Research paper Acta Neurobiol Exp 2013, 73: 289–303

The number decision task: Investigation of the representation of multi-digit numbers

Katharina Sass[1,2,3]*, Lena Theissen[1,5], Thomas F. Münte[4], Ute Habel[1,5], and Arie van der Lugt[3]

[1]Department of Psychiatry, Psychotherapy and Psychosomatics, Medical School, RWTH Aachen University, Aachen, German, *Email: k.sass@uq.edu.au y; [2]School of Psychology, The University of Queensland, Brisbane, Australia; [3]Faculty of Psychology, Maastricht University, Maastricht, Netherlands; [4]Department of Neurology, University of Lübeck, Lübeck, Germany; [5]JARA – Translational Brain Medicine, Aachen, Germany

The ability to process multi-digit numbers is an essential skill which we investigated using a number decision task. Subjects were asked to decide whether a target number (e.g., 649) is too small or too large to be the mean between two delimiter numbers that constituted the interval (e.g., 567 and 715). Three-digit numbers were presented vertically with (1) growing interval sizes (i.e., distance between the two delimiters; e.g., interval size between 567 and 715 is 148) and target gap to the mean (e.g., the gap between the 'real' mean 641 and the target 649 is 8) and (2) growing interval sizes with constant gap to the mean (i.e., for each interval size the gap between target and mean was held constant). The results showed that target gap to the mean "masked" the influence of interval sizes, i.e., subjects' performance improved with increasing interval sizes (distance effect). This effect was reversed when constant target gaps to the mean and growing interval sizes were presented. These results were replicated presenting the numbers horizontally and with two-digit numbers. Additionally, a significant influence of decade but no effect of unit compatibility on reaction times and error rates, on number magnitude (size effect) and response format was found. Overall, we showed that the number decision task is an efficient tool to investigate multi-digit number representation and the results from the experiments revealed evidence for a hybrid model of multi-digit number representation in which numbers are represented as a whole but also on separate mental number lines that interact with each other.

Key words: number representation, compatibility, number decision task, hybrid model, multi-digit numbers, distance effect, size effect

GENERAL INTRODUCTION

Different theories try to account for the representation and storage of numbers with the common assumption that numbers are transformed from their symbolic format to an analogical quantity representation on the so-called mental number line (MNL;). On the MNL, numbers are represented in a continuous and quantity-based analogical format and are organized by their numerical proximity. Thus, small numbers are located on the left-hand side and large numbers are located on the right, at least in cultures with a reading direction from left to right[1] (e.g., Moeller et al. 2009a, 2011). Although numbers are represented by their numerical proximity, there is increasing fuzziness in the mental representation for larger numbers resulting in less precise processing of larger numbers (Dehaene and Mehler 1992, Dehaene 2001). However, theories differ with regard to the number of MNLs. The holistic model assumes that numbers are represented as a "whole" on one single MNL (e.g., Dehaene et al. 1990, Brysbaert 1995). According to decompositional models (Nuerk et al. 2001, Verguts and Fias 2004, Ratinckx et al. 2005, Verguts and

Correspondence should be addressed to K. Sass,
Email: k.sass@uq.edu.au

Received 17 January 2013, accepted 08 April 2013

[1] Recent studies showed that the MNL is horizontal by default by could also be represented vertically when "magnitude is a salient component of the representation"(Zuber et al. 2009). For example, it has been shown that for responses defined in the vertical dimension, responding to relatively large numbers is faster with a top response and responding to relatively small numbers is faster with a bottom response (Shepard et al. 1975, Moeller et al. 2009b, c).

© 2013 by Polish Neuroscience Society - PTBUN, Nencki Institute of Experimental Biology

De Moor 2005) decade and unit digits of multi-digit numbers are represented separately on different mental number lines that interact with each other. The hybrid model combines both approaches (Nuerk and Willmes 2005) assuming that during number comparisons separate but connected comparisons take place for the overall magnitude of the numbers and for the decade and unit magnitudes, respectively. Overall, within the last years evidence for the hybrid model has become overwhelming (e.g., Korvorst and Damian 2008, Moeller et al. 2009a, 2011). However, detailed refinement of the model is still rare, for example, how do the overall magnitude and the individual magnitudes interact.

A well-established technique to test and refine these models and to investigate the spatial representation of number magnitude is the number bisection task (NBT; Nuerk et al. 2002). The subject is asked to "bisect" an interval of two numbers into two equal parts to get the numerical mean. For example, the numerical mean of the interval 7 and 13 is 10. Overall, different formats of the NBT have been suggested: In the production task (e.g., Zorzi et al. 2002) subjects have to name or write down the numerical mean. In the multiple-choice-task (e.g., van Herten 1999), they have to choose the correct answer from several alternatives. In the verification paradigm (e.g., Geppert 2005) triplets of three different numbers are presented (e.g., 2_5_8) with the outer numbers (delimiters) defining the numerical interval and the middle number being the target. The subject has to decide if the target is the numerical mean or not by pressing one of two buttons. Some advantages and disadvantages have been noted for the different paradigms (see also Nuerk et al. 2002). For example, the production task is sensitive to the length of syllables, whereas the multiple-choice task is influenced by the distractors, i.e., numbers that do not "bisect" the numerical interval and might increase variance of errors due to quantity and position. By contrast, the verification paradigm seems advantageous as there are only two response alternatives and reaction times are easy to calculate as subjects respond via button press. Chance performance is 50%, however, and the particular presentation of numbers may also influence performance. In addition, beside task specific influences there are also other factors that have an impact on the performance in the NBT: (1) Multiplicativity/ Multiplication knowledge: Performance for number triplets from multiplication tables (e.g., 3_6_9) is better than for triplets in non-multiplicative trials (e.g., 5_9_13; van Herten 1999, Nuerk et al. 2002). (2) Parity: Bisection of two numbers into their numerical middle is only possible if the two numbers have the same parity (e.g., 4 and 8 are both even numbers). Hence, "checking" the parity of the two outer numbers (delimiters) can be used as a strategy that simplifies the task (Nuerk et al. 2002). (3) Delimiter distance: Distance between the two outer numbers and gap between the target and the numerical mean influences performance (van Herten 1999). Reaction time increases with larger intervals (i.e., distance between the delimiters), possibly because the two delimiter numbers activate a larger segment of the MNL (Nuerk et al. 2002). In contrast, if the gap between the target and the numerical mean is large, performance improves (Geppert 2005), i.e., it is easier to reject 6_7_14 than 6_9_14. To conclude, despite its advantages, the NBT is thus faced with a number of problems calling for an improvement like influence of parity, multiplicativity or influence of reading direction that might mask the main effects of number representation and processing.

