concerns the collinearity of the sub-base-five effect and the numerical size effect. Indeed, even if the numerical size effect is assumed to influence response times in a continuous manner, while the sub-base-five effect is assumed to do so in a categorical manner, both effects predict an increase in response times for larger pairs. To circumvent this issue of collinearity, Domahs and colleagues (2010) used a regression analysis approach that consisted in removing the variance explained by the mean value of the pairs before comparing the subsequent residuals associated with *No-Break* pairs (i.e., pairs composed of two digits ≤ 5) and *Break-5* pairs (i.e., pairs with at least one digit > 5). The residuals were found to be larger for the Break-5 than for the No-Break pairs, suggesting a sub-base-five effect influencing response times beyond the mere effect of numerical size. While the critical test only concerned pairs of single digits (pairs 1-3 to 7-9), the residuals associated with these pairs were obtained by regressing response times on numerical size over an extended range including two-digit numerals (all pairs from 1-3 to 18-20). This modelling method is based on the implicit assumption that there is no discontinuity in the effect of numerical size on response times measured for single and two-digit numerals. However, such assumption overlooks the fact that, given their lexico-syntactic properties, comparing two-digit numerals involve different processes or strategies than comparing single digits (e.g., see Ganor-Stern et al., 2009; Tzelgov et al., 2015). A close look at the results of Domahs and colleagues (2010) shows for example that pairs 8-10 and 9-11 are associated with a drop in response times, presumably because participants based their judgement on the perceptual difference in length between single-digit and two-digit numerals rather than on the comparison of their magnitude. The discontinuity of the size effect over single- and two-digit numeral comparison is further supported by independent evidence showing that the magnitude of two-digit numbers is processed in parallel for decades and units and is thus not comparable to the holistic processing of single-digit magnitude (Nuerk et al., 2001; Verguts & De Moor, 2005, Zhang & Wang, 2005). We are concerned that the influence of factors other than numerical size on response times may have affected the regression and, consequently, the analysis of subsequent residuals that have not been completely purged of its effect. Because the sub-base-five effect is tested by analysing the residuals in the range of single digits, it would be advisable to obtain these residuals by regressing response times on numerical size in that specific range (Parkman, 1971). Thus, the question arises as to whether the sub-base-five effect could be an artefact of the modelling method used to control for the collinear effect of numerical size in the study of Domahs and colleagues (2010).

In the present study, we varied the range of numerals used to regress response times on numerical size and showed its critical influence on the results of the residual analysis, calling into question previous evidence of a sub-base-five effect. To circumvent the issue of collinearity, we also designed a totally different approach, on the basis that the numerical size effect influences response times in a continuous manner, whereas the sub-base-five effect, if it exists, should influence response times in a categorical manner. As illustrated in Figure 1 (panel A), the sub-base-five effect should be associated with a steeper increase between response times for pairs 3-5 and 4-6 (which are located on either side of the boundary that segregates No-Break and Break-5 pairs), that should contrast with the steady and continuous effect of numerical size alone on the other pairs. Under the hypothesis that there exists a sub-base-five effect affecting differently the processing of numerals ≤ 5 *vs.* > 5, we predict that the difference in response times between pairs 3-5 and 4-6 should be significantly larger than any other difference between adjacent pairs, after taking into account the logarithmic nature of the size effect (Brysbaert, 1995; Dehaene, 2007).

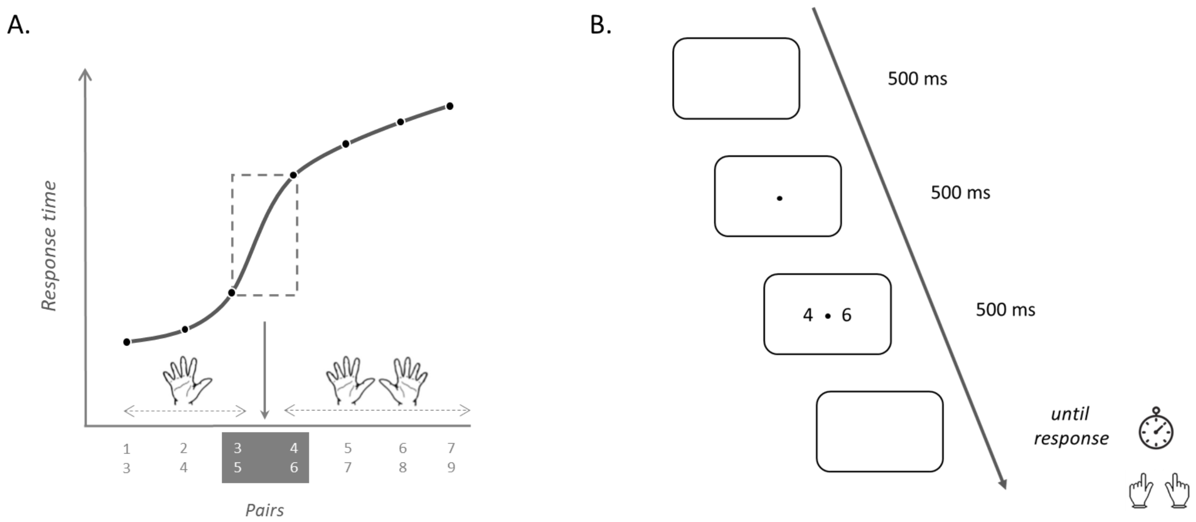

Figure 1. A. Predicted distribution of the response times (RTs) as a function of the single-digit pairs under the hypothesis of the existence of a sub-base effect. The sub-base-five effect is characterised by a slowdown affecting categorically the processing of digits larger than five, due to the cost of mentally representing them on two hands (*vs.* one for digits ≤ 5). It should thus result in a steeper increase in RTs between pairs 3-5 and 4-6 than between the other adjacent pairs, where the only source of RT increase is the continuous size effect. **B.** Time course of a trial in the Arabic numeral comparison task. Numerals ranged from 1 to 20 in Experiment 1, and from 0 to 10 in Experiment 2.

Experiment 1

1.1 Method

1.1.1 Participants

Participants were recruited among students from the Université catholique de Louvain (Belgium). The inclusion criteria were: (1) having no history of psychiatric, neurological or learning disorders, (2) being a native French speaker and (3) being under 40 years of age. Studies have shown that the counting strategy in the participants' culture of origin consists of raising the fingers of one hand to count to 5, and then continuing with the other hand to reach 10 (Lindemann et al., 2011). Forty-seven students participated and seven were excluded a priori because they did not meet the inclusion criteria. One additional participant was excluded a posteriori because they committed more than 20% of errors, most of them on pair 5-7. The final sample included 39 participants (37 females; 2 males; 2 left-handed; mean age ± *SD*: 19 ± 1 years). The minimum sample size required to conduct a two-tailed paired t-test comparing the residuals for the Break-5 and for the No-Break pairs was estimated at 38 participants, on the basis of a power analysis using G*Power (version 3.1.9.7), with dz = 0.5 (moderate effect size; Domahs et al., 2010), $\alpha = .05$, and power = .85. A two-tailed *t*-test was chosen on the basis of strict statistical criteria. Participants were all naïve about the hypotheses being tested. The experiment was conducted with the approval of the ethical committee of the Psychological Sciences Research Institute of UCLouvain (Belgium). All procedures complied with the ethical standards set out in the Declaration of Helsinki.

1.1.2 Apparatus and Stimuli

The experimental task was programmed using PsychoPy (Version 2022.2.1; Pierce et al., 2019) and run on a PC equipped with a 23.8-inch LCD screen (1920 x 1080 pixels; refresh rate: 60 Hz). The eye-screen distance was 60 cm. As in the previous study by Domahs et al. (2010), the stimuli consisted of pairs of Arabic numerals that were two numerical units apart (for example, 1-3 or 5-7). All numerals between 1 and 20 were used to compose 18 pairs presented either in ascending (1-3 to 18-20) or descending (3-1 to 20-18) order, giving a total of 36 stimuli. Numerals (font: *Open Sans*; height: 1.5° of visual angle) were displayed side by side horizontally, in white on a grey background, each at 2° from the centre (marked by a white fixation dot).

