Running head: SYMBOLIC VERSUS NON-SYMBOLIC MAGNITUDE ESTIMATIONS

Symbolic versus non-symbolic magnitude estimations in children and adults

Mirjam Ebersbach, Petra Erz

Abstract

The ability of children and adults to produce symbolic and non-symbolic magnitude estimations was examined and related to children's familiarity with numbers. Six-year-old kindergartners, 7-year-old first graders, 9-year-old third graders and adults made symbolic estimations either by saying number words that matched to numbers of dots (i.e., perception task) or by generating numbers of dots that matched to given number words (i.e., production task). In the non-symbolic estimation task, participants generated the corresponding numbers of dots they had seen previously (i.e., reproduction task). In line with the bi-directional mapping hypothesis (Castronovo & Seron, 2007), children and adults made underestimations in the perception task, overestimations in the production task and intermediate estimations in the reproduction task. However, the performance of kindergartners and first graders showed significant deviations from the predictions of the bi-directional mapping hypothesis. They performed significantly poorer in the production task than in the perception task, implying that these tasks are not mirrored processes in young children. In addition, they made systematic overestimations in the non-symbolic reproduction task, which suggests that a biased mapping occurs here as well. The results are discussed in view of children's familiarity with numbers and potential estimation strategies.

Symbolic versus non-symbolic magnitude estimations in children and adults

Estimating is an important skill that is required in many everyday activities as well as in mathematical settings. Two types of estimations can be distinguished: (1) Numerical (or symbolic) estimations that involve the mapping between number symbols and non-symbolic magnitudes (e.g., estimating the number of elements in a set by saying a numeral) and (2) non-numerical (or non-symbolic) estimations that involve the mapping between two non-symbolic magnitudes (e.g., estimating a number of dots by a number of key presses; Siegler & Booth, 2005).

Previous research concerning symbolic estimations of adults demonstrated systematic biases depending on the direction of the estimations. If numerals had to be assigned to given non-symbolic magnitudes (i.e., *perception* tasks), the actual magnitudes were often underestimated (e.g., Castronovo & Seron, 2007; Izard & Dehaene, 2008; Krueger, 1972, 1982). In contrast, overestimations occurred if non-symbolic magnitudes had to be generated that matched given numerals (i.e., *production* tasks; e.g., Cordes, Gelman, Gallistel, & Whalen, 2001; Krueger, 1984; Whalen, Gallistel, & Gelman, 1999). The bi-directional mapping hypothesis (Castronovo & Seron, 2007; Crollen, Castronovo, & Seron, 2011) has been put forward to explain these systematic biases. It proposes that the accuracy of the underlying mental representations involved in symbolic estimations differs (cf. Izard & Dehaene, 2008; Piazza, Pinel, Le Bihan, & Dehaene, 2007). The mental representation of symbolic magnitudes is supposed to be relatively precise at least in adults, that is, increasing linearly with larger magnitudes. The mental representation of non-symbolic magnitudes, in contrast, is supposed to be distorted, that is, either logarithmically compressed with decreasing distances between larger magnitudes (Brysbaert, 1995; Dehaene, 1992, 1997; Dehaene & Mehler, 1992) or linearly shaped with a proportionally increasing variability for larger magnitudes (i.e., scalar variability; Cordes et al., 2001; Gallistel & Gelman, 1992,

2000; Whalen et al., 1999). The bi-directional mapping hypothesis states that the transcoding between the relatively accurate symbolic representation and the distorted non-symbolic representation results in the typical estimation biases, as larger non-symbolic magnitudes on the distorted mental number line correspond to smaller symbolic numbers. As a result, the actual magnitudes are overestimated in *production* tasks but underestimated in *perception* tasks (for an illustration, see Crollen et al., 2011, p. 41).

Reproduction tasks, in contrast, that involve non-symbolic estimations are assumed to be more accurate as they do not require access to the symbolic representation. Instead, reproduction tasks can be solved by perceiving and immediately reproducing a non-symbolic magnitude. Alternatively, the non-symbolic magnitude might be first internally transcoded into a symbolic number (i.e., perception), which is then transcoded back in the reproduction phase into a non-symbolic magnitude (i.e., production). However, if perception and production were symmetrically mirrored processes, as implicitly assumed by the bi-directional mapping hypothesis, the over- and underestimations would cancel each other out (Crollen et al., 2011). Thus, estimations in the reproduction task should be in any case more accurate compared to the perception and production task.