Hence, the first aim of the current experiments is the validation of a new paradigm called "Number Decision Task" (NDT) as efficient and alternative tool to investigate the representation of multi-digit numbers. In this modified version of the NBT, we presented three-digit numbers as a triplet. The middle number was the target and the other two delimiters defined the numerical interval. The participants were instructed to decide whether the target was too small or too large to be the numerical mean. In Experiment 1 (pilot experiment) and Experiment 2 we used three-digit numbers that were presented vertically to avoid any influence of direction bias. In Experiment 3, we used the same task but with two-digit numbers, in order to adapt the NDT for the investigation of numerical representations in patients with different disorders (e.g., schizophrenia, depression; Liotti and Mayberg 2001, Cavezian et al. 2007b). The advantages of the NDT are that using large, multi-digit numbers the influence of parity and multiplicativity is minimized (Nuerk et al. 2002). Also, if subjects are instructed to decide whether the target is too small or too large to be the numerical mean, multiplication fact knowledge cannot be used and the influence of parity is eliminated, as parity does not help to answer the smaller/ larger question. For the validation, we focused on the classical effects of numerical representation, namely the size and the distance effect. The size effect describes decreased performance if the distance between two numbers is constant but the numerical magnitude increases (e.g., it is easier to distinguish between 5 and 9 than between 54 and 58 even though the numerical distance is identical; Brysbaert 1995, Dehaene 2001). The distance effect refers to the fact that performance in numerosity discrimination decreases if the distance between two numbers is minimized (e.g., the performance is enhanced if subjects have to compare 1 and 5 in comparison to 4 and 5; Moyer and Landauer 1967), i.e., presentation and perception of numbers follow Weber's law (Shepard et al. 1975) . If the NDT is an efficient tool, we should find the same effects as known from the NBT.

The second aim is the comparison of different models of number representation (decomposition vs. hybrid) to refine current theories and accordingly, get insight into the representation of multi-digit numbers on the MNL. Recent research suggested that for larger numbers a decomposition is mandatory (e.g., Nuerk et al. 2001, Ratinckx et al. 2005). Korvorst and Damian (2008) assumed that "the way in which we categorize larger quantity information, by means of place-value, necessarily entails decomposition, thereby minimizing redundancy. […] if we assume that decomposition occurs with larger numbers, then we would expect Stroop-like interference to be introduced between the constituent digits when numbers have to be compared". To investigate this hypothesis that refers to the organization of the MNL, the influence of unit-decade compatibility must be tested (Nuerk et al. 2001). A two-digit number is called compatible when both the unit and the decade digit of one number are larger than the unit and the decade digit of another number (for example 65 and 32). A pair of numbers is called incompatible when the decade number but not the unit number is larger than the decade number of another number, for example 65 and 37. Subjects are more likely to respond correctly to compatible number pairs than to incompatible number pairs because of a "stroop-like" interference (Nuerk et al. 2001). In other words, the compatibility effect is an "inhibition effect" as the overall magnitude might lead to the correct response but is inhibited by incompatible decade/ unit magnitudes. According to the current study and the models of number representation, we hypothesized the following effects for three-digit numbers: If the decompositional model is right then the performance in compatible trials for decades and units will be better than in incompatible trials. Finally, if the hybrid model is the one to be retained then an influence of the whole number as well as an influence of compatibility for decades and for units is assumed, i.e., an interaction of overall and individual magnitudes occur.

EXPERIMENTS 1 AND 2

Introduction

The focus of interest in the first experiments was the examination of multi-digit numbers using the newly developed Number Decision Task. We hypothesized that the larger the numerical interval, the more errors should be made, and the more time would be needed to respond as a larger area of the MNL has to be activated (van Herten 1999). Therefore, we manipulated (a) the size of the interval (distance between the two delimiters) and (b) the size of the target distance to the numerical mean (gap), i.e., in Experiment 1, the target distances to the mean was increased to keep the demands for every interval comparable and in Experiment 2, the target distance to the mean was kept constant (see Table I for the exact manipulation). According to the classical size effect we expected that the larger the number magnitude the worse the performance. Less clear is the prediction concerning the distance effect: Following the suggestion of Geppert (2005) the larger the gap between target and numerical mean, the easier the task while van Herten (1999) assumed that with increasing delimiter distance the performance decreases. Hence, our working hypothesis for Experiment 1 is that an interaction effect of gap between target and numerical mean and delimiter distance occurs. Furthermore, due to complexity of the main analysis the responses of the subjects were pooled but to examine the spatial representation of numbers, we additionally compared both types of responses. We suggested that if the target number is smaller than the real numerical mean between the two delimiters, the target is easier to reject than targets that are larger than the mean because subjects underestimate the real numerical mean and show a leftward bias ("pseudoneglect"; e.g., Nuerk et al. 2002, Hoeckner et al. 2008, Lourenco and Longo 2009).

Table I

Conditions used in the experiments

Interval	Delimiter Distance	Target distance to the mean	
THREE DIGIT NUMBERS		Experiment 1 and A*	Experiment 2 and B*
1st	110–150	± 8–12	constant target distance to the mean (gap): ±11–15 and ±32–36
2nd	180–220	± 16–24	
3rd	250–290	± 24–36	
4th	320–360	± 32–48	
TWO DIGIT NUMBERS		Experiment C*	Experiment 3
1st	11–18	± 1–2	constant target distance to the mean (gap): ±1–2
2nd	22–29	± 2–4	
3rd	32–39	± 3–6	

The exact distances between the interval numbers and the distance to the mean were chosen randomly (e.g., a random number between 8 and 12); * Experiment A, B and C can be found in the supplementary materials (see Appendix).

Method

Participants

Experiment 1: Twenty-eight students were recruited from the Otto-von-Guericke University, Magdeburg, Germany (19 women; age = 23.4 years, S.D. = 3.0 years).

Experiment 2: Sixteen new participants were recruited (6 women; age = 24.5 years, S.D. age = 1.7 years) from the Otto-von-Guericke University, Magdeburg, Germany.

All subjects were right-handed native speakers of German, had normal or corrected-to-normal vision and gave informed consent before participating. The study was approved by the local ethics committee.

Stimuli and Design

For the NDT, two three-digit numbers were presented to the subjects (e.g., 567 and 715), followed by a third number (target; e.g., 641) placed between the delimiters. The lower number of the delimiters (e.g., 567) was always presented at the top of the screen, the higher number (e.g., 715) at the bottom. Subjects were instructed to decide whether the target was too small or too large to be the mean of the two delimiters (e.g., 631 is too small) by pressing one of two buttons.

Four experimental conditions were presented based on the delimiter distance: first interval = 110–150, second interval = 180–220, third interval = 250–290, fourth interval = 320–360 (see Table I for the exact distances between the delimiters and the corresponding distances to the mean). Each condition comprised 60 stimuli with 30 targets being too small and 30 too large. For each subject an individual set of stimuli was generated by using a MATLAB® script.