1.1.3 Task and Procedure

Each trial began with the display of a blank grey screen for 500ms. Then, a white central fixation dot (0.4° diameter) appeared for 500ms. Next, the pair of numerals was presented for 500ms, followed by another blank grey screen that remained until a response key was pressed (Figure 1, panel B). The next trial started immediately after the response. Participants were instructed to look at the fixation dot throughout the experiment and to respond to the presentation of pairs of numerals, as quickly and accurately as possible, by pressing the response key on the side of the largest number in one half of the experiment and on the side of the smallest number in the other half of the experiment (response keys were ‘s’ or ‘l’, both marked with a coloured sticker on the computer keyboard). The small/large instruction changed halfway through the experiment and order was counterbalanced across participants. In total, participants went through 720 trials: i.e., 2 runs (one per instruction) composed of 5 blocks of 36 stimuli repeated 2 times. Each run was preceded by eighteen practice trials, with written feedback about accuracy, to allow participants to familiarise themselves with the given instruction. Participants took a short break between blocks, and a longer one between the two runs. They were reminded to place their index fingers on the two response keys before starting or resuming the task.

This procedure faithfully replicates the original study of Domahs and colleagues (2010), with the following exceptions: (1) the number of trials was 720 *vs*. 432 in the original study (and 36 *vs.* 72 practice trials in the original study); (2) the stimulus display time was 500ms *vs*. maximum 2s in the original study; (3) the transition with the next trial was automatic vs. self-paced in the original study; and lastly, (4) the fixation cross preceding the stimulus presentation lasted 500ms *vs*. 200ms followed by a 200ms blank screen in the original study. The reason for these changes was to increase the number of trials, while maintaining a sustained pace to keep participants engaged throughout the experiment.

1.1.4 Data Analysis

The response time analyses were performed on correct trials (mean participants' accuracy ± *SD*: 93% ± 3%), after filtering out extreme values using the median absolute deviation (MAD; Leys et al., 2013), computed individually for all blocks sharing the same instruction (small or large). Then, trials for which response time (RT) deviated by 3 MAD or more from the median RT were excluded. The analyses were performed on the remaining trials after aggregating all the blocks (89% of the dataset). The rationale was to confront the results obtained from the analysis of residuals from regressions of RTs on numerical size, as described in the study of Domahs et al. (2010), with the results obtained from a different approach, specifically testing the existence of a step in RTs for pairs with number(s) larger than 5, as predicted by a categorical sub-base-five effect. We describe both approaches in details below.

Although participants made very few errors, we also conducted exploratory analyses on accuracy. The results replicated those observed in the RT analysis and are presented as supplemental material (see Supplemental 1 – Experiment 1).

Residual analysis approach. This approach strictly replicates the original analysis of Domahs and colleagues (2010). We first regressed the RTs of each participant on the mean value of the two numerals composing the pair, using the following power function: ln(mean RT) = a * ln(mean value of the pair) + b. Then, the standardised residuals were aggregated per participant across Break-5 pairs (i.e., 4-6, 5-7, 6-8 and 7-9) and No-Break pairs (i.e., 1-3, 2-4 and 3-5); pairs with two-digit numerals were not included in the residual analysis. Then, to assess the existence of a sub-base-five effect at the group level, these two sets of residuals were contrasted using a paired *t*-test, and the residuals for the Break-5 pairs were tested against zero using a one sample *t*-test (Domahs et al., 2010). In the first instance, these analyses were conducted on the residuals resulting from regressions modelling the effect of numerical size using data from the whole range of the numeral pairs presented (i.e., from pair 1-3 to pair 18-20) as in Domahs and colleagues' (2010) study. We however suspect that conducting such regressions would fail to provide an adequate fit for the effect of numerical size over the range of single-digit pairs, because of the inclusion of pairs with two-digit numerals. Indeed, the effect of numerical size might not be homogeneous over the range of single-digit and two-digit numeral pairs. To assess whether this was effectively the case, we tested the existence of a remaining correlation between the mean value of the numerals of the pairs and the corresponding residuals (averaged over the participants), using a Pearson correlation coefficient. And, in the second instance, in an attempt to better capture the effect of numerical size over the range of single-digit pairs, we replicated the analyses, but this time on the residuals resulting from regressions performed using data from this range of pairs only (i.e. from pair 1-3 to 7-9). It should be noted that the results remained similar without applying the ln-transformation of the RTs (i.e., using a log-function instead of a power one to model the effect of numerical size; e.g., Brysbaert, 1995; Dehaene, 2007; Krajcsi et al., 2016).

Step-at-five Approach. As an alternative to the *residual analysis approach* used by Domahs and colleagues (2010), we proposed a different method to investigate a categorical sub-base-five effect, by testing the existence of a step in RTs between pairs with and without number(s) larger than 5, i.e., the *Step-at-five approach*. The analysis was restrained to the range of single-digit pairs. The rationale was to consider the difference in RTs between adjacent pairs (dRTs) and to test whether the difference between pairs 3-5 and 4-6 stands out. First, for each participant, we computed the difference in ln-transformed RTs between adjacent pairs (e.g., dRT for the couple of adjacent pairs 3-5 and 4-6 = mean ln(RTs) for pair 3-5 – mean ln(RTs) for pair 4-6). Then, we conducted individual regressions of the dRTs on the mean of the ln-transformed values of the adjacent pairs (e.g., $\frac{\ln(4)+\ln(5)}{2}$ for adjacent pairs 3-5 and 4-6). We added this regression step upon the recommendation of a reviewer to align the present analyses with those of the previous residual approach, which were based on a logarithmic model of the numerical size effect (e.g., Brysbaert, 1995; Dehaene, 2007). This modelling implies that the dRTs should decrease as the mean value of the adjacent pairs increases. Importantly, the collinearity between the numerical size effect and the sub-base-five effect is not a concern in the present approach, as analysing the dRTs circumscribes the impact of the sub-base five effect to a single couple of pairs: i.e, [3-5]~[4-6]. The rationale is the following: if digit comparison is only influenced by the continuous size effect, the dRTs should align with the predictions of the regression model. In contrast, under the hypothesis of a sub-base five effect superimposed on the size effect, the dRTs for the couple [3-5]~[4-6] should stand out, and the residuals associated with this couple should deviate from the predictions in a more important manner than the residuals associated with the other couples. To test whether the residuals associated with the couple [3-5]~[4-6] differed from the predictions of the regression model, they were tested against zero. Then, to test whether the residuals associated with this couple differed more from the predicted values than the residuals associated with the other couples, we performed a one sample *t*-test comparing the absolute mean value of the residuals associated with couple [3-5]~[4-6] to the distribution of the absolute mean values of the residuals associated with the other couples.

1.2 Results

1.2.1 Residual analysis after regression over pairs 1-3 to 18-20

Figure 2 (panel A) represents the mean RTs as a function of the numeral pairs, along with the regression line obtained after regressing RTs on the mean value of all pairs from 1-3 to 18-20 (extended range).