Crollen et al. (2011) tested the bidirectional mapping hypothesis (Castronovo & Seron, 2007) with adults using three computerized tasks. In the *perception* task, arrays of 21 to 98 dots were presented for 250 ms and the participants had to estimate their numerosity turning a potentiometer that generated numerals in an ascending order. In the *production* task, numerals were presented visually and the task was to generate the matching numerosity of dots by turning a potentiometer that launched the emergence of dots on the screen, as long as the accordant number was reached. In the *reproduction* task, arrays of dots were presented for 250 ms and participants had to reproduce their number in the same way as in the production task. Counting was prohibited and prevented by the short presentation of the dots in the

4

perception and reproduction task and the rapid emergence of the dots in the production task (1 dot per 1.4° of angular rotation). In line with the bi-directional mapping hypothesis, the perception task yielded underestimations and the production task overestimations, while the estimations in the reproduction task were relatively accurate (cf. Crollen & Seron, 2012).

The question arises whether the bi-directional mapping hypothesis is also suited to explain symbolic and non-symbolic estimations of children, whose mental representation of number is still developing (e.g., Booth & Siegler, 2006). Mejias and Schiltz (2013) addressed some aspects of bi-directional mapping in their study with children. Preschoolers estimated four sets of dots (i.e., 8, 16, 34, 64) either symbolically by producing an Arabic number (i.e., perception task) or non-symbolically by reproducing the quantity (i.e., reproduction task). The estimations increased in both tasks for larger numerosities and exhibited the signature of scalar variability, that is, the variability of the estimations increased proportionally with the mean estimations. However, contrary to the bi-directional mapping hypothesis, the preschoolers overestimated the actual numerosities in both tasks, which led the authors to conclude that there is a development from over- to underestimations in children. They explained the overestimations by a noisier representation of large symbolic numbers –

although the overestimations emerged for the two smaller targets in their study (i.e., 8, 16), too. An alternative explanation of the overestimations could be that the children performed four training trials with numerosities that were on average larger than in the test trials (i.e., 15, 15, 50, 75). Moreover, the children received feedback on their accuracy in these test trials. They might have used it to calibrate their estimations towards larger numbers. Izard and Dehaene (2008) showed that feedback on the accuracy of symbolic estimations can lead to more accurate but also to systematically biased calibrations. In addition, no production task was conducted and children's symbolic number knowledge was assessed only in the range from 1 to 12.

The present study aimed at investigating all aspects of the bi-directional mapping hypothesis with children before and after entering school to gain a more complete picture of how the mapping between symbolic and non-symbolic magnitudes and also between two non-symbolic magnitudes develops. We tested kindergartners, first and third graders, and adults to examine whether the typical effects (i.e., underestimation in the perception task, overestimations in the production task, relatively accurate estimation in the reproduction task) emerge in children in a similar way as in adults. The numerosities to be estimated were presented too short to be counted. To get further insights, not only the accuracy in terms of the mean error rate (ER) but also the mean absolute error rates (AER) and the shape and variability of the estimations were inspected. Furthermore, children's familiarity with numbers was assessed and related to their performance in the symbolic and non-symbolic estimation tasks. The aspect of children's familiarity with numbers has often been neglected so far in studies concerned with children's estimation skills (e.g., Barth & Paladino, 2011; Siegler & Opfer, 2003; but see Lipton & Spelke, 2005).

We expected to find a task effect in children, too, similar as in adults. However, due to children's limited familiarity with larger numbers, kindergartners and first graders should perform poorer compared to the older age groups in the symbolic estimation tasks (i.e., production and perception) in particular, by discriminating less sharply between target magnitudes and thus producing larger errors (cf. Barth, Starr, & Sullivan, 2007). It is also probable that they perform poorer in the non-symbolic reproduction task, too, as the accuracy of the underlying mental representation has been shown to increase with age (Halberda & Feigenson, 2008).