For the number triplets, both delimiters were randomly generated. For the interval, four different distances were defined (first to fourth interval; see Table I). Dehaene (2001) hypothesized that the distance between the numbers has to be large if the same discrimination performance is to be accomplished as seen for small numbers. Therefore, for the interval we defined a minimum delimiter distance of 111 and a maximum delimiter distance of 360. The exact distances were randomized within the intervals (e.g., the delimiters were 567 and 715 with an interval size of 148, i.e., condition: first interval) and the numerical mean between both delimiters was assessed (e.g., 641). *Via* addition and subtraction, the gap (target distance to the mean) was calculated to obtain the response options “too small” and “too large”. For example, in Experiment 1 a randomized number between for example 8 and 12 for the first interval was subtracted or added to the numerical mean (e.g., 641 + 8 = 649). For Experiment 2 a constant number was subtracted or added, i.e., two different constant gaps between target and real numerical mean (±11–15 and ±32–36) were employed[2] (every interval was combined with both distances).

For the compatibility it was ensured that 50% of the delimiter numbers were compatible and 50% were incompatible. In addition, because of the "Unit-Zero"-effect[3] (Dehaene et al. 1990), the zero was replaced by a "1" within the intervals (e.g., the interval 130 was replaced with 131). For the delimiters, a random number between "11" and "19" was added. The ending "90" was an exception. Here, a number between 21 and 29 was added.

To assess the presentation time of the number triplets, two pretests were conducted with five subjects who did not participate in the main experiment. In the first pretest, the numbers were presented for 1 000 ms (both delimiters for 1 000 ms, followed by the target for 1 000 ms). In the second pretest, the numbers were presented for 1 500 ms. Error rates were significantly different from chance for the 1 500 ms version (P<0.003) but not for the 1 000 ms version (P=0.584). Thus, we used a 1 500-ms presentation.

Procedure

The stimuli were presented in a randomized order in a vertical perceptual arrangement in black Arial 20 bold font on a light grey background in the center of the screen. The average viewing distant was approximately 80 cm between screen and subject so that the visual angle occupied by each three-digit number would be about 2.0 × 2.0 degrees. For every subject, a new set of triplets was constructed. The stimulus presentation was controlled using the Presentation® software package (Neurobehavioral Systems, http://www.neurobs.com/).

Each session began with a short instruction and practice block of 64 items. After thorough familiarization (i.e., all subjects had the same amount of training), the main experiment took place. Each trial began with the presentation of the two delimiters for 1 500 ms. Then, the target was shown together with the delimiter for 1 500 ms. The triplet was replaced by a fixation cross (2 500 ms, Fig. 1). Subjects were instructed to press one of two buttons according to the target size (too small – left button, too large – right button) as fast and as accurately as possible. The button presses were executed with the index fingers of the left and right hand. In Experiment 2, each subject received both constant gaps in two separated sessions. The order of sessions was pseudorandomized across subjects.

Data analysis

Raw reaction time data were trimmed by eliminating responses exceeding the mean by more than two standard deviations, in order to reduce skew (Ratcliff 1993; Experiment 1: 5% of the data; Experiment 2: 5.7%). Paired t-tests were conducted comparing the error rates with chance performance (50%). Trimmed reaction time and error data were entered into a repeated-measure ANOVA with INTERVAL (110–150, 180–220, 250–290, and 320–360) and COMPATIBILITY for DECADES (compatible, incompatible) and UNITS (compatible, incompatible) as within-subject factors. Compatibility was defined between the delimiters, i.e., for the interval numbers 567 and 715 both decades (6 vs. 1) and units (7 vs. 5) were incompatible. For Experiment 2, the additional within-subject factor CONSTANT_GAP (small, large) was added. To assess whether the error rates or reaction times followed a linear trend, a contrast analysis was carried out. *Post-hoc* pairwise comparisons (corrected for multiple comparisons) were used to reveal significant differences. In an additional

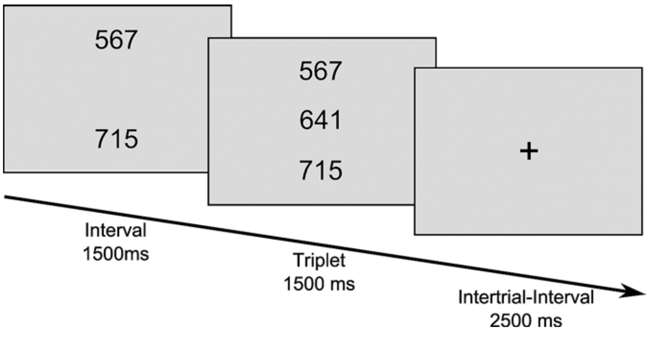

Fig. 1. Schematic display of the number decision task. After presenting the delimiters for 1 500 ms, the target was presented lasting 1 500 ms followed by the inter-trial interval (ITI).

[2] Two gaps (distances to the mean) were used to ensure that the task difficulty was neither too high nor too low. The different gaps were presented in different sessions.

[3] "Unit-Zero" effect: "A number ending with a zero prompts the same reaction time as a number ending with five and belonging to the same decade" (Dehaene et al. 1990), i.e., the distance effect postulates decreasing reaction times with increasing distance between two numbers (e.g., if 55 is the standard number, 58 is faster than 56) but this not true for numbers including a zero as unit (e.g., 60; see ibid. for further discussion of the effect).

Table II

Mean reaction times and error rates of Experiment 1, Experiment 2 and Experiment 3

Interval	Mean RT	S.E. RT	Mean Error	S.E. Error
Experiment 1 – Increasing distance to the mean (Three-digit numbers)				
1st	1 290 ms	61ms	38%	2%
2nd	1 229 ms	62 ms	32%	2%
3rd	1 196 ms	62 ms	24%	2%
4th	1 180 ms	58 ms	22%	2%
Experiment 2 – Constant distance to the mean (Three-digit numbers)				
1st	1 338 ms	125 ms	28%	2%
2nd	1 351 ms	125 ms	28%	3%
3rd	1 357 ms	125 ms	30%	2%
4th	1 368 ms	118 ms	37%	2%
Experiment 3 – Constant distance to the mean (Two-digit numbers)				
1st	1 020 ms	72 ms	28%	4%
2nd	1 097 ms	90 ms	30%	5%
3rd	1 134 ms	93 ms	37%	5%

(RT) Reaction time; (S.E.) standard error; detailed information about compatibility can be found in Figure 2.

ANOVA the spatial representation was investigated by adding the within-subjects factor RESPONSE (too small, too large)

For reaction time analyses, error trials were excluded (Experiment 1: 29%, Experiment 2: 30%).