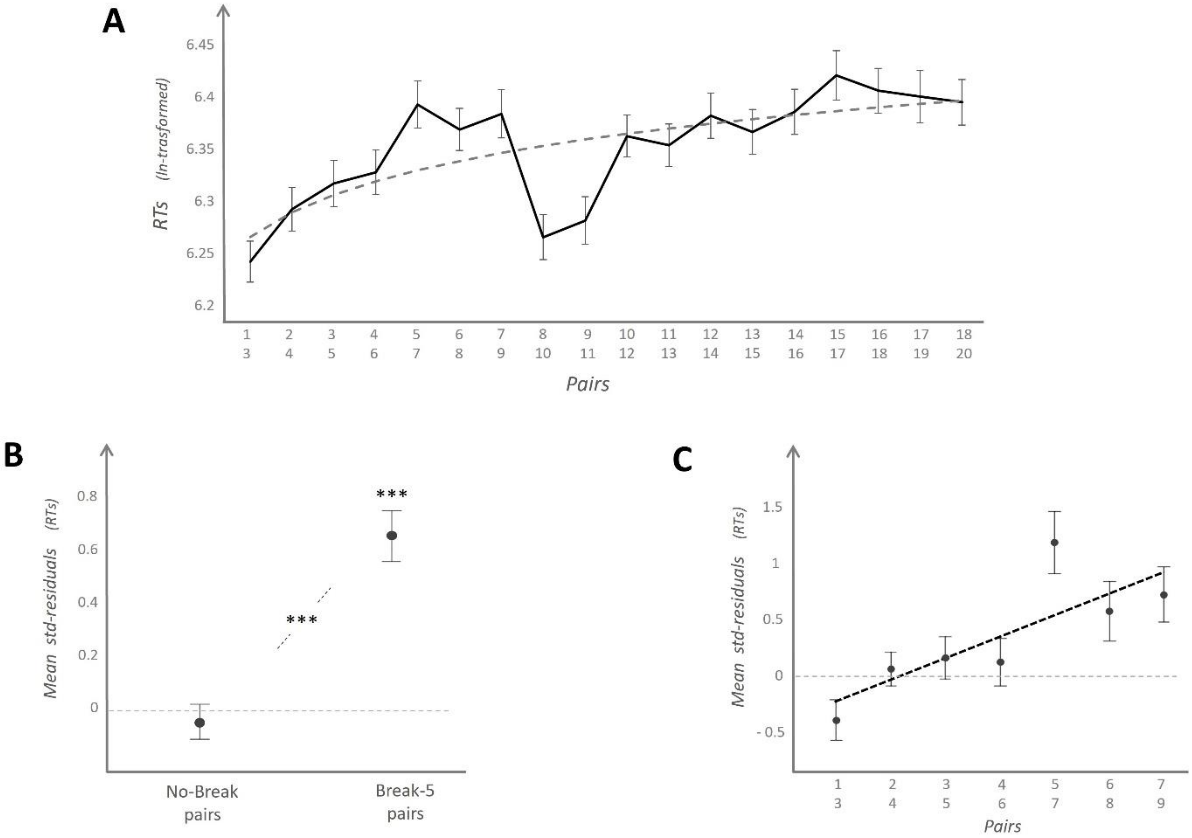

Figure 2. Residual approach after individual regressions over the extended range of pairs. A. Mean RTs (ln-transformed) as a function of the digit pairs. Observed values are connected with black plain lines; error bars represent the standard error of the mean. The dashed grey line represents the model of the effect of numerical size averaged at the group level, after fitting the individual data from all pairs between 1-3 and 18-20 with a function of the type: ln(mean-RT) = a * ln(pair-mean-value) + b. **B.** Mean standardised residuals for the Break-5 and No-Break pairs; error bars represent the 95% IC. The residuals for the Break-5 pairs are superior to zero and superior to those for the No-Break pairs. This constitutes the signature of a sub-base-five effect according to Domahs et al. (2010). **C.** Positive correlation between the mean residuals and the mean value of the numerals to be compared (dashed black line); error bars represent the standard error of the mean. *** $p \leq .001$.

The residuals from the regression analysis for Break-5 pairs were found to be significantly larger than zero ($M \pm SD$: 0.67 ± 0.30), $t(38) = 13.78$, $p < .001$, $d = 2.21$, and also significantly larger than the residuals for No-Break pairs (-0.06 ± 0.21), $t(38) = 9.24$, $p < .001$, $dz = 1.48$. These results are illustrated in Figure 2 (panel B). These results evidence a difference between Break-5 and No-Break pairs, using the same method as Domahs and colleagues (2010). However, a positive remaining correlation was observed between the residuals and the mean values of the pairs of single digits, $r = .80$; $p = .029$, indicating that, when conducted over the range of pairs 1-3 to 18-20, the regressions failed to fully capture the effect of numerical size over these pairs (see Figure 2, panel C).

1.2.2 Residual analysis after regression over pairs 1-3 to 7-9

To improve the fit of the regressions over the range that is critical for testing the sub-base-effect, we replicated the regression analyses using only the data from the pairs of single digits (i.e., pairs 1-3 to 7-9). In other words, the pairs including two-digit numeral(s) were considered as fillers. Figure 3 (panel A) represents the mean RTs as a function of the numeral pairs, along with the regression line obtained after regressing RTs on the mean value of single-digit pairs (over the range of pairs from 1-3 to 7-9).

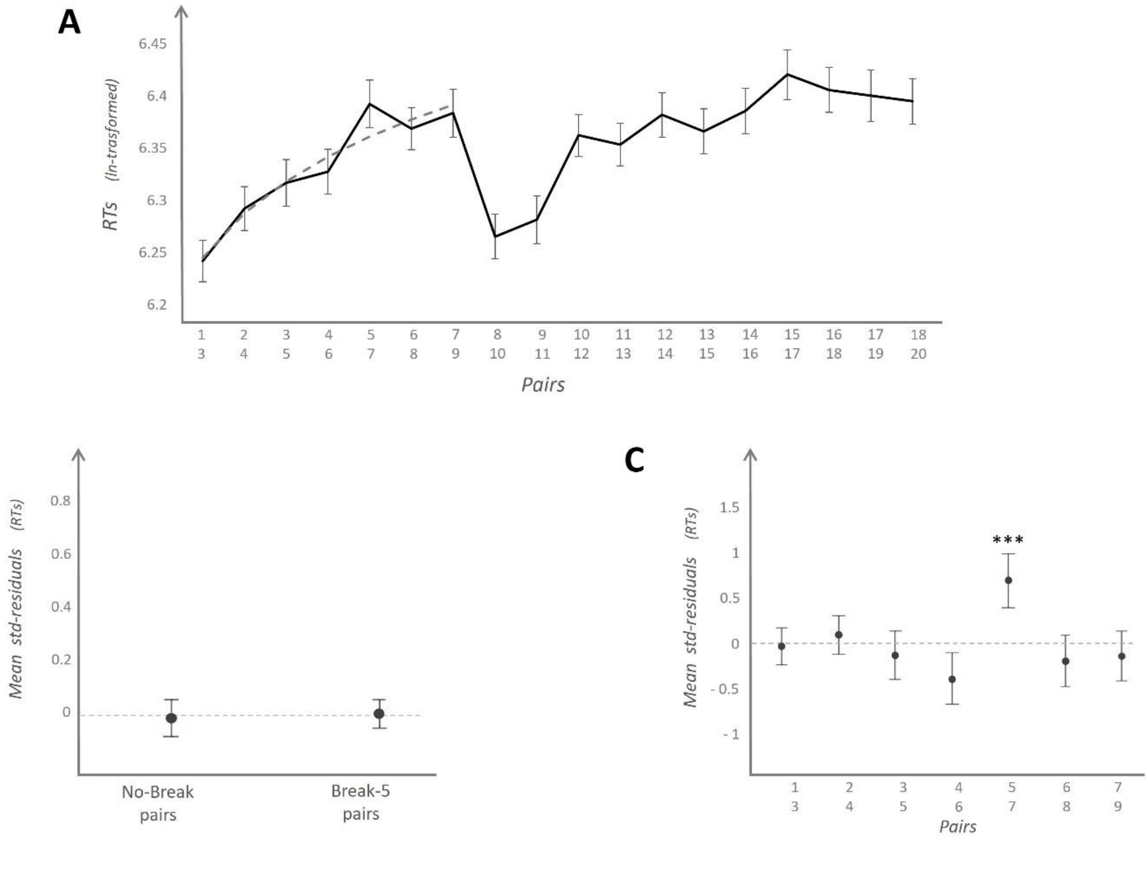

Figure 3. Residual approach after individual regressions over the single-digit range of pairs. A. Mean response times (RTs; ln-transformed) as a function of the digit pairs. Observed values are connected with black plain line; error bars represent the standard error of the mean. The dashed grey line represents the model of the effect of numerical size averaged at the group level, after fitting the individual data from pairs of single digits with a function of the type: ln(mean-RT) = a * ln(pair-mean-value) + b. **B.** Mean standardised residuals for the Break-5 and No-Break pairs; error bars represent the 95% IC. The residuals for the Break-5 pairs do not differ from zero or from those for the No-Break pairs.

