

NUMBER COMPARISON AND NUMBER LINE ESTIMATION RELY ON DIFFERENT MECHANISMS

Delphine Sasanguie* & Bert Reynvoet

The performance in comparison and number line estimation is assumed to rely on the same underlying representation, similar to a compressed mental number line that becomes more linear with age. We tested this assumption explicitly by examining the relation between the linear/logarithmic fit in a non-symbolic number line estimation task and the size effect (SE) in a non-symbolic comparison task in first-, second-, and third graders. In two experiments, a correlation between the estimation pattern in number line estimation and the SE in comparison was absent. An ANOVA showed no difference between the groups of children with a linear or a logarithmic representation considering their SE in comparison. This suggests that different mechanisms underlie both basic number processing tasks.

Introduction

The way that children and adults process number has attracted a lot of research interest in the past few years. Different paradigms have been used to investigate number processing in children, including number comparison (e.g., Holloway & Ansari, 2009) and number line estimation (e.g., Booth & Siegler, 2006). In a number comparison task, children have to indicate the larger/smaller of two presented numbers. Typically, a *distance effect* (DE) is observed in comparison (Moyer & Landauer, 1967): children are slower and less accurate in differentiating two numbers that are numerically close (e.g., 8 vs. 9) than two numbers that are numerically more distant (e.g., 1 vs. 9). This DE is accompanied by a *size effect* (SE): when the numerical distance is kept constant but the size of the numbers increases (e.g. 2 and 4 versus 8 and 10), error rates and response times increase with age and experience, the size of these effects decreases (e.g., Holloway & Ansari, 2009; Laski & Siegler, 2007). In a number line estimation task, two variants are possible: in a

* Delphine Sasanguie, Laboratory of Experimental Psychology, KU Leuven, Subfaculty of Psychology and Educational Sciences, KU Leuven Kulak; Bert Reynvoet, Laboratory of Experimental Psychology, KU Leuven, Subfaculty of Psychology and Educational Sciences, KU Leuven Kulak.

The authors thank all children, parents and schools that took part in this study.

We also want to thank Eva Van den Bussche for her helpful comments considering the statistical analyses. This research was supported by the Research Fund KU Leuven.

Correspondence concerning this article should be addressed to Delphine Sasanguie, Subfaculty of Psychology and Educational Sciences – KU Leuven, Kulak, Etienne Sabbelaan 53; 8500 Kortrijk, Belgium. E-mail: Delphine.Sasanguie@kuleuven-kulak.be

number-to-position (NP) task, children have to place a number on an empty number line and in a position-to-number (PN) task, they are shown a position on a number line and asked to estimate the number that corresponds to it. Children's estimates have been observed to increase logarithmically with numerical magnitude on the NP task and exponentially with numerical magnitude on the PN task (Siegler & Opfer, 2003). Most studies however, use the number-to-position variant to investigate the underlying representation (Booth & Siegler, 2006; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013). In this task, young children initially generate estimates that closely fit a *logarithmic function* (i.e. small numbers are put further away from one another on this empty line than larger numbers). With increasing age, children learn to use the number line linearly, although the age at which this criterion is reached depends on the scale of the number line (e.g., Bertelletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010).

The DE and the SE in comparison and the logarithmic mapping in number line estimation are explained by assuming a *common* underlying magnitude representation, i.e. a left-to-right oriented "mental number line" (Dehaene, 1997; Gallistel & Gelman, 1992). On this mental number line, magnitudes are represented as a Gaussian distribution around the true location of each specific number, with partially overlapping representations for nearby numbers. This implies that, whenever a number is presented, not only the representation of that specific number will be activated, but also partially the representation of the nearby numbers. Moreover, the overlap between nearby numbers increases with increasing number size, resulting in a logarithmic compressed number line. Such a representational organisation leads to more difficult discrimination of nearby numbers (i.e. DE), increased reaction times for larger numbers (i.e. SE), and a logarithmic mapping in the number line estimation task in young children. It is suggested that with increasing age, the logarithmic compression gradually disappears and the mental number line becomes linear and more precise. This developmental change results in smaller DEs and SEs (e.g., Holloway & Ansari, 2009; Laski & Siegler, 2007; Sasanguie, De Smedt, Defever, & Reynvoet, 2012) and a better linear fit on the number line estimation task (e.g., Booth & Siegler, 2006; Siegler & Booth, 2004).

To test the hypothesis of a common mechanism responsible for the findings in comparison and number line estimation, Laski and Siegler (2007) examined the performance on both tasks in a sample of children. They observed decreasing DEs and SEs in comparison and more reliance on a linear fit in number line estimation with increasing age. Moreover, a correlation was found between the accuracy in the comparison task and the individual variance of estimations in the number line estimation task, supporting the idea of a common underlying representation in both tasks. However, we believe that this conclusion is premature. First, older children and adults typically use