It was furthermore assumed that the estimations in all tasks and age groups would exhibit the signature of scalar variability (Whalen et al., 1999), which would reflect the characteristics of the underlying approximate number system (e.g., Cantlon, Platt, & Brannon,

6

2009; Cordes et al., 2001; Deheane, 1997; Feigenson, Dehaene, & Spelke, 2004; Huntley-Fenner, 2001). This signature is expected to emerge in particular in such tasks where the magnitude of the estimations is virtually unlimited, such as in free estimations, compared to other tasks that provide an upper anchor, such as in number line estimation tasks (e.g., Siegler & Opfer, 2003; cf. Cohen & Blanc-Goldhammer, 2011). Finally, the estimation performance in the symbolic and non-symbolic tasks should be related – perhaps stronger between the two symbolic tasks (i.e., perception and production) than between the symbolic and non-symbolic tasks as both perception and production are based on symbolic number knowledge. Accordingly, we also expected that children's familiarity with numbers was closer related to the estimation performance in the perception and production tasks than in the reproduction task.

Method

Participants

Participants were 22 kindergartners (13 girls, 9 boys; mean age: $M = 6$ years, 2 months, $SD = 7$ months), 22 first graders (10 girls, 12 boys; mean age: $M = 7$ years, 6 month, $SD = 7$ months), 21 third graders (10 girls, 11 boys; mean age: $M = 9$ years, 5 months, $SD = 4$ months), and 21 adults (14 women, 7 men; mean age: $M = 24$ years, 3 months, $SD = 32$ months). The sample was predominantly Caucasian and lived in a medium-sized city in Germany.

The rationale for choosing these age groups was to cover subsamples with a relatively broad range of numerical skills, starting with kindergartners who have not yet received formal instruction and possess only a limited and quite variable knowledge of numbers. While there are some kindergartners, who can already relatively robustly produce numbers to 100, others are hardly able to count to 10 (cf. Ebersbach, in press). First graders in Germany learn to calculate with numbers to 20 and become acquainted with the decades to 100. Third graders

were expected to have a relatively consolidated number knowledge in the number range to 100 (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008). In addition, the mapping between symbolic and non-symbolic magnitudes develops in particular in this age range (Mundy & Gilmore, 2009). Adults served as basis for normative comparisons and for replicating the results of Crollen et al. (2011). The children were recruited at public kindergartens and schools in middle-class neighbourhoods and were not selected with regard to their mathematical or cognitive skills. They took part voluntarily with informed consent of their parents. The adults were students from the local university. All participants spoke German fluently.

Stimuli and Procedure

Twelve target magnitudes had to be estimated (i.e., 15, 18, 22, 30, 37, 46, 54, 63, 70, 78, 85, and 94). In the reproduction and perception task, the target magnitudes were non-canonical arrays of green dots on a white background presented too shortly to be counted. The location of the single dots was randomized in each trial. We did not control for the total area of the dots (each with a diameter of 5 mm) or other continuous variable to have a more valid and realistic indicator of people's estimation skills as they are required in everyday life (e.g., estimating the number of apples or humans). Usually, numerosity is correlated with total area or size and, accordingly, people use this visual information for numerosity judgments (e.g., Gebuis & Reynvoet, 2012). However, we are aware that the results apply only to such stimuli and should not be generalized to other non-symbolic stimuli controlled for continuous variables, for instance (cf., Ebersbach, Luwel, & Verschaffel, 2013). The estimations in the perception task were made by saying the corresponding number word. In the production task, the target magnitudes were presented verbally as number words, too. A verbal format rather than Arabic numbers were used as at least kindergartners and first graders were not familiar with reading and writing numerals up to 100.

The participants were tested individually in a quiet room, sitting in a distance of about 50 cm from the computer screen (15"). The instruction was to make estimations without counting, which was accomplished by presenting the numerosities too short to be counted. The non-symbolic reproduction task was completed by all participants first. Thereafter, half of the participants of each age group proceeded with the symbolic perception task and the other half with the symbolic production task. The fixed order of the non-symbolic and symbolic tasks was chosen to avoid that the participants activated numerals in the symbolic tasks that they used later as a basis for their non-symbolic estimations. Each task consisted of two blocks and within each block the twelve target magnitudes were presented in random order.