C. Mean standardised residuals for the pairs of single-digit numerals; error bars represent the standard error of the mean. The residuals are close to zero, except for pair 5-7, which is associated with larger residuals than the other pairs. *** $p \leq .001$.

When only data from pairs of single digits were included in the regression analyses, no difference was found between the residuals for Break-5 pairs (0.01 ± 0.14) and those for No-Break pairs (-0.01 ± 0.19), $t(38) = 0.33$; $p = .741$. Also, the residuals for Break-5 pairs did not differ from zero, $t(38) = 0.33$; $p = .741$. These results are illustrated in Figure 3 (panel B). No correlation was found between the standardised residuals and the mean value of the pairs of single digits, $r = -.01$; $p = .976$, suggesting that the regressions better captured the effect of numerical size over the pairs of single digits when pairs including two-digit numeral(s) were excluded from the analyses. The distribution of the residuals (averaged by pairs, at a group level) also indicates that the model was better adjusted to the single-digit comparison data when the regression of RTs on numerical size does not mix pairs of single digits and pairs of two-digit numerals (see Figure 3, panel C).

An exploratory analysis using repeated measures ANOVA on the residuals, with pairs as a within-subjects variable, revealed a significant effect of the pair, $F(6,228) = 6.10$, $p < .001$. Post-hoc comparisons indicated that none of the pairs differed significantly from the overall mean of the residuals (0 ± 0.93), except for pair 5-7 which was associated with larger residuals (0.76 ± 0.99), $t(38) = 4.79$, $p < .001$ (after Bonferroni correction for multiple comparisons). Responses to pair 5-7 were slower than what was predicted by the mean value of the digits to be compared.

1.2.3 Step-at-five analysis

One could argue that the analysis of the residuals from regressions pairs of single digits may have led to an overestimation of the effect of numerical size, due to its collinearity with the putative sub-base-five effect. This could have masked the effect of sub-base-five in the subsequent analyses. In order to circumvent this collinearity issue, we analysed the differences in ln-transformed RTs between each adjacent pairs within participants (dRTs). In the following analyses, we focused on pairs containing exclusively single-digit numerals. Figure 4 (panel A) represents the dRTs for each couple of pairs, along with the regression line obtained from individual regressions. As a reminder, the individual dRTs were regressed on the difference between the means of the ln-transformed values of the two contrasted pairs, in order to align with the previous analyses accounting for a logarithmic impact of numerical size on the performance. The subsequent residuals for the couple [3-5]~[4-6] (-0.18 ± 0.88) did not differ from zero, $t(38) = 1.24$; $p = .221$. Furthermore, the absolute value of their mean (0.18), did not differ from the distribution of the absolute-mean-values of the residuals associated with the five other couples (0.33 ± 0.31) , $t(4) = 1.14$; $p = .317$. These results mean that the difference between RTs for pairs 3-5 and 4-6 did not deviate more than the others - and actually did not deviate at all - from the continuous evolution predicted on the basis of increasing numerical size (see Figure 4, panel B and C). Besides, a one sample *t*-test conducted on the observed dRTs for the couple [3-5]~[4-6] revealed that this difference did not even differ from 0 (0.01 ± 0.04), $t(38) = 1.41$; $p = .166$. No significant increase in RTs was found between responses to these two pairs.

Exploratory analyses showed that the difference in RTs between pairs 4-6 and 5-7 stood out. These analyses not only showed that the residuals for the couple of pairs [4-6]~[5-7] were larger than predicted by the regression (0.77 ± 1.02), $t(38) = 4.69$; $p < .001$ (after Bonferroni correction for multiple comparisons), but also that the absolute value of their mean (0.77) was the only one that differed from the distribution of the others (0.22 ± 0.20), $t(4) = 6.24$; $p = .02$ (all other *p*-values > .49, after correction). Furthermore, this steeper RT increase between pairs 4-6 and 5-7 was followed by a plateau: dRT for the couple [5-7]~[6-8] did not differ from 0 (-0.03 ± 0.06), $t(38) = 2.67$; $p = .066$ (after Bonferroni correction), despite the increasing numerical size of the pairs; and, the residuals associated with this couple after regression of the size effect were found to be negative (-0.56 ± 0.92), $t(38) = 3.78$; $p = .003$ (after Bonferroni correction). Lastly, the other dRTs did not deviate from the predictions of the regression model, as the associated residuals did not differ from 0 (all *p*-values > .807, after Bonferroni correction).

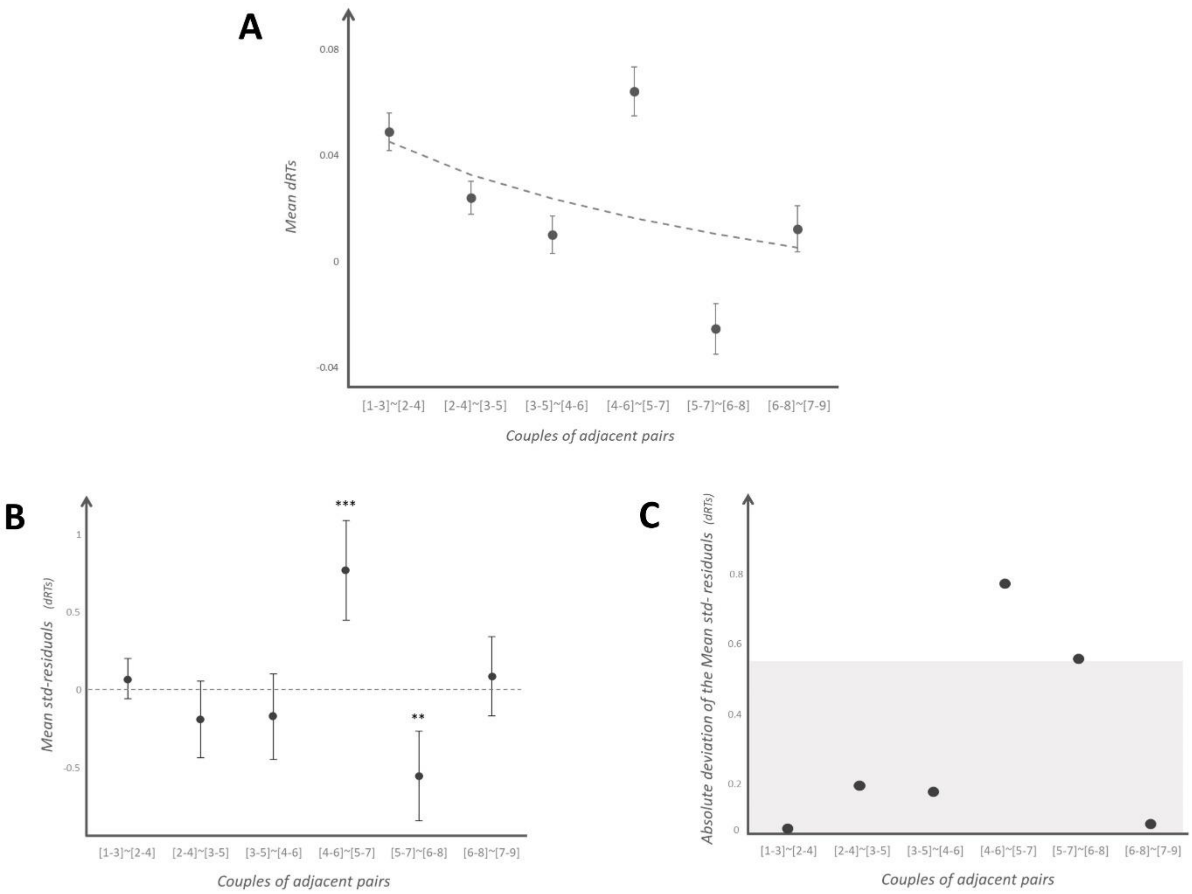

Figure 4. Step-at-five approach. A. Mean differences in ln-transformed response times within couples of adjacent pairs (dRTs); error bars represent standard error of the mean. The dashed grey line represents the model of the effect of numerical size averaged at the group level, after fitting the individual dRTs with a function of the type: dRT = a * (mean ln-values of the pairs) + b. **B.** Mean standardised residuals for the couples of pairs; error bars represent 95% IC. Std-residuals are close to zero, except for couples [4-6]~[5-7] and [5-7]~[6-8], which are respectively associated with residuals larger and smaller than 0. **C.** Absolute deviation (from 0) of the averaged standardised residuals. The grey zone represents the 95% IC for the absolute deviation range. The absolute deviation for mean std-residuals associated with couple [4-6]~[5-7] is outlier; dRTs for this couple of pairs deviate more from the predictions than the dRTs for the other couples. ** $p \leq .01$; *** $p \leq .001$.