linear mappings in the number line task. If number is represented linearly, no SE in comparison should be present; small numbers and large numbers are equally discriminable. Contrary to this expectation, a SE is still observed in comparison tasks (Laski & Siegler, 2007; Schwarz & Stein, 1998). This inconsistency between both tasks is difficult to explain on the basis of a common mechanism. Second, in the study by Laski and Siegler (2007), the association between the comparison and the number line estimation task is investigated by computing the correlation between the comparison accuracy and the child's estimates. However, overall accuracy performance in comparison is not a direct manifestation of the underlying logarithmic compression of the mental number line; the DE and the SE are. Moreover, although the performance in number comparison and number line estimation are thought to reflect characteristics of the same underlying mental number line, recent work also pointed towards a different origin of these effects (e.g., connectionist models about comparison; Verguts, Fias, & Stevens, 2005; and non-numerical processes such as proportion judgments – Barth & Paladino, 2011, attention resources – Anobile, Cicchini, & Burr, 2012 or strategy use – Petitto, 1990; White & Szűcs, 2012 in case of number line estimation performance). Finally, Laski and Siegler (2007) presented symbolic numbers between 0-100. As already indicated by the authors themselves, the performance in a comparison task with these numbers is influenced by unit-decade compatibility effects typically observed in comparison (e.g., Nuerk, Weger, & Willmes, 2001), whereas such compatibility effects do not affect the performance in number line estimation. The presence of this compatibility effect in comparison can be avoided by using non-symbolic stimuli (Moeller, Klein, Nuerk, & Cohen-Kadosh, 2012).

In the present study, we therefore re-examined the correlations between the indices of mental number line compression in a number comparison task and a number line estimation task, both with non-symbolic stimuli (i.e. dot patterns). More precisely, we investigated whether the individual variance in the SE (i.e. the difference between reaction times on large versus small trial pairs – a more direct manifestation of the underlying logarithmic compression of the mental number line than overall accuracy –) on the one hand, and the logarithmic and the linear fit on a number-to-position variant of the number line estimation task on the other hand, were related in children. If performance in comparison and number line estimation is driven by the same underlying representation, a positive correlation between the *logarithmic* fit in number line estimation and the reaction time difference on large and small trial pairs is expected: more logarithmic compression should result in higher reaction times for larger pairs. Moreover, a negative correlation between the *linear* fit in number line estimation and the difference between large and small trial pairs is predicted: the more linear children their estimation pattern

is, the smaller the reaction time difference between the large and small number pairs is expected to be. We also explicitly tested whether there is a difference between children relying on a logarithmic versus a linear estimation pattern in a number line estimation task and their performance on the comparison task. According to Laski and Siegler's (2007) claim, it is expected that children relying on a logarithmic representation will show a larger DE or SE in the comparison task^[1], while this difference is expected to be absent in case of the children with a linear estimation pattern (see Figure 1).

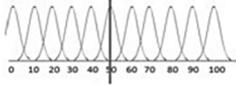
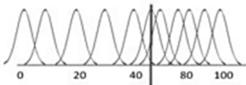
Experiment 1		
REPRESENTATION IN NUMBER LINE ESTIMATION TASK	COMPARISON TASK	PREDICTIONS
LINEAR 	RT on small pairs (1-4) = RT on large pairs (6-9) → smaller SE	Negative correlation $R^2_{lin} - SE$
LOGARITHMIC 	RT on small pairs (1-4) < RT on large pairs (6-9) → larger SE	Positive correlation $R^2_{log} - SE$
Experiment 2		
LINEAR 	RT on small pairs (smaller than 24) > RT on large pairs (larger than 24) → larger SE	Positive correlation $R^2_{lin} - SE$
LOGARITHMIC 	RT on small pairs (smaller than 24) = RT on large pairs (larger than 24) → smaller SE	Negative correlation $R^2_{log} - SE$

Figure 1

Visual representation of the hypotheses of Experiment 1, based on Laski and Siegler's (2007) claim. On top, a linear representation is presented. Below, a logarithmic representation is visualised

1. However, to test our hypothesis for the DE, three difference scores are needed (i.e. a difference score between the large and the small distances within the small trial pairs, a difference score between the large and the small distances within the large trial pairs and a difference score between these two first difference scores) and we believe that this measure, based on three difference scores, contains a lot of noise and might be unreliable (Siegrist, 1995; Strauss, Allen, Jorgensen, & Cramer, 2005). In a difference score, the measurement errors of both compounds are combined, reducing the possibility to evoke a significant correlation (Strauss et al., 2005). Therefore, we chose to only use the SE (which is only one difference score between large and small trial pairs) as a reflection of the underlying logarithmic compression of the mental number line.

Experiment 1

Method

Participants

Ninety-two participants were recruited from an elementary school in Flanders, Belgium, in a middle to higher income neighbourhood. Subjects that were outliers ($>3SD$ below or above the group average) in one of the experimental tasks were removed ($N = 5$). The final sample consisted of 87 typically developing children, comprising 28 first graders (16 males, $M_{\text{age}} = 6.6$ years, $SD = .33$), 29 second graders (10 males, $M_{\text{age}} = 7.7$ years, $SD = .29$) and 30 third graders (13 males, $M_{\text{age}} = 8.6$ years, $SD = .25$).

Materials and Procedure

Children were tested in small groups of 7 to 10 children in a quiet room accompanied by two experimenters. All children first completed the number line estimation task, followed by the number comparison task. To prevent fatigue, a short break between the two tasks was provided.