Trials in all tasks started with a fixation cross (1000 ms), followed by a blank screen (500 ms). In the reproduction and perception task, the target magnitude (i.e., dots) followed, masked by an array of irregular grey lines to avoid visual after effects. The presentation time for targets and masks were 250 ms for adults (cf. Crollen et al., 2011). As pilot testing suggested that these durations were too short for children, presentation times were prolonged in these age groups to 1000 ms (cf. Mejias & Schiltz, 2013). Subsequent to the mask, a blank screen followed (500 ms) and then either a question mark appeared as go-signal to indicate that the verbal response was requested, which was recorded by the experimenter (i.e., perception task) or a single dot was presented as go-signal to start with the reproduction. The reproduction was accomplished by means of a potentiometer. Turning the knob of the potentiometer slightly to the right (i.e., 4.9°) elicited one to three dots at once on the screen (this number was varied randomly to suppress counting strategies), while turning it to the left deleted one dot. The estimation was finished when the participant pressed the knob. The location where the dots appeared on the screen was randomized. Before the reproduction task

started, two training trials were conducted without feedback showing arrays with 3 and 110 dots to familiarize the participants with the potentiometer.

In the production task, the target magnitude was expressed as number word after the fixation cross by the experimenter, followed by a blank screen (500 ms). Thereafter, a single dot appeared and the procedure was the same as in the reproduction task. No more than 192 dots could be generated per trial. The participants were not informed about the maximum number of dots and received no feedback about the accuracy of their estimations – neither in the training trials nor in the main test. Only encouraging comments were made during the test, irrespectively of the accuracy.

The children additionally completed two tests to assess their familiarity with numbers in the tested range of 100. They were first asked to count as far as they could (i.e., counting range). They were stopped if they reached 100. Otherwise, the largest number they could count to was recorded. Second, a test focusing on the decade changes was administered to tap children's robust and flexible number knowledge that goes beyond the simple, often practiced repetition of successive numbers starting with "one" (Ebersbach et al., 2008; Lipton & Spelke, 2005). The experimenter started to count (e.g., 16, 17, 18…) and asked the child to count on. If a child completed the decade change correctly, he or she was interrupted and the experimenter began to count numbers preceding the following decade change (e.g., 26, 27…). If the child failed to complete the decade change, the same numbers were repeated. If he or she failed again, the highest decade change that was successfully completed was recorded (e.g., 20). If a child reached 100, the test was discontinued.

<div align="center">Results</div>

Preliminary analyses

Missing data due to inattention or incorrect handling of the potentiometer occurred in 2.4% of the trials of kindergartners, 2.8% of first graders, 1.5% of third graders, and 0.2% of

adults. In addition, the data of one kindergartner and one first grader in the reproduction task were excluded due to an insufficient understanding of the task and the data of one first grader in the reproduction task and one in the perception task were missing due to technical problems. To enhance the comparability between the tasks, verbal estimations in the perception task that were larger than 192 were excluded (i.e., 2% of the trials of kindergartners, 0.2% of first and third graders, respectively, 0% of adults). The mean estimations for each task and age group, computed from the remaining data, are shown in Figure 1.

Insert Figure 1 about here

The averaged decade change of kindergartners as a group was $M = 33.2$ ($SD = 30.0$) and they counted on average to $M = 41.0$ ($SD = 28.3$), while first graders completed decade changes on average up to $M = 74.6$ ($SD = 36.3$) and counted to $M = 77.0$ ($SD = 30.2$). Third graders completed all decade changes and were able to count to 100; the same was supposed for adults.

Assessing the signatures of the approximate number system

It was tested first whether the estimations exhibited the signatures of the approximate number system, that is, that mean estimations (Ms) and mean standard deviations (SDs) increased with target magnitudes and that the coefficient of variation ($CV = SD/M$) was constant across target magnitudes (i.e., scalar variability). Scalar variability was expected as the estimations virtually had no upper anchor. To account for the relatively large number of target magnitudes, linear regressions instead of ANOVAs were computed.