1.3 Interim discussion

The sub-base-five hypothesis of Domahs and colleagues (2010) predicted a difference in response times (RTs) between pairs of numbers ≤ 5 (i.e., No-Break pairs: 1-3, 2-4, 3-5) and pairs with a least one number > 5 (i.e., Break-5 pairs: 4-6, 5-7, 6-8, 7-9). To distinguish this sub-base-five effect from the mere effect of numerical size, the authors proposed to regress RTs on the mean value of the pairs and then to test the difference between the residuals associated with Break-5 and No-Break pairs.

Our results strictly replicate those from the original study, as such difference was observed when RTs were regressed on numerical size over an extended range of pairs including two-digit numerals (i.e., from 1-3 to 18-20). However, additional analyses showed that the effect disappeared when the regressions were conducted over the range of single digits. The regression of RTs over the extended range indeed failed to fully capture the effect of numerical size, as attested by the positive correlation between this latter and the residual RTs. The goodness-of-fit improved when RTs were regressed over the single-digit range, but the effect of sub-base-five was then no longer significant. Although the error rate was very low, we conducted exploratory analyses on accuracy that strictly replicated these findings (see Supplemental 1 – Experiment 1).

These results showed that the collinearity with numerical size plays a critical role in the assessment of the sub-base-five effect. To overcome this issue, we decided to approach the problem differently, by looking at RT differences between the adjacent pairs. The sub-base-five hypothesis predicts a steeper increase in RTs when crossing the boundary between No-Break and Break-5 pairs. The difference in RTs between pairs 3-5 and 4-6 should consequently be larger than the others (at least after having considered the logarithmic decrease of dRTs with pair size due to the continuous size effect). The results revealed that the difference in RTs for this particular couple of pairs did not deviate more from the predictions allowed by the numerical size effect than the other differences between couples of adjacent pairs. To sum up, no steeper increase in RTs was found when crossing the boundary of the putative sub-base-five. Even more so, we found no significant difference in RTs between pairs 3-5 and 4-6, which means that there was no RT increase at all between the last pair of numbers representable on one hand and the first pair with numbers requiring two hands. These results contradict the predictions of the sub-base-five hypothesis.

An unexpected finding was that participants were particularly slow to discriminate 5 and 7. The difference in RTs observed between pair 4-6 and pair 5-7 was larger than predicted by the model of the numerical size effect, and this deviation was particularly notable as it was larger than what was observed for the other couples of adjacent pairs. The results indicate a spike in RTs confined to pair 5-7, rather than an actual step in the RT increase, since they further showed that, despite larger numerical size, pair 6-8 was actually judged slightly faster than pair 5-7.

Experiment 2 aimed to replicate these observations using a task that specifically assessed the comparison of pairs of digits within the first decade. We reasoned that, in Experiment 1, the inclusion of two-digit numbers might have focussed participants' attention on the decade-unit structure, making the sub-base-five representation less salient. The task used in Experiment 2 also varied the response associated with digits 1 and 2, by adding pairs 0-1 and 0-2 to the stimulus list, in order to minimize the use of strategies based on statistical learning and encourage semantic processing of all pairs. Hence, in this new experiment, the range of pairs presented to participants closely matched the range of pairs included in the analysis, and the digits constituting these pairs were equally often associated with a small or a large response.

Experiment 2

1.4 Method

1.4.1 Participants

A new pool of participants was recruited on the university campus through posts on social media. The inclusion criteria were identical to Experiment 1. Twenty-eight participants were included, but one was excluded a posteriori because of an outlier global error rate (20%). The final sample therefore consisted of 27 participants (18 females and 9 males; 2 left-handed; mean age ± *SD*: 24 ± 3 years). The minimum sample size was estimated at 7 participants, based on a power analysis using G*Power (version 3.1.9.7), with $dz = 1.48$ (as estimated based on the results of Experiment 1), $\alpha = .05$, and power = .85.

1.4.2 Apparatus and Stimuli

The same apparatus as in Experiment 1 was used. Stimuli still consisted of pairs of Arabic numerals, but in this second experiment, numerals from 0 to 10 were used. More specifically stimuli set consisted of the single-digit numeral pairs of Experiment 1 (i.e., from 1-3 to 7-9), to which were added pairs 0-1, 0-2, 8-10 and 9-10, so that, in this second experiment, digits from 1 to 9 all appeared as many times as the largest or as the smallest of the pair. In total there were 11 pairs, presented in ascending (0-1 to 9-10) or descending order (10-9 to 1-0), resulting in 22 stimuli.

1.4.3 Task and Procedure

Task and procedure were similar to those in Experiment 1. In total, participants went through 880 trials: i.e., 2 runs (instruction) composed of 10 blocks of 22 stimuli repeated 2 times. Each run was preceded by eighteen practice trials, with written feedback about accuracy.

1.4.4 Data Analysis

Analyses focused exclusively on pairs of single-digit numerals separated by a numerical distance of two units, excluding pairs with ‘0’ and ‘10’ (for sake of consistency with Experiment 1): i.e., pairs 1-3, 2-4, 3-5, 4-6, 5-7, 6-8, and 7-9. The four other pairs (i.e., 0-1, 0-2, 8-10, 9-10) were considered as fillers. The RTs analyses were performed on correct trials (mean participants’ accuracy ± *SD*: 93% ± 4%), after filtering out extreme values using the MAD, computed individually for all blocks sharing the same instruction (small or large; Leys et al., 2013). Analyses were performed on the remaining trials after aggregating all the blocks (90% of the dataset). Both Residual approach and Step-at-five approach analyses were conducted as in Experiment 1, except that the Residual analysis was conducted after regressions over the single-digit range only.

Although participants made very few errors, we also conducted exploratory analyses on accuracy. The results replicated those observed in the response time analysis and are presented as supplemental material (see Supplemental 1 – Experiment 2).

1.5 Results

1.5.1 Residual analysis

Figure 5 (panel A) represents the mean RTs as a function of the numeral pairs, along with the regression line obtained after regressing RTs on the mean value of the pairs. The residuals from the regression analysis did not differ between the No-Break (-0.04 ± 0.18) and the Break-5 pairs (0.03 ± 0.14), $t(26) = 1.10$, $p = .283$, and the latter did not differ from zero either, $t(26) = 1.10$; $p = .283$. These results are illustrated in Figure 5 (panel B). No remaining correlation was found between the standardised residuals and the mean value of the pairs of single digits, $r = -.004$, $p = .994$. The distribution of the residuals supports a fitting of the numerical size effect that is decently adjusted to the data (see Figure 5, panel C).

An exploratory analysis using repeated measures ANOVA on the residuals, with pairs as a within-subjects variable, revealed a significant effect of the pair, $F(6,156) = 16.49$, $p < .001$. Post-hoc comparisons indicated that three pairs differed from the overall mean of the residuals (0 ± 0.93): first, pair 5-7, which was associated with larger residuals (1.18 ± 0.66), $t(38) = 9.32$, $p < .001$, as in Experiment 1; but also, pairs 4-6 and 7-9, which were associated with smaller residuals (respectively -0.76 ± 0.76, $t(26) = 5.16$, $p < .001$, and -0.50 ± 0.81, $t(26) = 3.23$, $p = .023$, after Bonferroni correction).