Number line estimation task. This task was administered with pen and paper. Children were presented with 25 cm long lines in the centre of white A4 (210 × 297 mm) sheets. Dot patterns were generated with a MatLab script (Dehaene, Izard, & Piazza, 2005) that controlled for total area in order to prevent consistent use of this feature (larger numerosities contained smaller dots). Beginning and end points were labelled by a figure of 0 dots on the left and 100 dots on the right respectively. The to-be-positioned dot pattern was shown in the centre of the sheet, 6 cm above the number line. The dot patterns shown were 2, 3, 4, 6, 18, 25, 48, 67, 71, 86 (corresponding to sets A and B for the same interval used in Siegler and Opfer, 2003). A different random order of presentation for the dot patterns was generated for each child and each line was presented on a separate sheet. Children were instructed to mark on the line where they thought the quantity had to be positioned. To ensure that the child was aware of the interval size, an example was provided by the experimenters solving the first item of the task while saying: “This line goes from 0 dots to 100 dots. If here is 0 and here is 100, where would you position this number of dots?”. After that, the children were able to go through all sheets at their own pace.

Number comparison task. This task was presented on laptops with 14-inch screens. Stimulus presentation and the recording of behavioural data were controlled by E-prime 1.1 (Psychology Software Tools, <http://www.pst->

net.com). Participants had to select the larger of two dot patterns that were presented on the left and on the right side of the screen, by pressing a key at the side of the largest quantity ('a' and 'p' on an AZERTY keyboard). Stimuli were two white-filled circles (radius 3.5 cm) each containing a dot pattern, simultaneously presented on a black background. Children were asked to respond as quickly as possible without making errors. Five practice trials were included, to make the children familiar with the task requirements. Stimuli involved dot patterns ranging from 1 through 9, but only combinations of stimuli with a maximum distance of 5 were presented to the children, which resulted in 60 trials. A trial started with a fixation cross for 600 ms, after which the two stimuli that had to be compared appeared. Stimuli remained on the screen until the child responded. The inter-trial interval was 1000 ms.

Results

Number line estimation task

The percentage absolute error (PAE) was calculated per child as a measure of children's estimation accuracy. This was done by using the following formula by Siegler and Booth (2004):

$$\left| \frac{\text{Estimate} - \text{Estimate Quantity}}{\text{Scale of Estimates}} \right|$$

For example, if a child was asked to estimate 18 on a 0-100 number line and placed the mark at the point on the line corresponding to 30, the PAE would be $(30-18) / 100$ or 12%. Table 1 shows the mean PAE per grade. Accuracy on the number line task increased with grade, $F(2,84) = 4.43$, $p = .02$, $\eta_p^2 = .10$.

We further analysed children's estimation patterns by fitting linear and logarithmic functions for each individual child (see also Siegler & Opfer, 2003). Then, a paired t -test was conducted on the mean R^2 linear and mean R^2 logarithmic for each grade, in order to investigate whether there was a difference. The model with the highest R^2 was logarithmic for the first graders ($R^2_{\log} = .87$), but did not differ from the model with the linear fit ($R^2_{\text{lin}} = .86$; $t(27) = -.46$, $p = .65$). For the second graders, the fit of the linear model was the best ($R^2_{\text{lin}} = .89$) and also did not differ from the logarithmic fit ($R^2_{\log} = .88$; $t(28) = .74$, $p = .47$). For the third graders, the linear model clearly fit the best ($R^2_{\text{lin}} = .97$), but also did not differ significantly from the logarithmic model ($R^2_{\log} = .86$; $t(29) = 1.89$, $p = .07$).

Table 1
Mean percent absolute error (PAE) and the corresponding standard deviations of the number lines, per grade

Number line estimation task	
First graders (N=28)	.15 (.07)
Second graders (N=29)	.13 (.05)
Third graders (N=30)	.11 (.05)

Number comparison task

We computed adjusted reaction times (RT) to reflect both speed and accuracy (ACC) of performance in one measure by combining the median reaction times and mean error rates using the formula $RT/(1-\text{error rate})$ or RT/ACC . This way, the reaction times remain unchanged with 100% accuracy and increase in proportion with the number of errors (see Luculano, Tang, Hall, & Butterworth, 2008 for a similar method^[2]). The adjusted RTs are shown per distance and per grade in Table 2.

The adjusted RTs were submitted to a repeated measures analysis of variance (ANOVA) with distance as within-subject variable (5 levels) and grade as between-subjects factor (3 levels). There was a main effect of distance, $F(4,81) = 123.83, p < .0001, \eta p^2 = .86$, showing faster latencies with increasing distance. There was also a main effect of grade, $F(2,84) = 13.27, p < .0001, \eta p^2 = .24$, indicating that the reaction times decreased with increasing grade. Furthermore, there was a distance by grade interaction, $F(4,82) = 2.34, p = .02, \eta p^2 = .10$: differences between grades were more pronounced at the smallest distances, but the distance effects remained significant in each of the grades separately, $F(4,24) = 36.13, p < .001, \eta p^2 = .86$ for the first grade; $F(4,25) = 28.84, p < .001, \eta p^2 = .82$ for the second grade and $F(4,26) = 81.69, p < .001, \eta p^2 = .93$ for the third grade.

To quantify individual differences in the SE, we calculated the SE by subtracting the adjusted RT on small pairs (both numbers less or equal than 4) from the adjusted RT on large pairs (both numbers more or equal to 6). The RT difference was then divided by the overall adjusted RT. This mean calculated SE (.32; $SD = .23$) was significantly different from zero, $t(86) = 13.10, p < .001$. No grade differences on the SE were present, $F(2,84) = 2.05, p = .14, \eta p^2 = .05$: the mean SEs were .31 ($SD = .22$), .27 ($SD = .24$) and .39 ($SD = .21$) for the first-, second- and third graders respectively.