To examine the size effect, the smallest and largest numbers were compared, i.e., based on the largest number of one triplet (e.g., for the triplet 567_641_715, the last number is the largest) we defined three different number magnitudes each comprising around 30% of all stimuli: 100–599, 600–799 and 800–999. In a *post-hoc* pairwise comparison we contrasted the error rates and reaction times for the smallest and largest number magnitudes.

Results

Errors

The percentage of errors for every condition and both experiments differed significantly from chance (all *P's*<0.001).

Experiment 1: The analysis of error rates showed that the percentage of error decreased with increasing interval size (i.e., larger distance between the two delimiters) while incompatible trials induced more errors than compatible. The ANOVA revealed significant main effects of INTERVAL, $F_{3, 81}$=36.08, P<0.001, and COMPATIBILITY DECADES, $F_{1,27}$=17.08, P<0.001, with more errors being made for incompatible trials (32%) in comparison to compatible trials (28%). The interaction between INTERVAL × COMPATIBILITY DECADES was also significant, $F_{3, 81}$=9.83, P<0.001. All other effects and interactions were not significant, F's<1.68. To assess the possibility of a linear distribution of the error rates, a contrast analysis was carried out which revealed linear trends for INTERVAL, $F_{1,27}$=62.09, P<0.001, COMPATIBILITY DECADES, $F_{1,27}$=17.08, P<0.001, and INTERVAL × COMPATIBILITY DECADES, $F_{1,27}$=13.56, P<0.001. In other words, the error rate decreased for increasing interval sizes with this effect being more pronounced for compatible than for incompatible decades. The *post-hoc* pairwise comparison revealed significant differences between all intervals (P<0.05) except third and fourth intervals (P=0.64), and a significant COMPATIBILITY DECADE effect for the fourth interval (19% vs. 32%; P<0.001; see

Table II and Fig. 2A upper panel). The second ANOVA on spatial representation revealed that REPONSE had a main effect, $F_{1,\ 27}$=8.44, P<0.01, with less error being made for "too small" (28% vs. 31%) and with a larger compatibility effect for "too large" in comparison to "too small" responses (RESPONSE × COMPATIBILITY DECADES, $F_{1,\ 27}$=6.48, P<0.05; 1.5% vs. 4.9%). The analysis of number magnitude was not significant (P=0.49).

Experiment 2: In contrast to the first experiment, error rates increased with increasing interval size while compatible trials induced less error. The ANOVA revealed a significant effect of INTERVAL, $F_{3,\ 45}$=4.86, P<0.01, and COMPATIBILITY DECADES, $F_{1,15}$=22.05, P<0.001, with more error being made for incompatible than compatible trials (28% vs. 33%). All other effects and interactions did not reach significance, F's<2.64 (see Table II and Fig. 2B upper panel). The contrast analysis revealed a linear trend for INTERVAL, $F_{1,15}$=7.14, P<0.05, and COMPATIBILITY DECADES, $F_{1,15}$=22.05, P<0.001, i.e., error rates increased with increasing interval size for compatible and incompatible trials. The *post-hoc* analyses showed significant differences for the first interval – small distance (25% vs. 32%) and for the second interval – large distance (25% vs. 31%). Analysis of RESPONSE revealed that subjects tended to make less errors for "too small" in comparison to "too large" ($F_{1,\ 6}$=4.04, P=0.09). The analysis of number magnitude revealed a significant difference between performance for the smallest vs. the largest numbers (28% vs. 32%; P<0.05).

Reaction times

Experiment 1: The pattern of reaction times was similar to error rates with decreasing reaction

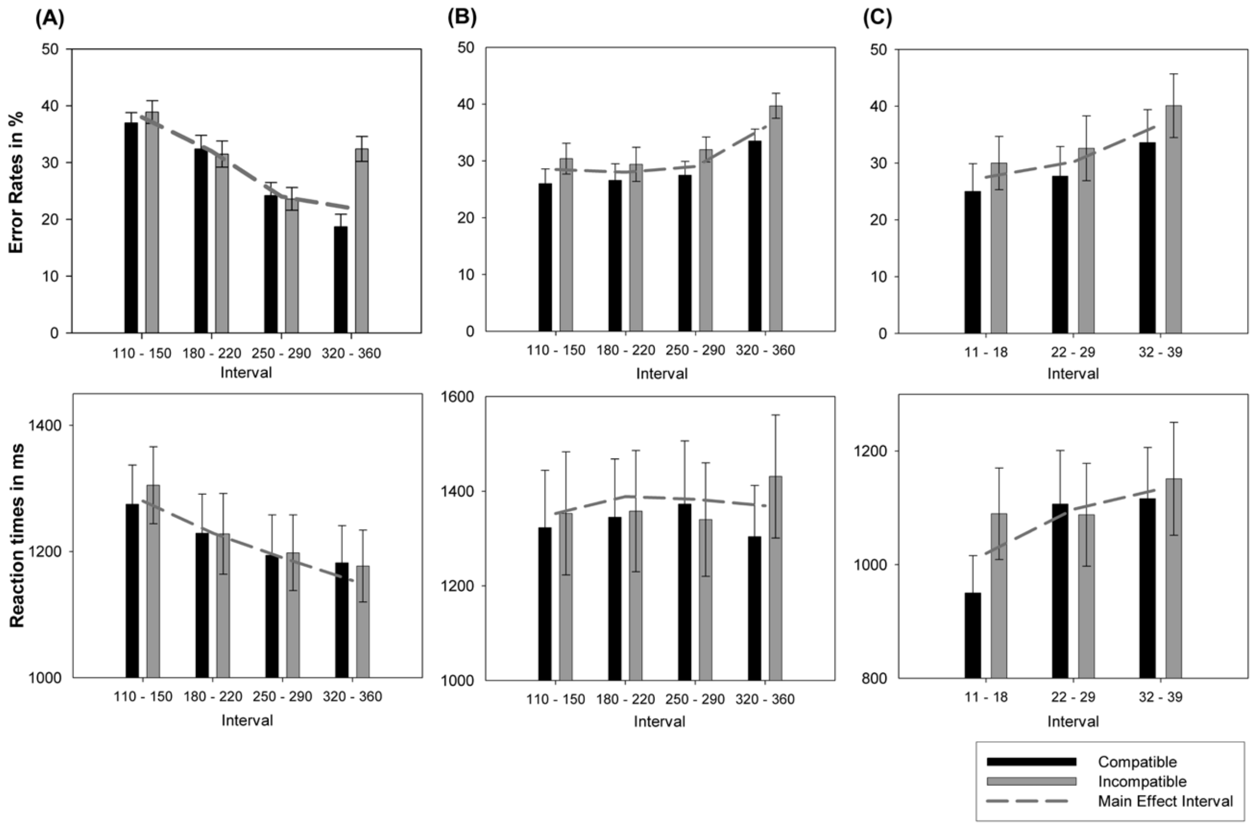

Fig. 2. Results of Experiment 1, 2 and 3. (A) Results of the first experiment (increasing target distance to the mean). (B) Results of the second experiment (constant target distance to the mean). (C) Results of the third experiment (two-digit numbers, constant target distance to the mean). The upper panel shows the error rates and the lower panel the reaction times. The compatibility refers to the decades as the units showed no influence on reaction time or error rates. The dashed line refers to the main effect of interval.