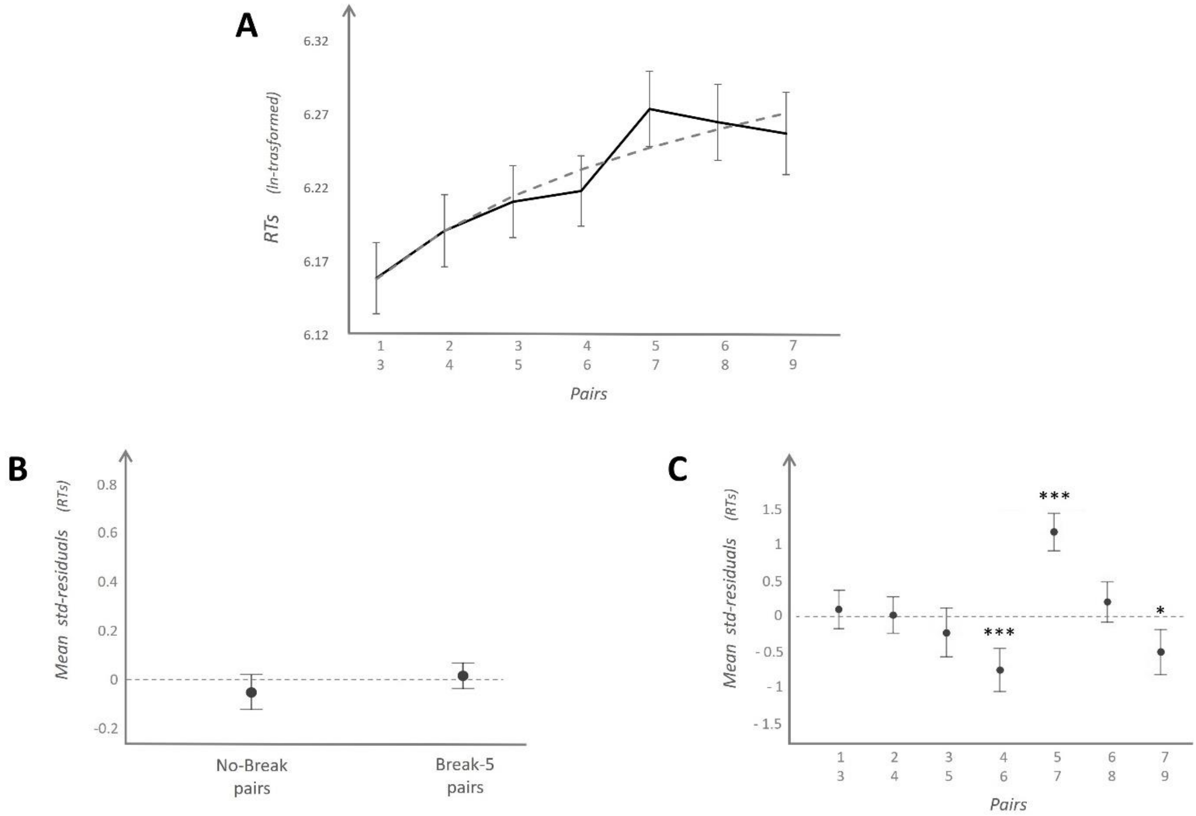

Figure 5. Residual approach after individual regressions over the single-digit range of pairs. A. Mean response times (RTs; ln-transformed) as a function of the digit pairs. Observed values are connected with black plain line; error bars represent the standard error of the mean. The dashed grey line represents the model of the effect of numerical size averaged at the group level, after fitting the individual data from pairs of single digits with a function of the type: ln(mean-RT) = a * ln(pair-mean-value) + b, excluding pairs with '0'. **B.** Mean standardised residuals for the Break-5 and No-Break pairs; error bars represent the 95% IC. The residuals for the Break-5 pairs do not differ from zero, or from those for the No-Break pairs. **C.** Mean standardised residuals for the pairs of single-digit numeral; error bars represent the standard error of the mean. The residuals are close to zero, except for pairs 4-6 and 7-9, which are associated with smaller residuals than the other pairs, and pair 5-7, which is associated with larger residuals than the other pairs. *** $p \leq .001$.

1.5.2 Step-at-five analysis

The following analyses aimed at examining the differences in RTs between each couple of adjacent pairs from 1-3 to 7-9 within participants (dRTs). Figure 6 (panel A) represents the distribution of these dRTs, along with the regression line obtained after regressing them on the mean value of the couples (after ln-transformation). Following the regressions, the residuals for the couple [3-5]~[4-6] were not found to be larger than zero, but instead they were significantly negative (-0.33 ± 0.73), $t(26) = 2.36$; $p = .026$ (see Figure 6, panel B). The absolute value of their mean (0.33) did not differ from the distribution of the absolute-mean-values of the residuals associated with the five other couples (0.48 ± 0.53), $t(4) = 0.65$; $p = .552$ (see Figure 6, panel C). As in experiment 1, these results show that the difference in RTs between response for pairs 3-5 and 4-6 did not deviate from the continuous evolution predicted on the basis of increasing numerical size. Besides, a one sample *t*-test conducted on the observed dRTs for the couple [3-5]~[4-6] revealed again that this difference did not even differ from 0 (0.01 ± 0.02), $t(26) = 1.55$; $p = .134$ (see Figure 6, panel A), indicating no significant increase in RTs between the responses to these two pairs.

As in Experiment 1, exploratory analyses showed that the difference in RTs between pairs 4-6 and 5-7 stood out. These analyses not only showed that the residuals for the couple of pairs [4-6]~[5-7] were larger than predicted by the regression (1.38 ± 0.49), $t(26) = 14.6$; $p < .001$ (after Bonferroni correction for multiple comparisons), but also that the absolute value of their mean (1.38) was the only one that differed from the distribution of the others (0.28 ± 0.17), $t(4) = 14.39$; $p < .001$ (all other *p*-values > .61, after Bonferroni correction). Furthermore, this steeper RT increase between pairs 4-6 and 5-7 was followed by a plateau: dRT for the couple [5-7]~[6-8] did not differ from 0 (-0.01 ± 0.03), $t(26) = 1.99$; $p = .345$ (after Bonferroni correction), despite the increasing numerical size of the pairs; and, the residuals associated with this couple after regression of the size effect were found to be negative (-0.50 ± 0.81), $t(26) = 3.19$; $p = .022$ (after Bonferroni correction). None of the other dRTs was found to significantly deviate from the predictions, the residuals they are associated with did not differ from 0 (all remaining *p*-values > .080, after Bonferroni correction).