2. Bruyer and Brysbaert (2011) suggested that it is not a good idea to limit the analyses to adjusted RTs without further checking the data. Therefore, we reanalysed the data of both Experiment 1 and Experiment 2 using unadjusted RTs instead of adjusted ones. We observed that none of the results changed.

Table 2

Adjusted reaction times (ms) and corresponding (standard deviations) for the five distances of the number comparison task, per grade

	Number Comparison task				
	d1	d2	d3	d4	d5
First graders (N=28)	1535.84 (368.53)	1232.48 (260.03)	1155.47 (274.66)	1019.44 (201.17)	976.39 (204.01)
Second graders (N=29)	1274.07 (256.94)	1025.20 (221.40)	921.99 (150.83)	902.48 (221.94)	863.85 (185.48)
Third graders (N=30)	1227.90 (185.72)	987.82 (169.37)	874.59 (155.03)	817.60 (131.73)	797.64 (145.21)

Relation between both number processing tasks

Correlation analysis. A correlation analysis was conducted to explore the relation between the performance on both tasks. For the number line estimation task, the individual logarithmic (R^2_{\log}) and the linear fit (R^2_{lin}) were used as indices. For comparison, the calculated SE was used. Confidence intervals (CIs) of the obtained correlation were calculated according to the formula suggested by Quertemont (2011, p. 126). This immediately gives readers information about the highest and lowest correlations at the population level. If the entire confidence interval falls below a sensible threshold value, it will be possible to conclude that the correlation at the population level is negligible. The correlation between the logarithmic function and the performance on the comparison task was not significant: $r(85) = -.07, p = .50$; 95% CI = [-.28; .14]. Similarly, the correlation between the linear fit and the performance on the comparison task was not significant: $r(85) = -.09, p = .43$; 95% CI: [-.30; .12]^[3].

Independent samples t-test. Children were classified as linear or logarithmic depending on the best fitting model (i.e. the highest significant R^2 , see Siegler & Opfer, 2003). Moreover, if both models failed to reach significance, children were classified as having ‘no representation’ (see Berteletti et al., 2010

- When these correlations were examined per grade, the pattern of the results did not change. Because the small sample sizes per grade imply that the results of the correlation analysis might be more sensitive to outliers, Cook’s values were calculated. To interpret Cook’s distance, the sensitive cutoff value of $4/n$ recommended by Bollen and Jackman (1990) was used. Based on this criterion, one participant was removed from the first grade, three participants from the second grade and two participants from the third grade. None of the correlations per grade reached significance: the correlations between the logarithmic function and the SE in comparison were $r(25) = -.22, p = .27$ for the first grade, $r(24) = -.28, p = .16$ for the second grade and $r(26) = -.05, p = .82$ for the third grade respectively. With regards to the linear fit and the SE, correlations were $r(25) = .03, p = .90$ for the first grade, $r(24) = .04, p = .74$ for the second grade and $r(26) = -.29, p = .13$ for the third grade respectively.

for a similar method). In case of the first graders, 13 children were classified as having a linear representation, 14 as having a logarithmic one and 1 child as having no representation. From the second graders, 17 children had a linear and 12 had a logarithmic representation and from the third graders, 20 children had a linear and 10 a logarithmic representation. An independent samples *t*-test was conducted to examine the differences in the performance on the comparison task between the group of children who were classified as having a linear representation and the group of children who were classified as having a logarithmic representation. No significant difference could be detected between both groups, $t < 1$. The mean SEs were .32 ($SD = .23$) for the children with a linear representation ($N = 50$) and .33 ($SD = .24$) for the children with a logarithmic representation ($N = 36$).

Discussion

The results of Experiment 1 do not confirm the hypotheses we described in the introduction of this study. No correlation was observed between the logarithmic or the linear fit in number line estimation and the reaction times on small and large trial pairs (i.e. the SE) in number comparison. Moreover, no difference in the SE observed in the comparison task could be detected between children who relied on a logarithmic representation and children who relied on a linear representation. This suggests that the effects obtained with these two different tasks do not stem from a common underlying mechanism.

It should be noted that the different number ranges used in the number comparison and the number line estimation tasks may have led to the absence of a correlation between the two tasks. In the comparison task, small numbers (1-9 dots) were used, whereas in the number line estimation task, stimuli between 0 and 100 dots were presented. However, it should be pointed out that the use of different number ranges makes our results even more convincing. In the present study we observed that most of the children already performed linearly on the 0-100 number line task. Previous studies (e.g., Berteletti et al., 2010) have shown that linear mappings occur even earlier with smaller number lines (e.g., 1-10). On the basis of these previous findings, no SE in comparison is expected when only numbers up to 9 are shown. This expectation is however not met: a SE is present. Although we believe that the different number ranges are even more convincing to argue in favour of different mechanisms in both tasks, we conducted an additional experiment in which large numerosities were shown in the number comparison task to test the dependency of both tasks directly.