times from the first to the fourth interval. The ANOVA indicated a significant influence of INTERVAL, $F_{3,8}$=16.80, P<0.001, and a significant interaction of INTERVAL × COMPATIBILITY DECADES, $F_{3,81}$=5.52, P<0.005. All other effects and interactions were not significant, F's<2.61. A significant linear trend for INTERVAL, $F_{1,27}$=22.77, P<0.001, indicated that reaction times decreased with increasing interval size. *Post-hoc* analyses revealed that the first interval differed significantly from all other intervals (P<0.005) as reaction times were slowest for the first interval and fastest for the fourth interval. The COMPATIBILITY DECADES was only significant for the fourth interval (1 134 ms vs. 1 229 ms; P<0.001). The second ANOVA revealed a trend for the interaction between RESPONSE and COMPATIBLITY DECADES, $F_{1,26}$=2.97, P=0.10, with a larger compatibility effect for "too small" in comparison to "too large" (31 ms vs. 1 ms). The magnitude of numbers had no influence on performance. See Table II and Figure 2A lower panel for mean reaction times.

Experiment 2: Reaction time analysis revealed nearly no changes in reaction times for the different intervals. The ANOVA showed a significant interaction of INTERVAL × COMPATIBILITY DECADES, $F_{3,36}$=3.49, P<0.05, and a trend for the main effect of COMPATIBILITY DECADES, $F_{1,12}$=4.34, P=0.06, with compatible trials being faster than incompatible trials. All other main effects and interactions were not significant, F's<2.33, P>0.11. The contrast analysis showed no significant linear trends and RESPONSE had no influence on reaction times. However, there was a trend for reaction time difference related to number magnitude (1 406 ms vs. 1 429 ms; P=0.09)

Discussion

The first two experiments had the aim to validate and explore the effects of the NDT and to investigate the processing of multi-digit numbers.

In the first experiment, the results revealed that reaction times and error rates decreased with increasing interval size, i.e., the larger the distance between to delimiters the better the performance of subjects. These results suggest that if the gap between target and mean increases as a function of interval size, trials with large a gap between target and numerical mean (and consequently, large delimiter intervals) become the easiest trials (c.f., the distance effect; Moyer and Landauer 1967, Dehaene et al. 1990). In contrast, number magnitude had no influence on performance, i.e., for Experiment 1 the size effect was not found. Beside interval size, decade compatibility was highly significant for error rates and interacted with interval size for reaction times. Hence, for compatible trials a strong influence on subjects' decision could be found with performance improving with growing interval size. Taken together, the first experiment supports the hybrid model featuring "overall" as well as "decade" number representations whereas the decompositional model is unable to explain the results. However, it remained unclear whether increasing interval size or larger distances to the mean caused better performance.

To investigate this question we conducted a second experiment with constant gaps to the mean (gap: ±11–15 or ±32–36) to examine the effect of interval size. In contrast to the first experiment, we found decreasing performance with increasing intervals. Thus, when target gaps to the mean are held constant (independent of small or large constant gap), van Herten′s (1999) assertion, that reaction times and error rates increase with growing interval sizes, is supported (at least for error rates). Furthermore, taken number magnitude into account, the performance of subjects decreased with increasing number magnitude (c.f. size effect; Brysbaert 1995). Again, performance was better if decades were compatible but there was no influence of the compatibility of units.

In the first two experiments, stimuli were presented in a vertical arrangement. The results showed that increasing distances to the mean mask the effect of interval size, i.e., the distance effect is also true for three-digit numbers (Experiment 1), and that – when distances to the mean are held constant – increasing interval size leads to increasing error rates and reaction times with more errors being made for larger numbers (size effect; Experiment 2). Moreover, decade compatibility but not unit compatibility had an influence on behavior which supports the hybrid model (Nuerk and Willmes 2005) featuring an "overall" number representation and, in addition, a representation of decades on a separate MNL. However, the lack of a unit effect might be due to the fast presentation of the numbers as well as the size of the numbers, i.e., it was not necessary to process the unit number to solve the task. Furthermore, the overall pattern on spatial representation of multi-digit numbers analyzed by response format replicated earlier findings that subjects made fewer errors for numbers smaller than the real numerical mean, i.e., the target is easier to reject than targets that are larger than the numerical mean (e.g., Nuerk et al. 2002, Hoeckner et al. 2008). This result shows that effects on the MNL are comparable between the NBT and the NDT as both lead to a leftward bias in healthy controls ("pseudoneglect"; e.g., Lourenco and Longo 2009).

To exclude any influence of stimulus arrangement (i.e., the influence of perceptual encoding effects related to the presentation of numbers in terms of column-wise comparisons; Nuerk et al. 2004) we conducted two replication experiments using a horizontal arrangement of numbers (Experiment A and B in Appendix). Again, increasing performance was found for increasing target gap to the mean and decreasing performance for increasing interval sizes (with target gap to the mean held constant). However, the effect of target gap to the mean masking the impact of interval size was more pronounced when stimuli were presented vertically suggesting a strong influence of number magnitude on task performance (Zuber et al. 2009).

While the previous experiments demonstrated robust effects of target distance to the mean and interval size, the multi-digit numbers used in these experiments render the task too difficult for use in patient populations. As the NDT avoids common problems inherent in the number bisection task (see introduction), we performed a further experiment with two digit numbers to investigate visual-spatial perception and make it usable for patient groups. The rationale behind is that, for example, also patients with psychiatric disorders like depression or schizophrenia show a bias ("pseudoneglect") either to the left or right hemispace (e.g., Asthana et al. 1998, Cavezian et al. 2007a,b).

EXPERIMENT 3

Introduction

For the third experiment, the gap between the target and the numerical mean was held constant (±1–2). Furthermore, the timing was changed in order to simplify the task.

Method

Participants

Fourteen new healthy subjects (7 women; age = 29.4 years; S.D. = 2.9 years) were recruited from the staff of the RWTH Aachen University Hospital. All participants were native German speakers, right-handed according to the Edinburgh Inventory of Handedness (Oldfield 1971). The study was approved by the local ethical committee and participants all signed a written informed consent prior to participation.