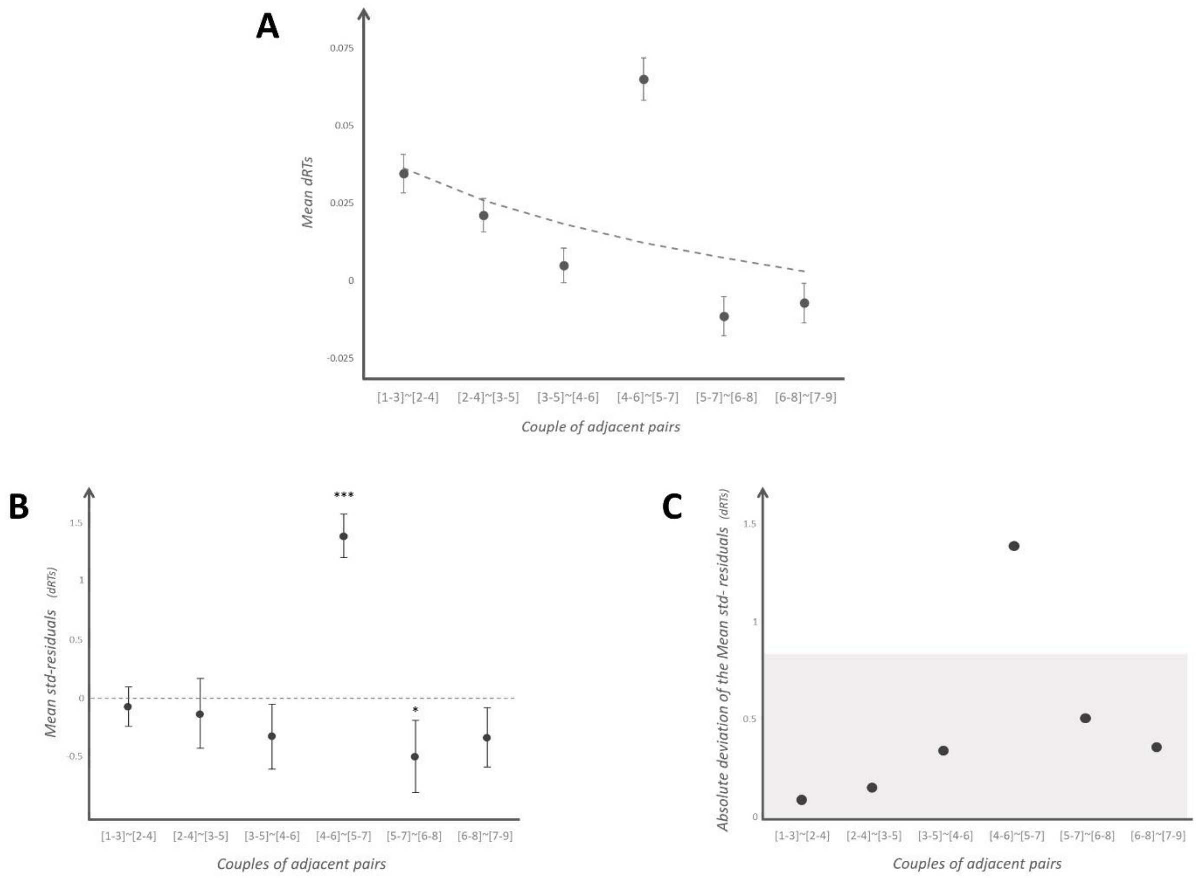

Figure 6. Step-at-five approach. A. Mean differences in ln-transformed response times within couples of adjacent pairs (dRTs); error bars represent standard error of the mean. The dashed grey line represents the model of the effect of numerical size averaged at the group level, after fitting the individual dRTs with a function of the type: dRT = a * (mean ln-values of the pairs) + b. **B.** Mean standardised residuals for the couples of pairs; error bars represent 95% IC. Std-residuals are close to zero, except for couples [4-6]~[5-7] and [5-7]~[6-8], which are respectively associated with residuals larger and smaller than 0. **C.** Absolute deviation (from 0) of the mean standardised residuals. The grey zone represents the 95% IC for the absolute deviation range. The absolute deviation for mean std-residuals associated with couple [4-6]~[5-7] is outlier; dRTs for this couple of pairs deviate more from the predictions than dRTs for the other couples. * $p \leq .05$; *** $p \leq .001$.

1.6 Interim discussion

While the task emphasized the processing of digit magnitude in pairs ranging from 1-3 to 7-9, the results of Experiment 2 also did not reveal any evidence that a sub-base-five effect would influence the response times (RTs). The results were strictly similar to those of Experiment 1. First, after regressing RTs on numerical size, no difference was observed between the residuals obtained for No-Break and Break-5 pairs (exploratory analyses on accuracy yielded the same results, see Supplemental

1 – Experiment 2). Second, the contrast between pairs 4-6 and 3-5 (which are located on either side of the boundary segregating No-Break and Break-5 pairs), revealed no significant increase in RTs. Third, we replicated the finding of a selective difficulty to discriminate 5 and 7. As in Experiment 1, the pair 5-7 gave rise to particularly slow responses. This slowdown did not reflect a sub-base-five effect, however, because it was not observed with other pairs containing number(s) larger than 5.

General Discussion

The present research aimed to investigate the representations underlying the processing of symbolic number magnitude. In particular, we tested the hypothesis that the performance in symbolic number comparison is influenced by a sub-base-five effect, using a new approach designed to better control for its collinearity with the numerical size effect (see Moyer & Landauer, 1967; or Parkman, 1971). Our results challenge the existence of a sub-base-five effect in symbolic number comparison. The results of Experiment 1 showed that this effect was likely an artefact resulting from the modelling method used in the original study by Domahs and colleagues (2010), and, more specifically, from the underestimation of the numerical size effect when response times were regressed on numerical size over an extended range of pairs including two-digit numerals. This regression method rests on the postulate that there is a continuity of the size effect over single and two-digit numbers, which is debatable in light of empirical evidence showing that the processing of two-digit numbers is not holistic (e.g., Tzelgov et al., 2015; Ganor-Stern et al., 2009; Nuerk et al., 2001; Verguts & De Moor, 2005; Zhang & Wang, 2005). We showed that a regression analysis including only pairs of single digits better captures the numerical size effect in this range and expunges the residuals from the difference between No-Break and Break-5 pairs, meaning that this difference was driven by the inclusion of the two-digit pair data. Our argument that single-digit comparison does not activate representations organised according to a sub-base-five structure is further supported by the absence of RT increase when contrasting directly pairs 3-5 and 4-6, which marks the transition between No-Break and Break-5 pairs. These results were replicated in Experiment 2, after restricting the range of comparisons to single digits and adding fillers to ensure that all digits were associated equally often with the "small" or "large" responses.

How, then, can we explain the fact that a sub-base-five effect was found in German but not Chinese participants in Domahs and colleagues' original study (2010)? We argue that this discrepancy arised because the statistical artefact caused by the regression method used by the authors was less pronounced in the Chinese group than in the German group. The regression slope for the Chinese group (slope = 0.094) was actually steeper by 11% than the slope of the German group (slope = 0.085). As a result, it is plausible that the underestimation of the size effect within the single-digit range was more pronounced in the German than in the Chinese group. We can only speculate on the origin of the between-group differences, but this larger slope in the Chinese group was possibly due to the leverage effect of their RTs on pairs with two digit-numbers, especially pair 10-12. This leverage effect might reflect verbal interference from Chinese number words which are typically longer above 10 than below, as reported by Domahs and colleagues (2010). In any case, our data points to the statistical method rather than the finger counting pattern being responsible for the observed discrepancy between the two groups.

Among the evidence supporting a link between finger and number representations, the sub-base-five-effect had a particular status as it was considered as key evidence that finger counting experience could shape the processing of symbolic number magnitude. The present results have shown that response times in symbolic number comparison do not reflect a sub-base-five effect, and therefore calls for new evidence of the involvement of finger-based representations in the processing of number magnitude. Two studies have explored the influence of the sub-base-five effect in the arithmetic domain, but their results need to be considered with caution. A first study showed that second-grade children sometimes make split-five errors, i.e., errors distant of five units from the correct answer, as they start solving addition and subtraction problems mentally. This pattern suggests that they rely on a mental image of their fingers to represent intermediary results, occasionally omitting one hand as they update their mental image (Domahs et al., 2008). However, the results of this study do not provide tangible evidence that bodily experience has a long-lasting impact on numerical cognition, because split-five errors were rare

(57% of the children never committed such errors), inconsistent (they do not repeat across testing sessions), and transient (split-five errors significantly decreased as children reached the third grade). This pattern led the authors to conclude that the sub-base-five structure imposed by finger counting progressively loses its influence on arithmetic performance during development. A related study nevertheless indicated that adults take more time to perform mental additions when the calculation implies to cross the middle of a decade (11+2 vs. 12+4), suggesting an invisible boundary around 5 analogue to the gap between the two hands (Klein et al., 2011). However, a quick look at the stimuli used in this study (see Appendix, p.7) reveals that the second operands were not systematically the same in no-break and break-5 problems, and their differences in size suffice to explain the extra time cost in the break-5 condition. This is further evidenced by a substantial correlation between the numerical size of second operands (+2, +3 or +4) and the response time ($r = 0.44$, $p < .001$). Further studies are thus necessary to elucidate whether and how the sub-base-five structure that characterises some finger counting practices influences number processing and calculation.