The log of the Ms and SDs increased with target magnitudes in all age groups and tasks except for kindergartners in the production task, where the increase of the log Ms was

only marginally significant (i.e., $p = .077$) and the log SDs remained constant (Table 1). Scalar variability could be assumed in all age groups and tasks except of in adults in the production task, where the CV increased slightly with larger target magnitudes ($p = .049$). Such a tendency was also shown for third graders in the reproduction task ($p = .077$). Moreover, the CVs varied as a function of age, $F(3, 937) = 99.51, p < .001,$ $_{part}\eta^2 = .24$, and task, $F(1.98, 1858.90) = 3.85, p = .022,$ $_{part}\eta^2 = .004$[1]. They were largest in kindergartners ($M_{CV} = .36, SD_{CV} = .12$) and significantly smaller in each of the following age groups, first graders ($M_{CV} = .31, SD_{CV} = .13$), third graders, ($M_{CV} = .26, SD_{CV} = .13$), and adults, ($M_{CV} = .17, SD_{CV} = .13$), all $ps < .001$[2]. In addition, CVs were larger in the reproduction task ($M_{CV} = .29, SD_{CV} = .22$) than in the perception task ($M_{CV} = .26, SD_{CV} = .26$), $p = .024$, that did not deviate from the CVs in the production task ($M_{CV} = .27, SD_{CV} = .22$), $p = .80$. No other effects were significant.

Insert Table 1 about here

Estimation accuracy as a function of task and age group

Two measures of the estimation accuracy were calculated: error rates, $ERs =$ (estimation – target magnitude) / target magnitude) and absolute error rates, $AERs =$ |(estimation – target magnitude) / target magnitude)| (see Table 2). While the ERs served to indicate over- and underestimations but are susceptible towards the fact that over- and underestimations might cancel each other, the $AERs$ refer to the general accuracy of the estimations irrespectively of over- and underestimations.

---

[1] Greenhouse-Geisser is reported if sphericity could not be assumed.

[2] Bonferroni was used to correct for multiple comparisons of within-subjects variables and Tukey-HSD was applied for multiple comparisons of between-subjects variables.

Preliminary analyses examined whether the mean ERs differed significantly from 0. This was confirmed for each age group and task, $ps < .026$, except of for third graders and adults in the reproduction task, $ps > .11$. The mean $AERs$ differed significantly from 0 in all age groups and tasks, $ps < .001$.

It was then tested whether each of the two accuracy scores differed as a function of age group and task. For the mean ERs, there were main effects of task, $F(1.38, 107.72) = 89.58$, $_{part}\eta^2 = .54$, and age group, $F(3, 78) = 6.55$, $_{part}\eta^2 = .20$, and an interaction of task and age group, $F(4.14, 107.72) = 1.71$, $p < .001$, $_{part}\eta^2 = .18$, $ps \leq .001$. The task effect was confirmed when computing the analyses separately for each age group (Table 2). Multiple comparisons revealed significant differences between all three tasks in each age group ($ps < .005$) with smallest, negative mean ERs – and thus underestimations – in the perception task, largest, positive mean ERs – and thus overestimations – in the production task, and intermediate values in the reproduction task. Only the mean ERs of third graders in the reproduction and perception task did not differ significantly, $p = .24$. To specify the age differences concerning the mean ERs for each single task, a MANOVA was computed. It confirmed a main effect of age group, $F(9, 185.12) = 3.28$, $p = .001$, $_{part}\eta^2 = .11$, that was present in each task (reproduction: $F(3, 78) = 8.28$, $p < .001$, $_{part}\eta^2 = .24$; perception: $F(3, 78) = 3.99$, $p = .011$, $_{part}\eta^2 = .13$; production: $F(3, 78) = 5.84$, $p = .001$, $_{part}\eta^2 = .18$). Pairwise comparisons suggested for the reproduction task similar mean ERs in kindergartners and first graders as well as in third graders and adults, $p = 1.00$, whereas the mean ERs of the two younger age groups were significantly larger than those of the two older age groups, $ps < .05$. The only significant difference in the perception task was between kindergartners and third graders with larger mean ERs in the first group, $p = .012$. In the production task, kindergartners had significantly larger mean ERs than third graders, $p = .009$, and adults, $p = .003$.