Stimuli and design

In the modified version of the NDT, two-digit numbers were presented in a vertical arrangement with a constant target distance to the mean and varying interval sizes. First, an attention cue (fixation cross) was presented for 500 ms. Then the two delimiters that constitute the interval were shown for 1 000 ms followed by the target stimulus (1 000 ms). The number triplet was replaced by a hash mark that was shown for a jittered time range of 4 000–5 000 ms (Gilmore et al. 2007). Three conditions were presented (see Table I): first interval = 11–18, second interval = 22–29, third interval = 32–39. The target distance to the mean was held constant by adding or subtracting 1–2 from the real numerical mean.

Procedure and data analysis were identical to Experiment 1 (7% of the data were excluded because they exceeded the mean by more than two S.D.; 30% were excluded for reaction time analyses). For the pairwise comparison of number magnitude, numbers between 11 and 59 were compared with numbers between 80 and 99.

Results

Errors

Error rates were different from chance for all three intervals (all $P<0.05$) and increased with increasing interval sizes. Analysis of variance showed that INTERVAL had a significant influence on error rates, $F_{2,2}=3.70$, $P<0.05$. All other effects and interactions were not significant, $F<2.1$, $P>0.18$. Additionally, error rates increased in a linear fashion with increasing INTERVAL, $F_{1,11}=8.11$, $P<0.05$, with a significant difference between first and third interval (27% vs. 38%, $P<0.05$; see Fig. 2C upper panel). The analysis of responses revealed an interaction between RESPONSE × INTERVAL, $F_{2, 22}=3.36$, $P=0.06$, with the first interval showing less error for "too large" while second and third interval showed less error for "too small". Furthermore, larger numbers induced more errors than smaller numbers (28% vs. 33%; $P<0.05$).

Reaction times

The factor INTERVAL induced an increase in reaction time. INTERVAL, $F_{2, 22}=11.07$, $P<0.001$, and COMPATIBILITY, $F_{1, 11}=9.60$, $P<0.01$, had a significant influence on reaction times. The interaction between INTERVAL × COMPATIBILITY was significant, $F_{2, 22}=4.19$, $P<0.05$. Reaction times increased linearly from first to third interval, $F_{1, 11}=15.18$, $P<0.005$, with a fast increase between first and second interval for compatible trials and an increase for incompatible trials between the second and the third interval (interaction INTERVAL × COMPATIBILITY: $F_{2, 22}=6.22$, $P<0.05$). The pairwise comparisons revealed a difference between first and second interval (1 018 ms vs. 1 091 ms; $P<0.01$) and first and third interval (1 018 ms vs. 1 137 ms; $P<0.01$; see Table II and Fig. 2C lower panel). In the second ANOVA, the main effect of RESPONSE showed a clear trend, $F_{1, 11}=4.60$, $P=0.06$, with slower reaction times for "too small" in comparison to "too large" responses (1 111 ms vs. 1 065 ms). Number magnitude had no influence on reaction times.

Discussion

As expected, performance was worse with growing interval size if the target gap to the mean was held constant and in general, the leftward bias in subjects could be replicated with less errors being made for the "too small" response (only for second and third interval). For error rates, the size effect was replicated. Thus, the results are in line with the earlier experiments and underscore the importance of considering the target gap to the mean. Moreover, subjects were faster in compatible trials as compared to incompatible trials (mainly in the first interval) maybe because "the magnitudes of the decade and the unit distances influence the compatibility effect in a specific way and in a reverse direction. Small decade distances and large unit distances tend to lead to the largest compatibility effects" (Nuerk and Willmes 2005).

In an additional experiment (Experiment C in Appendix) we also investigated the influence of the target distance to the mean in the two-digit number version of the NDT. Again, a linear trend for decreasing reaction times with growing interval size was observed but no effect was found for error rates. However, also for two-digit numbers, there seems to be a "masking effect" of increasing target gap to the mean on interval sizes. Importantly, Experiment 3 demonstrates that the NDT works efficiently for two-digit numbers.

GENERAL DISCUSSION

The main aims of the current experiments were (a) the exploration and validation of the NDT as efficient tool to investigate numerical processing in its two- as well as three-digit versions and (b) to compare the different models of number representation. Thus, in three main Experiments we investigated the influence of interval size, target gap to the mean (distance effect), number magnitude (size effect), compatibility and spatial representation (response format). Experiment 1 showed decreased error rates and reaction times with increasing interval size, whereas Experiment 2 revealed that with constant target gaps to the mean, error rates and reaction times increased with growing interval size. Hence, target gap to the mean and distance effect, respectively, may mask the effect of interval. Results of Experiment 3 moreover showed that the modified version of the NDT with a vertical arrangement also works with two-digit numbers. Compatibility of decades is important for three-digit numbers as subjects were faster and more accurate in compatible trials (for a summary of all results, please see Table I in the Appendix).

Vertical stimulus arrangement yielded more clear-cut results than horizontal display (see additional results in the supplemental materials) maybe because reading direction is confounding results with horizontal arrangements. Indeed, data from different visuospatial tasks suggest that reading direction has an impact on performance (Kazandjian and Chokron 2008). Also, participants may be more prone to calculate rather than estimate their responses for horizontally arranged stimuli. However, the NDT seems to confirm classical effects of the NBT

Table III

Comparison of NDT and NBT

	NDT	NBT
ADVANTAGES		
Two response alternatives	x	x
Assessment of - reaction times - error rates	 x x	 x x
Suitable for different number sizes	x	x
Suitable for different patient groups	x	x
Assessment of classical number effects (e.g., size effect, distance effect, compatibility)	x	x
DISADVANTAGES		
Sensitive to factors independent from number processing (e.g., influence of presentation format)		x
Chance performance of 50%	x	x
Multiplicativity		x
Parity		x

(NBT) Number bisection task – Verification version; (NDT) Number decision task

(e.g., distance effect, size effect) by avoiding important disadvantages (see Table III for comparison of both tasks) and additionally, compared to the classical number comparison task, the information that can be obtained by the NDT are not limited as we are able to manipulate different conditions like interval size, gap between target and numerical mean, number magnitude and compatibility.

The second aim of the experiments was related to the organization of the MNL in terms of compatibility and different theoretical models. Importantly, the compatibility effects as well as the interaction between compatibility and increasing interval sizes support earlier results showing that numbers are not represented on one single MNL but rather on separate MNLs for unit and decade digits. Decade compatibility in particular had an influence on performance, highlighting the role of decades in numerical processing of three-digit numbers. In contrast to the study of Korvorst and Damian (2008), we did not find an effect of unit compatibility. In their study, they asked participants to compare two three-digit numbers (number comparison task) and choose the larger one while compatibility was manipulated. The results of the study showed that compatibility effects for units were only present when the decades were compatible (Korvorst and Damian 2008). In *post-hoc* analyses we therefore looked for an interaction of unit and decade compatibility but still could not find a unit compatibility effect. At the very least this suggests that unit compatibility exerts only a weak effect that may be dependent on the specific paradigm used.