While our results contradict the existence of a sub-base-five effect in single-digit comparison, they do reveal a previously unnoticed effect, characterised by a particular slowdown affecting specifically the comparison between 5 and 7. Indeed, our results showed that numerical size decently predicts response times in single-digit comparison, except for the pair 5-7 that gave rise to a marked slowdown in both experiments. It should be noted that this effect is distinct from the “odd effect” reported by Hines (1990), which refers to the finding that it is easier to decide whether an Arabic digit is odd or even when this digit is even than when it is odd. The slower judgments on odd digits were related to the fact that "odd" is a linguistically marked and "even" a linguistically unmarked concept (Hines, 1990, 2010). If parity or markedness had played a role in the present results, then we should have observed a RT slowdown for all pairs of odd numbers compared to even numbers. The present data rather indicate that the slowdown was specific to the pair 5-7. This pair is special because it is made up of two prime numbers, which in addition do not belong to any series of multiples in the range of single-digit numbers (such as 3-6-9 or 2-4-6-8). This lack of information about their relations with other digits would make 5 and 7 more difficult to discriminate, especially when they are associated within the same pair. Supportive evidence comes from a study of Shepard and colleagues (1975) showing that 5 and 7 are isolated within the multidimensional space representing the perceived similarity between single-digit numbers, presumably because they are prime numbers that do not share any multiplicative relationship with the other single-digit numbers. In sum, performance in symbolic number comparison would also be influenced by semantic properties that are not directly related to numerical quantity, such as the density of multiplicative relationships. This calls for a refinement of models of symbolic number comparison which posit that performance exclusively relies on the representation of numerical quantity (e.g., Dehaene & Cohen, 1995; Dehaene, 2007).

In conclusion, our study challenges previous results suggesting that finger-based representations influence symbolic number comparison. In addition, our study revealed a novel effect: a relative difficulty in comparing 5 and 7, beyond the effect of numerical size. This finding questions the idea that the content of the representations underlying the comparison of numerals is reducible to numerical quantity alone. This calls for further investigation into which and how other numerical properties may influence the processing of symbolic number magnitude.

Statement

Declaration of interest

None.

Data availability statement

During the review process, data supporting the results and the R scripts used for data analyses are accessible to the reviewers through the following anonymised OSF link. They will be made publicly available, under the CC BY license, at the time of publication.

https://osf.io/2q9bc/?view_only=38aebacf8d044a29a38067cd51b371f7

References

Berteletti, I., & Booth, J. R. (2016). Finger Representation and Finger-Based Strategies in the Acquisition of Number Meaning and Arithmetic. In *Development of Mathematical Cognition* (pp. 109-139). https://doi.org/10.1016/b978-0-12-801871-2.00005-8

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of experimental psychology: General*, *124*(4), 434. https://doi.org/10.1037/0096-3445.124.4.434

Dehaene, S. (1997) *The Number Sense*. Oxford University Press, New York.

Dehaene, S. (2001). Précis of The Number Sense. *Mind & Language*, *16*(1), 16-36. https://doi.org/https://doi.org/10.1111/1468-0017.00154

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. *Sensorimotor foundations of higher cognition*, *22*, 527-574.

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, *5*(4), 390-407.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical cognition*, *1*(1), 83-120.

Domahs, F., Krinzinger, H., & Willmes, K. (2008). Mind the gap between both hands: Evidence for internal finger-based number representations in children's mental calculation. *Cortex*, *44*(4), 359-367. https://doi.org/10.1016/j.cortex.2007.08.001

Domahs, F., Moeller, K., Huber, S., Willmes, K., & Nuerk, H.-C. (2010). Embodied numerosity: Implicit hand-based representations influence symbolic number processing across cultures. *Cognition*, *116*(2), 251-266. https://doi.org/https://doi.org/10.1016/j.cognition.2010.05.007

Fischer, M. H., & Brugger, P. (2011). When digits help digits: spatial-numerical associations point to finger counting as prime example of embodied cognition. *Frontiers in Psychology*, *2*, 260. https://doi.org/10.3389/fpsyg.2011.00260

Ganor-Stern, D., Pinhas, M., & Tzelgov, J. (2009). Short Article: Comparing Two-Digit Numbers: The Importance of being Presented Together. *Quarterly Journal of Experimental Psychology*, *62*(3), 444-452. https://doi.org/10.1080/17470210802391631

Hines, T. M. (1990). An odd effect: Lengthened reaction times for judgments about odd digits. *Memory & Cognition*, *18*, 40-46. https://doi.org/10.3758/BF03202644

Hines, T. M. (2010). The odd effect and multiple meanings in English of the words "odd" and "even". *Psychological reports*, *107*(2), 443-446. https://doi.org/10.2466/04.10.11.22.PR0.107.5.443-446

Klein, E., Moeller, K., Willmes, K., Nuerk, H. C., & Domahs, F. (2011). The influence of implicit hand-based representations on mental arithmetic. *Frontiers in psychology*, *2*, 197. https://doi.org/10.3389/fpsyg.2011.00197

Krajcsi, A. (2017). Numerical distance and size effects dissociate in Indo-Arabic number comparison. *Psychonomic Bulletin & Review*, *24*(3), 927-934. https://doi.org/10.3758/s13423-016-1175-6

Krajcsi, A., Lengyel, G., & Kojouharova, P. (2016). The source of the symbolic numerical distance and size effects. *Frontiers in psychology*, *7*, 1795. https://doi.org/10.3389/fpsyg.2016.01795

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764-766. https://doi.org/10.1016/j.jesp.2013.03.013

Lindemann, O., Alipour, A., & Fischer, M. H. (2011). Finger counting habits in middle eastern and western individuals: an online survey. *Journal of Cross-Cultural Psychology*, *42*(4), 566-578. https://doi.org/10.1177/0022022111406254

Marinova, M., Sasanguie, D., & Reynvoet, B. (2021). Numerals do not need numerosities: robust evidence for distinct numerical representations for symbolic and non-symbolic numbers. *Psychological Research*, *85*(2), 764-776. https://doi.org/10.1007/s00426-019-01286-z

Moyer, R. S., & Landauer, T. K. (1967). Time required for Judgements of Numerical Inequality. *Nature*, *215*(5109), 1519-1520. https://doi.org/10.1038/2151519a0

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366-382.

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual review of neuroscience*, *32*(1), 185-208.

Nuerk, H.-C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, *82*(1), B25-B33. https://doi.org/https://doi.org/10.1016/S0010-0277(01)00142-1

Núñez, R. E. (2017). Is There Really an Evolved Capacity for Number? *Trends in Cognitive Sciences*, *21*(6), 409-424.

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, *91*(2), 191-205. https://doi.org/10.1037/h0031854

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.* 10.3758/s13428-018-01193-y

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex. *Neuron*, *53*(2), 293-305. https://doi.org/https://doi.org/10.1016/j.neuron.2006.11.022

Reynvoet, B., & Sasanguie, D. (2016). The Symbol Grounding Problem Revisited: A Thorough Evaluation of the ANS Mapping Account and the Proposal of an Alternative Account Based on Symbol–Symbol Associations. *Frontiers in Psychology*, *7*.

Rips, L. J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and brain sciences*, *31*(6), 623-642.

Roesch, S., & Moeller, K. (2015). Considering digits in a current model of numerical development [Perspective]. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.01062

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive psychology*, *7*(1), 82-138. https://doi.org/https://doi.org/10.1016/0010-0285(75)90006-7

Sixtus, E., Krause, F., Lindemann, O., & Fischer, M. H. (2023). A sensorimotor perspective on numerical cognition. *Trends in Cognitive Sciences*, *27*(4), 367-378. https://doi.org/10.1016/j.tics.2023.01.002

Soylu, F., Lester Jr, F. K., & Newman, S. D. (2018). You Can Count on Your Fingers: The Role of Fingers in Early Mathematical Development. *Journal of Numerical Cognition*, *4*(1), 107-135. https://doi.org/10.5964/jnc.v4i1.85

Tzelgov, J., Ganor-Stern, D., Kallai, A., & Pinhas, M. (2015). Primitives and non-primitives of numerical representations. *The Oxford handbook of numerical cognition*, 45-66.

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

Verguts, T., & Van Opstal, F. (2014). A delta-rule model of numerical and non-numerical order processing. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1092-1102. https://doi.org/10.1037/a0035114

Zhang, J., & Wang, H. (2005). The effect of external representations on numeric tasks. *The Quarterly Journal of Experimental Psychology Section A*, *58*(5), 817-838. https://doi.org/10.1080/02724980443000340