The analysis of the mean *AER*s yielded also main effects of task, $F(1.31, 101.99) = 36.31$, $_{part}\eta^2 = .32$, age group, $F(3, 78) = 23.33$, $_{part}\eta^2 = .47$, and an interaction of the two variables, $F(3.95, 101.99) = 5.27$, $p = .001$, $_{part}\eta^2 = .17$, $ps \leq .001$. Separate analyses confirmed a significant main effect of task in each age group (Table 2). Multiple comparisons suggested larger mean *AER*s in the production task compared to the reproduction task in all age groups, $ps < .024$. In addition, the production task yielded larger mean *AER*s than the perception task in kindergartners and first graders, $ps < .005$, but not in third graders and adults, $ps > .23$. Finally, the mean *AER*s of the reproduction task were only in adults smaller than in the perception task, $p = .001$, but not in the other age groups, $ps > .10$.

To specify the differences between age groups, a MANOVA was conducted, with mean *AER*s of each task as dependent measure. It confirmed the main effect of age group, $F(9, 185.12) = 11.69$, $p < .001$, $_{part}\eta^2 = .31$, that emerged in each task (reproduction: $F(3, 78) = 18.06$, $_{part}\eta^2 = .41$; perception: $F(3, 78) = 37.63$, $_{part}\eta^2 = .59$; production: $F(3, 78) = 12.76$, $_{part}\eta^2 = .33$; all $ps < .001$). Pairwise comparisons specified that the mean *AER*s of third graders and adults did not differ in any of the tasks, $ps > .57$. The mean *AER*s of kindergartners and first graders did neither differ in the reproduction task nor in the production task, $ps > .38$. The remaining differences were significant with larger mean *AER*s in younger age groups, all $ps < .05$.

Insert Table 2 about here

Estimation accuracy as a function of task and familiarity with numbers

The task effect on the two accuracy measures was also examined separately for the familiar and unfamiliar number range of kindergartners and first graders. This was done as Figure 1 suggested relatively large estimation biases in the familiar number range in the two

younger age groups. We thus solely considered the mean *ER*s of kindergartners and first graders who were not familiar with numbers to 100. The familiar number range included only those trials where the target magnitude was within each child's range of successful decade changes (i.e., 24% of the trials in this subsample), while the unfamiliar number range included the remaining trials.

An ANOVA with repeated measures confirmed a main effect of task in the familiar number range, $F(1.15, 21.87) = 28.78$, $p < .001$, $_{part}\eta^2 = .60$. Multiple comparisons revealed significant differences between all three tasks, $ps \leq .001$. However, the pattern deviated from the previous findings as the familiar target magnitudes in the reproduction task were significantly overestimated, $t(19) = 5.07$, $p < .01$, whereas the mean *ER*s in the perception task did not differ significantly from 0, $p = .18$. In addition, the typical overestimation in the production task was revealed in the familiar range, too, $t(20) = 5.35$, $p < .01$ (see Figure 2).

A task effect emerged also in the unfamiliar number range, $F(1.42, 35.51) = 13.93$, $p < .001$, $_{part}\eta^2 = .36$, again with significant differences between all tasks, $ps < .015$. However, the pattern here was in line with the bi-directional mapping hypothesis (Castronovo & Seron, 2007), with significant overestimations in the production task, $t(26) = 3.38$, $p = .012$, marginally significant underestimations in the perception task, $t(26) = -2.62$, $p = 0.08$, and mean *ER*s that did not differ significantly from 0 in the reproduction task, $p = .60$.

Insert Figure 2 about here

Mathematical relationship between estimations and target magnitudes

To examine the relationship between estimations and target magnitudes, a power model was fitted to the data (cf. Crollen et al., 2011). We abstained from fitting other mathematical models (e.g., logarithmic, 2-phase, linear models) to the estimations as the

question, which model would yield the best fit depends largely on the estimation paradigm (e.g., bounded vs. unbounded estimation tasks; midpoint provided or not; cf. Ebersbach et al., 2013). Thus, studies are often not comparable. Moreover, the differences between the best fitting models are often small and have only a minor explanatory value with regards to the underlying mechanisms of the estimation processes (cf. White & Szucs, 2012).