In addition, the analysis of spatial representation *via* investigation of responses showed that subjects underestimate the real numerical mean and respond "too large" more often leading to more errors ("pseudoneglect"). Hence, the results support the compressive rather than linear representation of multi-digit numbers. First, the subjects estimate whether the target is too small or too large and hence, no precise determination of the numerical mean is necessary. Second, Lourenco and Longo (2009) suggested that linear scaling is only necessary when "precise discriminations are […] necessary for larger numerical values […] where compressive scaling would most certainly lead to biased judgments" like the ones we found in the current study.

Limitations

The aims of the study were the validation of the NDT as well as the investigation of representation of multi-digit numbers and the comparison of different models. However, with the current design we might not be able to exclude possible influences of arithmetic or estimation processes. Estimation can be defined as "ability to determine approximate numerostiy of stimulus sets in a manner that is appropriate when exact representations are […] not possible" (Beran et al. 2006). Following this suggestion, we would not expect specific effects like compatibility of decades or effects of decade because of the fuzziness of number representation. Hence, estimation processes might influence the current results but results are comparable with results from the classical NBT so that the impact of estimation is minimized.

CONCLUSIONS

The current experiments as well as the results by Korvorst and Damian (2008) argue for a decomposition of multi-digit numbers. The question remains whether the hybrid or the strict decomposed model better accommodates the current results. As the decomposed model proposes an explicit effect for decades (which we found) as well as for units (which we did not find), the current set of experiments rather argue for the hybrid model and a interaction of overall with individual magnitudes even for three-digit numbers.

More generally, the NDT has many advantages over the usual NBT. Neither the use of well-learned multiplication tables nor parity information has a beneficial influence on performance. The NDT is an efficient tool to investigate spatial (numerical) representations in healthy subjects as well as different patient groups (Cavezian et al. 2007b) and works in two and three digit versions. The current results support the hybrid model of Nuerk and Willmes (2005) as multi-digit numbers appear to be represented on several mental number lines that interact with each other.

ACKNOWLEDGEMENTS

We thank Klaus Willmes-von Hinckeldey for his help and inspiring discussions on the data and Sarah Bubenzer, Juliane Muehlhaus, Franziska Kintzel and Mario Senden for their assistance during data collection and analyses. Also, we thank Jane Banefield for her help on the manuscript. Finally, we wish to acknowledge the helpful comments of two anonymous reviewers.

Experiment 3 and Experiments A-C were supported by the START-Program of the Faculty of Medicine, RWTH Aachen University and IZKF Aachen (Interdisciplinary Center for Clinical Research within the Faculty of Medicine at the RWTH Aachen University, IZKF VV N3). KS is supported by the International Research Training Group (IRTG 1328) of the German Research Foundation (DFG) and funded by a grant from the German Research Foundation (DFG, SA 2221/3-1).

REFERENCES

Asthana HS, Mandal MK, Khurana H, Haque-Nizamie S (1998) Visuospatial and affect recognition deficit in depression. J Affect Disord 48: 57-62.

Beran JM, Taglialatela LA, Flemming TM, James FM, Washburn DA (2006) Nonverbal estimation during numerosity judgments by adult humans. Q J Exp Psychol (Hove) 59: 2065–2082.

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: 434-452.

Cavezian C, Danckert J, Lerond J, Dalery J, d'Amato T, Saoud M (2007a) Visual-perceptual abilities in healthy controls, depressed patients, and schizophrenia patients. Brain Cogn 64: 257–264.

Cavezian C, Rossetti Y, Danckert J, d'Amato T, Dalery J, Saoud M (2007b) Exaggerated leftward bias in the mental number line of patients with schizophrenia. Brain Cogn 63: 85–90.

Dehaene S (2001) Précis of " The number sense ". Mind and Language. 16: 16–36.

Dehaene S, Dupoux E, Mehler J (1990) Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. J Exp Psychol Hum Percept Perform 16: 626–641.

Dehaene S, Mehler J (1992) Cross-linguistic regularities in the frequency of number words. Cognition 43: 1–29.

Geppert BE (2005) Mental number representation as assessed by the number bisection task: a combined reaction time and fMRI study [Graduation thesis]. Faculty of Medicine. RWTH Aachen University, Aachen.

Gilmore CK, McCarthy SE, Spelke ES (2007) Symbolic arithmetic knowledge without instruction. Nature 447: 589–591.

Hoeckner SH, Moeller K, Zauner H, Wood G, Haider C, Gassner A, Nuerk HC (2008) Impairments of the mental number line for two-digit numbers in neglect. Cortex 44: 429–438.

Kazandjian S, Chokron S (2008) Paying attention to reading direction. Nat Rev Neurosci 9: 965; author reply: 965.

Korvorst M, Damian MF (2008) The differential influence of decades and units on multidigit number comparison. Q J Exp Psychol (Hove) 61: 1250–1264.

Liotti M, Mayberg HS (2001) The role of functional neuroimaging in the neuropsychology of depression. J Clin Exp Neuropsychol 23: 121-136.

Lourenco SF, Longo MR (2009) Multiple spatial representations of number: evidence for co-existing compressive and linear scales. Exp Brain Res 193: 151–156.

Moeller K, Fischer MH, Nuerk HC, Willmes K (2009a) Eye fixation behaviour in the number bisection task: evidence for temporal specificity. Acta Psychol (Amst) 131: 209–220.

Moeller K, Nuerk HC, Willmes K (2009b) Internal number magnitude representation is not holistic, either. Eur J Cogn Psychol 21: 672–685.

Moeller K, Pixner S, Kaufmann L, Nuerk HC (2009c) Children‘s early mental number line: logarithmic or decomposed linear? J Exp Child Psychol 103: 503–515.

Moeller K, Huber S, Nuerk HC, Willmes K (2011) Two-digit number processing: holistic, decomposed or hybrid? A computational modelling approach. Psychol Res 75: 290–306.

Moyer RS, Landauer TK (1967) Time required for judgements of numerical inequality. Nature 215: 1519–1520.

Nuerk HC, Geppert BE, van Herten M, Willmes K (2002) On the impact of different number representations in the number bisection task. Cortex 38: 691–715.

Nuerk HC, Weger U, Willmes K (2001) Decade breaks in the mental number line? Putting the tens and units back in different bins. Cognition 82: B25–33.

Nuerk HC, Weger U, Willmes K (2004) On the perceptual generality of the unit-decade compatibility effect. Exp Psychol 51: 72–79.