In a typical comparison task with large non-symbolic stimuli, stimulus pairs are created on the basis of relative distance (i.e. ratios) instead of abso-

lute distance (e.g., Inglis, Attridge, Batchelor & Gilmore, 2011; Libertus, Feigenson & Halberda, 2011) to compensate for logarithmic compression (and the SE that is said to be a result of this compression). In line with these studies, we also opted for a design where stimulus pairs were created on the basis of ratios. Such a manipulation, however, also changes the predictions if one assumes a common underlying mechanism in both tasks (see figure 1). In contrast to Experiment 1, a positive correlation between the linear fit in number line estimation and the reaction time difference between large and small trial pairs is expected: if children rely on a linear representation, larger number pairs are further away from one another than smaller number pairs, leading to faster latencies for larger pairs. Moreover, a negative correlation between the logarithmic fit in number line estimation and the reaction time difference between large and small trial pairs is predicted: if children have perfect logarithmic mappings, no RT difference is expected between small and large number pairs. Furthermore, when explicitly testing the difference between children relying on a logarithmic versus a linear estimation pattern in a number line estimation task considering the performance on the comparison task, it is predicted that children relying on a linear representation will show a large difference between large and small trial pairs on the comparison task – due to the faster latencies on the larger trial pairs –, while this difference will be absent in case of the children with a logarithmic estimation pattern.

Experiment 2

Method

Participants

Experiment 2 included 28 Flemish second graders (19 males, $M_{\text{age}} = 8.1$ years, $SD = .37$) recruited from an elementary school in a middle income neighbourhood. This age group was chosen because in Experiment 1 about half of the second graders performed logarithmically on the 0-100 number line estimation task, while the other half performed linearly. Moreover, this group was preferred above the group of first graders, because in this second experiment, the level of the number comparison task was more difficult. Subjects that were outliers ($>3SD$ below or above the group average; $N = 2$) in one of the experimental tasks were removed from the original subject group.

Materials and Procedure

Children were tested in groups of 10-15 in a quiet room, accompanied by two experimenters. All children first completed the non-symbolic number com-

parison task, followed by the non-symbolic number line estimation task. To prevent fatigue, a short break between the two tasks was provided.

Number comparison task. The number comparison task was conducted using laptops with 14-inch screens. Stimulus presentation and the recording of behavioural data were controlled by E-prime 1.1 (Psychology Software Tools, <http://www.pstnet.com>). Participants had to select the larger of two presented dot patterns, one on the left and one on the right side of the screen, by pressing a key at the side of the largest dot quantity ('a' and 'p' on an AZERTY keyboard). Children were asked to respond as quickly as possible without making errors. Five practice trials were included, to make the children familiar with the task requirements. Stimuli were pairs of dot patterns presented in white on a black background, generated by a Matlab script developed by Gebuis and Reynvoet (2011) that is perfectly suited to control for continuous visual variables with large non-symbolic stimuli. In this program, 5 visual properties are manipulated: (1) the area subtended (i.e. the smallest contour around the dot pattern), (2) the total surface of the dots, (3) the density (i.e., the total surface divided by the area subtended), the (4) the average diameter of the dots and (5) the total circumference (i.e. the total contour length of all dots). Regression analyses confirmed that there was no relationship between each visual cue and numerosity, all R^2 s < .03, all p s > .11. The different visual cues of the stimuli co-varied positively with numerosity in half of the trials and negatively with numerosity in the other half. One of the stimuli always contained the reference number of 24 dots. The other dot pattern contained either 12, 16, 19, 30, 36 or 48 dots, resulting in 3 different numerical distances (ratios of 1.25, 1.5 and 2). Each of the pairs was repeated ten times, resulting in 60 trials. A trial started with a fixation cross for 600 ms, after which the two stimuli that had to be compared appeared. Stimuli were presented simultaneously in the centre of the screen and remained on the screen until the child responded. The inter-trial interval was 1000 ms.

Number line estimation task. The number line estimation task was administered with pen and paper and was identical as in experiment 1.

Results

Number line estimation task

The model with the highest R^2 was linear for this group of second graders ($R^2_{\text{lin}} = .788$), but did not differ from the model with the logarithmic fit ($R^2_{\text{log}} = .787$; $t(27) = .03$, $p = .98$).

Number comparison task

A repeated measures analysis of variance (ANOVA) with distance (3 levels) as the within-subject factor on children's adjusted reaction times (i.e. RT/ACC) was conducted. There was a main effect of distance, $F(2,26) = 43.41$, $p < .0001$, $\eta p^2 = .77$, showing decreasing RTs with increasing numerical distance. The adjusted RT for ratio 1.25 was 1900.98ms ($SD = 681.44$ ms), for ratio 1.5 it was 1729.80ms ($SD = 507.26$ ms) and for ratio 2, the adjusted RT was 1288.40ms ($SD = 349.86$ ms).

As an index for the large number comparison task, a calculated difference measure similar to that of Experiment 1 was used. The difference measure in this case was calculated by subtracting the adjusted RT on the pairs smaller than 24 (i.e. 12-24, 24-12, 16-24, 24-16, 19-24 and 24-19) from the adjusted RT on the pairs larger than 24 (i.e. 24-30, 30-24, 24-36, 36-24, 24-48 and 48-24). The difference RT was then divided by the overall adjusted RT. Unexpectedly, this mean calculated SE (-.13; $SD = .14$) was significantly different from zero, $t(27) = -4.77$, $p < .001$, indicating that the participants responded faster to the trial pairs larger than 24.