Nuerk HC, Willmes K (2005) On the magnitude representations of two-digit numbers. Psychology Science 47: 52–72.

Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9: 97–113.

Ratcliff R (1993) Methods for dealing with reaction time outliers. Psychol Bull 114: 510–532.

Ratinckx E, Brysbaert M, Fias W (2005) Naming two-digit arabic numerals: evidence from masked priming studies. J Exp Psychol Hum Percept Perform 31: 1150–1163.

Shepard RN, Kilpartric DW, Cunningham JP (1975) The internal representation of numbers. Cogn Psychol 7: 82–138.

van Herten M (1999) Number processing [Graduation thesis]. Department of Psychology, University of Maastricht, Maastricht, NL.

Verguts T, De Moor W (2005) Two-digit comparison: decomposed, holistic, or hybrid? Exp Psychol 52: 195–200.

Verguts T, Fias W (2004) Representation of number in animals and humans: a neural model. J Cogn Neurosci 16: 1493–1504.

Zorzi M, Priftis K, Umilta C (2002) Brain damage: neglect disrupts the mental number line. Nature 417: 138–139.

Zuber J, Pixner S, Moeller K, Nuerk HC (2009) On the language specificity of basic number processing: transcoding in a language with inversion and its relation to working memory capacity. J Exp Child Psychol 102: 60–77.

APPENDIX

Additional experiments on the Number Decision Task

The goal of the following experiments was to investigate whether the results of the main experiments were replicable when stimuli were presented horizontally, i.e. to exclude any possible influence of stimulus orientation. The prediction was that if the orientation of the triplet (i.e., horizontal or vertical) has no influence, the same results as in the first two experiments should be found. Otherwise, the results should differ in specific patterns like the target distance to the mean or the interval size. Both experiments (increasing target gap to the mean and constant target gap to the mean) were executed with horizontal stimulus presentation.

Experiment A

The experiment was equivalent to Experiment 1 but in contrast to the first experiment, the numbers were presented horizontally. We hypothesized that results of Experiment 1 should be replicated if triplet orientation has no influence on performance. Contradictory to our hypothesis, there was no "linear" decrease of error rates but error rates differed significantly between the first and the last interval, i.e., there was an "overall" trend for linearity of error rates (P=0.09; however effect of interval was significant: $F_{3,93}$=6.59; P<0.001). Reaction time analysis shows that reaction times decrease with increasing interval size (main effect of interval: $F_{3,93}$=6.47; P<0.001). This decrement in reaction time is especially pronounced between first and second interval (P<0.05). For compatibility, the results showed that subjects are faster (main effect of compatibility on RT: $F_{3,92}$=47.88; P<0.001) but not better in compatible in comparison to incompatible trials (interaction interval × compatibility: ($F_{3,96}$= 4.46; P<0.05). Although analyses yielded results that support our hypotheses, the general picture of the current experiment is not as clear as the results of vertical stimulus representation.

Experiment B

The aim of the experiment was the replication of Experiment 2. Thus, the distance between target stimulus and real numerical middle ranged from ±11 to ±15. Analyses showed that interval had a significant influence on error rates ($F_{3,96}$=8.88; P<0.001) and reaction times ($F_{3,96}$=5.56; P<0.005). In accordance with our hypothesis, error rates and reaction times increased in third and fourth interval with highest values in fourth interval. In contrast, error rates and reaction times decreased from first interval to second interval. One possible explanation could be that subjects tried to calculate the correct response in trials falling into first interval because the first interval seemed to be "easier" than the other three leading to enlarged reaction times. Although the results of the second interval are difficult to explain the general picture tends toward our expectations: subjects made more errors and need more time in the course of first to fourth interval. Regarding compatibility, there was neither an effect on reaction times nor on error rates.

Experiment C

The following experiment was done to expand the methodological establishment of the NDT as an efficient tool to investigate numerical processing of multi-digit numbers. Therefore, in contrast to the previous experiments, we used two-digit numbers and presented them vertically. In Experiment C, the distance between target and numerical mean was increased. Furthermore, timing was changed to simplify the task making it applicable in different subject populations and testing situations. We hypothesized that this modified version of the NDT (vertical presentation of two-digit numbers) works effectively with two-digit numbers, showing that the effect of interval size is masked by target distance to the mean. The results revealed that interval had a significant influence on error rates ($F_{2, 24}$=7.14; P<0.01) and reaction times ($F_{2, 24}$=39.94; P<0.001) with (mainly) increasing performance with growing interval size.

Table A1: Summary of the main results

Experiment	Target gap	presentation	Distance effect	Size effect	Compatibility	Spatial representation	Conclusion
1. Three-digit	increasing	vertical	significant effect with larger gaps between target and real numerical mean leading to better performance	no significant effect	influence of decades on error rates and in interaction with interval size on reaction times (mainly for compatible trials)	less errors for „too small“, i.e., leftward bias („pseudoneglect“; e.g., Lourenceo and Longo 2009)	Large gap between target and mean (in interaction with increasing interval size) lead to facilitation of task performance (cf. distance effect) and mask the effect of size
2. Three-digit	constant	vertical		significant effect with decreasing performance with increasing number magnitude	influence of decades on error rates and reaction times with compatible better than incompatible trials		Larger number magnitude induce a decrease in performance (cf. size effect)
3. Two-digit	constant	vertical		significant effect with decreasing performance (error rate) with increasing number magnitude	influence on reaction times with compatible better than incompatible trials		Larger number magnitude induce a decrease in performance (cf. size effect)
A. Three-digit	increasing	horizontal	significant effect with larger gaps between target and real numerical mean leading to better performance	no significant effect	influence of decades on reaction time with compatible better than incompatible trials		Large gap between target and mean (in interaction with increasing interval size) lead to facilitation of task performance (cf. distance effect) and mask the effect of size
B. Three-digit	constant	horizontal		significant effect with decreasing performance (error rate) with increasing number magnitude	influence of decades on error rates with compatible better than incompatible trials (P=0.10)	no effect of response format	Larger number magnitude induce a decrease in performance (cf. size effect)
C. Two-digit	increasing	horizontal	trend for distance effect with larger gaps between target and real numerical mean leading to better performance	reverse effect on reaction times with increasing number magnitude results in better performance	influence on error rates with compatible better than incompatible trials (P=0.09)	no effect of response format	Large gap between target and mean (in interaction with increasing interval size) lead to facilitation of task performance (cf. distance effect) and mask the effect of size

Notes: Experiment A–C can be found in the Appendix/ Distance effect: Defined as target gap to the real numerical mean (only measurable for Experiment 1, A and C)/ Size effect: Defined as increasing number magnitudes, based on comparison between the smallest and the largest 30% of stimuli/ Spatial representation: based on response format „too small and too large“