Relation between both number processing tasks

Correlation analysis. A correlation analysis was conducted to explore the relation between the performance on both tasks. For the number line estimation task, the individual logarithmic (R^2_{\log}) and the linear fit (R^2_{lin}) were used as indices. For comparison, the calculated SE was used. Similar as in Experiment 1, confidence intervals were calculated according to the formula provided by Quertemont (2011). There was no significant correlation between the logarithmic function and the performance on the comparison task, $r(26) = -.20$, $p = .31$, 95% CI = [-.60; .19]. Neither was the correlation between the linear fit and the performance on the comparison task significant, $r(26) = -.21$, $p = .27$, 95% CI = [-.61; .18].

Independent samples t-test. Regression analyses on the individual data of the children (similar method as in Experiment 1, see also Berteletti et al., 2010) showed that 14 children were classified as having a linear representation, 10 as having a logarithmic one and four children as having no representation. An independent samples *t*-test was conducted to examine the differences in the performance on the comparison task between the group of children who were classified as having a linear representation and the group of children who were classified as having a logarithmic representation. No significant difference could be detected between both groups, $t(22) = -1.37$, $p = .18$. The mean difference scores were -.17 ($SD = .14$) and -.09 ($SD = .12$) for the children

with a linear representation and the children with a logarithmic representation respectively.

General Discussion

Number comparison and number line estimation are common tasks to investigate the development of number processing in children. Performance on the comparison task results in a size effect (e.g., Holloway & Ansari, 2009; Schwarz & Stein, 1998): when numerical distance is kept fixed, discrimination performance is more difficult when the numerical size increases. The performance of young children on a number line estimation task is characterised by a logarithmic mapping (e.g., Siegler & Opfer, 2003): small numbers are put further away from one another on this empty line than larger numbers. The performance in both tasks is assumed to rely on the same underlying representation taking the form of a compressed mental number line (Laski & Siegler, 2007). An explicit test for this assumption is to examine the relation between the performance in both tasks. A more linear representation as measured in the number line task, should result in smaller SEs. We investigated this in first-, second-, and third graders with a non-symbolic number comparison and number line estimation task.

In both experiments of the present study, a relation between the SE in the comparison task and the performance on a number line estimation task was absent. The absence of a relation between the SE in the comparison task and the performance on a number line estimation task is not in line with the assumption of a common underlying principle in both tasks (Laski & Siegler, 2007) and suggests that different mechanisms may play a role in both tasks. Previous studies have indeed argued in favour of task-specific mechanisms to account for the performance in the number line estimation task and number comparison. For instance, Barth and Paladino (2011) demonstrated that children rely on the beginning and ending marks of the number line to estimate the correct position on a number line. This way, children estimate the part (e.g. 30) of a whole (e.g. 100), or in other words, a proportion. According to these authors, the requirement of a proportion judgment in this task provides a better explanation than does the idea of a logarithmic-to-linear representational shift. In another study, Anobile, Cicchini and Burr (2012) showed that when adults are distracted by a concurrent task, they perform logarithmically on a non-symbolic number line estimation task. This finding led the authors to suggest that the non-linearity arises from an intrinsic logarithmic representation and that attentional resources play a very important role in making the shift to linear representations. Both studies suggest that additional non-numerical processes may moderate the performance in the number line task

and suggest that number line mapping is not a pure reflection of the underlying representation.

Based on the present results, also the true nature of the SE in comparison can be questioned based. The SE is assumed to reflect the compressive nature of the number line. This compression can be controlled for by presenting small and large trial pairs with the same ratio (instead of same absolute value) and should result in the absence of a SE when performance on small and large trial pairs is compared. However, the results of Experiment 2 showed that despite the use of the same ratios in the small and large trial pairs, a SE was still present, which is not in line with the idea of a logarithmic scaling of the number line. This finding might be explained by either an even more compressed number line, or alternatively, the SE might be originated by alternative mechanisms. For instance, on the basis of connectionist modelling, Verguts and colleagues (Verguts et al., 2005; Van Opstal, Gevers, De Moor, & Verguts, 2008) demonstrated that the SE can be explained by decisional processes in comparison. They showed that representational overlap is not a necessary condition for this effect to emerge. The SE can alternatively be explained by the compressive pattern of connections between the magnitude representation and the response nodes (Verguts et al., 2005).

It should be noted that our conclusion that different mechanisms are involved in both tasks, is based on the absence of an effect. One factor that might have lead to the absence of a correlation, is the reliability of the tasks. The majority of studies, certainly studies including the number line estimation task, have used Arabic digits and not non-symbolic dot patterns. Moreover, this is also true for the Laski and Siegler's study (2007) who argued in favour of a common underlying mechanism in both task. So perhaps our non-symbolic variant is less reliable, leading to null results. However, we have strong arguments to believe that both tasks we used are highly reliable and valid. First, Gilmore, Attridge and Inglis (2011) found significant split-half reliability coefficients in non-symbolic comparison tasks, very similar to the one used in this study (see also Maloney, Risko, Preston, Ansari, & Fugelsang, 2010; Sasanguie, Defever, Van den Bussche, & Reynvoet, 2011). Second, the fact that the performance of children on the non-symbolic number line task is significantly related to their performance on a symbolic number line task (e.g., Sasanguie, De Smedt et al., 2012) and to their math achievement scores (e.g., Sasanguie, De Smedt et al., 2012; Sasanguie, Van den Bussche et al., 2012), demonstrates both the reliability and the validity of the non-symbolic number line task. Another factor that could be responsible for the absence of a positive finding is a power problem (i.e. insufficient number of participants to detect significant correlations). To verify this possibility, we calculated the confidence intervals of the correlations and statistically showed that also at the population level, it is very unlikely that there exists a relation

between the performance on both tasks (see especially the small CIs in Experiment 1). In Experiment 2, large CIs demonstrated that it is difficult to make conclusions at the population level based on this small sample (i.e. ‘statistical indeterminacy’; Quertemont, 2011), but the results point in the same direction as those of Experiment 1. Moreover, to further counter the power issue in Experiment 2, we also computed the correlations between the linear and logarithmic fit and the math achievement scores of the children (measured with a standardised math test; Achievement test for mathematics from the Flemish Student Monitoring System, Dudal, 2000). In line with Sasanguie, Van den Bussche et al. (2012), we observed (marginal) significant correlations between the performance on the number line estimation task and mathematics achievement (for the linear, $r(26) = .37, p = .058$; for the logarithmic fit, $r(26) = .41, p = .035$). This indicates that the sample size in Experiment 2 was large enough to detect significant correlations.

It thus seems that different mechanisms underlie number comparison and number line estimation. In contrast to what previously has been suggested (e.g., Laski & Siegler, 2007; Schneider, Grabner, & Paetsch, 2009), when it is explicitly tested whether the compression of the mental number line is underlying both tasks, no relation is found between both tasks. It should be noted, however, that we, in contrast to Laski and Siegler’s study, chose to use non-symbolic instead of symbolic stimuli. Caution is thus recommended in generalising the results, as other mechanisms may play a role in symbolic and non-symbolic number processing (e.g., Santens, Roggeman, Fias, & Verguts, 2010). However, the main finding of the current study is that different mechanisms underlie two wide-spread cognitive number processing tasks: the number comparison task and the number line estimation task. This finding is consistent with recent studies (e.g., Anobile et al., 2012; Barth & Paladino, 2011; Ebersbach, Frick, Luwel, Onghena, & Verschaffel, 2008; Verguts et al., 2005) which have argued in favour of alternative explanations to explain the results in number line estimation and comparison. As a result, we and others have argued in favour of alternative measures to investigate magnitude representation, such as priming (e.g., Defever, Sasanguie, Gebuis, & Reynvoet, 2011) and same-different judgments (e.g., Cohen Kadosh, Muggleton, Silvanto, & Walsh, 2010; Defever, Sasanguie, Vandewaetere, & Reynvoet, 2012; Van Opstal & Verguts, 2011). Although the fact that number comparison and number line estimation reflect the underlying magnitude representation may be highly questionable, it has been frequently observed that the performance on these tasks is related to mathematics achievement (e.g., De Smedt, Verschaffel, & Ghesquière, 2009; Sasanguie, De Smedt et al., 2012; Sasanguie, Van den Bussche et al., 2012; Sasanguie, Göbel et al., 2013). According to us, this may be explained by the fact that not so much the characteristics of the magnitude representation is important for mathematical

skills, but rather decisions and/or task specific mechanisms that act on these magnitude representations. The idea that the magnitude representation is not related to math achievement is supported by a recent study by Defever and colleagues (2012). In this study, a same-different paradigm was used (i.e. a task in which subjects have to decide whether two magnitudes are numerically the same or different and which is considered as a more appropriate task to investigate the mental representations of magnitudes – Van Opstal and Verguts, 2011) and results showed no relationship between the DE and math achievement. On the other hand, Holloway and Ansari (2008) demonstrated that the DE in numerical and non-numerical comparisons is similar, supporting the idea of the DE in comparison reflecting a decisional mechanism. Also neuroimaging studies showed that comparison of number and letter comparison activate very similar parieto-frontal networks in humans, nourishing the idea of common decision mechanisms (Fias, Lammertyn, Caessens, & Orban, 2007). A crucial test would be to investigate the relationship between non-numerical decision processes and mathematics achievement and dyscalculia. Evidently, general cognitive factors like intelligence, working memory and cognitive control have to be taken into account in these studies, as these factors may influence decisional aspects and strategies.

References

- Anobile, G., Cicchini, G.M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*, 454-459.
- Barth, H.C., & Paladino, A.M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*, 125-135.
- Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*, 545-551.
- Bollen, K.A., & Jackman, R. (1990). *Regression diagnostics: An expository treatment of outliers and influential cases*. In J. Fox & J. Scott Long (Eds.), *Modern methods of data analysis* (pp. 257–291). Newbury Park, CA: Sage.
- Booth, J.L., & Siegler, R.S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41*, 189-201.
- Booth, J.L., & Siegler, R.S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*, 1016-1031.
- Bruyer, R., & M. Brysbaert (2011). Combining speed and accuracy in cognitive psychology: is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, *51*, 5-13.
- Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition*, *118*, 35-47.

- Cohen Kadosh, R., Muggleton, N., Silvano, J., & Walsh, V. (2010). Double dissociation of format-dependent and number-specific neurons in human parietal cortex. *Cerebral Cortex*, *20*, 2166-2171.
- Defever, E., Sasanguie, D., Gebuis, T., & Reynvoet, B. (2011). Children's representation of symbolic and non-symbolic magnitude examined with the priming paradigm. *Journal of Experimental Child Psychology*, *109*, 174-186.
- Defever, E., Sasanguie, D., Vandewaetere, M., & Reynvoet, B. (2012). What can the same-different task tell us about the development of magnitude representations? *Acta Psychologica*, *140*, 35-42.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- Dehaene, S., Izard, V., & Piazza, M. (2005). Control over non-numerical parameters in numerosity experiments. *Unpublished manuscript (available on www.uni-cog.org)*.
- De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology*, *103*, 469-479.
- Dudal, P. (2000). *Leerlingvolgsysteem. Rekenen: Toetsen 1-2-3. Basisboek*. Leuven, Belgium: Garant.
- Ebersbach, M., Frick, A., Luwel, K., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology*, *99*, 1-17.
- Fias, W., Lammertyn, J., Caessens, B., & Orban, G.A. (2007). Processing of abstract ordinal knowledge in the horizontal segment of the intraparietal sulcus. *Journal of Neuroscience*, *27*, 8952-8956.
- Gallistel, C.R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43-74.
- Gebuis, T. & Reynvoet, B. (2011). Generating non-symbolic number stimuli. *Behavior Research Methods*, *43*, 981-986.
- Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *Quarterly Journal of Experimental Psychology*, *64*, 2099-2109.
- Holloway, I.D., & Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Developmental Science*, *11*, 644-649.
- Holloway, I.D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*, 17-29.
- Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin Review*, *18*, 1222-1229.
- Iuculano, T., Tang, J., Hall, C.W.B., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science*, *11*, 669-680.

- Laski, E.V., & Siegler, R.S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical-magnitude comparison. *Child Development, 78*, 1723-1743.
- Libertus, M.E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*, 1292-1300.
- Maloney, E., Risko, E.F., Preston, F., Ansari, D., & Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica, 134*, 154-161.
- Moeller, K., Klein, E., Nuerk, H.-C., & Cohen-Kadosh, R. (2012). A unitary or multiple representations of numerical magnitude? – The case of structure in symbolic and non-symbolic quantities. *Frontiers in psychology, 3*, 191, doi: 10.3389/fpsyg.2012.00191
- Moyer, R.S., & Landauer, T.K. (1967). Time required for judgments of numerical inequality. *Nature, 215*, 1519-1520.
- Nuerk, H.-C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting tens and units back into different bins. *Cognition, 82*, 25-33.
- Petitto, A.L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction, 7*, 55-78.
- Santens, S., Roggeman, C., Fias, W., & Verguts, T. (2010). Number processing pathways in human parietal cortex. *Cerebral Cortex, 20*, 77-88, doi: 10.1093/cercor/bhp080
- Sasanguie, D., Defever, E., Van den Bussche, E., & Reynvoet, B. (2011). The reliability of and the relation between non-symbolic numerical distance effects in comparison, same-different judgments and priming. *Acta Psychologica, 136*, 73-80.
- Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology, 30*, 344-357.
- Sasanguie, D., Van den Bussche, E., & Reynvoet, B. (2012). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind, Brain and Education, 6*, 119-128.
- Sasanguie, D., Göbel, S., Moll, K., Smets, K., & Reynvoet, B. (2013). Acuity of the approximate number sense, symbolic number comparison or mapping numbers onto space: what underlies mathematics achievement? *Journal of Experimental Child Psychology, 114*, 418-431.
- Schneider, M., Grabner, R.H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in Grades 5 and 6. *Journal of Educational Psychology, 101*, 359-372.
- Schwarz, W., & Stein, F. (1998). On the temporal dynamics of comparison processes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 1275-1293.
- Siegler, R.S., & Booth, J.L. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428-444.
- Siegler, R.S. & Opfer, J.E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237-243.

- Siegrist, M. (1995). Reliability of the Stroop test with single-stimulus presentation. *Perceptual and Motor Skills, 81*, 1295-1298.
- Strauss, G.P., Allen, D.N., Jorgensen, M.L., & Cramer, S.L. (2005). Test-retest reliability of standard and emotional Stroop Tasks. An investigation of color-word and picture-word versions. *Assessment, 12*, 330-337.
- Quertemont, E. (2011). How to statistically show the absence of an effect. *Psychologica Belgica, 51*, 109-127.
- Vanbinst, K., Ghesquière, P., & De Smedt, B. (2012). Numerical magnitude representations and individual differences in children's arithmetic strategy use. *Mind, Brain and Education, 6*, 129-136.
- Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: priming and comparison distance effects in numerical and non-numerical orders. *Psychonomic Bulletin & Review, 15*, 419-425.
- Van Opstal, F., & Verguts, T. (2011). The origins of the numerical distance effect: The same-different task. *Journal of Cognitive Psychology, 23*, 112-120.
- Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review, 12*, 66-80.
- White, S., & Szűcs, D. (2012). Representational change and strategy use in children's number line estimation during the first years of primary school. *Behavioral and Brain Functions, 8*:1.

Received December, 24, 2012
Revision received April, 3, 2013
Accepted April, 4, 2013