A power model is suited to describe linear (i.e., exponent of 1) as well as non-linear relationships (i.e., exponent smaller or larger than 1) between two variables. To avoid artifacts by averaging across the whole sample (cf. Bouwmeester & Verkoeijen, 2012), a power model was fitted to the estimations of each participant. The alpha level was raised to 10% to account for the small number of data per individual and to maintain the power (cf., Ebersbach, 2009; Schlottmann, 2001). Table 3 shows the percentage of participants for whom the power model yielded a significant fit. We also report the mean fit ($_{corr}R^2$), the mean constant, and the mean exponent ß as indicator of the linearity / non-linearity of the estimations – only of those participants for whom the model fit was significant.

A power model explained the data well in adults and older children but poorer in younger children. For kindergartners in the production task, in particular, the model yielded only in about 27% of the children a significant fit. The mean exponent was significantly smaller than 1 in all groups reflecting that the estimations increased slower and thus non-proportionally with the target magnitudes (see also Figure 1). One exception was the exponent of third graders in the production task that did not differ from 1 indicating that their estimations were linearly related to the target magnitudes. Another exception was the mean exponent of adults in the production that was significantly larger than 1 task implying that their estimations increased non-proportionally faster than the target magnitudes.

Insert Table 3 about here

A repeated measures ANOVA with the mean exponent as dependent measure, excluding kindergartners due to the small number of significant model fits, yielded effects of task, $F(2, 102) = 33.66$, $_{part}\eta^2 = .40$, age group, $F(2, 51) = 26.77$, $_{part}\eta^2 = .51$, and an interaction of task and age group, $F(4, 102) = 8.59$, $_{part}\eta^2 = .25$; all $ps < .001$. Analyses conducted separately for each age group revealed no significant differences between the exponents in the three tasks in first graders, $F(2, 24) = 1.22$, $p = .31$, but an effect of task in third graders, $F(2, 38) = 15.22$, $p < .001$, $_{part}\eta^2 = .45$, with an exponents approaching 1 in the production task compared to smaller exponents in both the perception and reproduction task, $ps < .01$. The effect of task was significant in adults, too, $F(1.40, 28.05) = 67.15$, $p < .001$, $_{part}\eta^2 = .77$, with a mean exponent even larger than 1 in the production task, an exponent smaller than 1 in the reproduction task and smallest exponents in the perception task, all $ps < .001$.

To unravel differences between age groups, a MANOVA was conducted with the mean exponent as dependent variable, again without the sample of kindergartners. It revealed a main effect of age group, $F(6, 112) = 19.27$, $_{part}\eta^2 = .51$, $p < .001$, that emerged in each task (reproduction: $F(2, 58) = 31.42$, $_{part}\eta^2 = .52$; perception: $F(2, 58) = 20.03$, $_{part}\eta^2 = .41$; production: $F(2, 58) = 27.64$, $_{part}\eta^2 = .49$, all $ps < .001$). Repeated contrasts indicated smaller exponents in first graders compared to third graders. who had smaller exponents than adults in all tasks, $ps < .01$. An exception were the exponents of third graders and adults in the production task that were similar, $p = .30$.

Relationship between symbolic and non-symbolic estimations and the impact of children's familiarity with numbers

Bivariate correlations were computed to examine whether the mean *ER*s in the symbolic (i.e., perception and production) and non-symbolic tasks (i.e., reproduction) were related. Furthermore, the counting range and the highest decade change of kindergartners and first graders were considered. As shown in Table 4, the mean *ER*s in the two symbolic tasks correlated significantly in all age groups. Moreover, the mean *ER*s in the non-symbolic reproduction task correlated with the mean *ER*s in the symbolic production task in kindergartners and first graders and with the mean *ER*s in the symbolic perception task in first and third graders and adults. Finally, the counting span and the decade change were marginally or significantly correlated with the mean *ER*s in the reproduction and production task in kindergartners but not in first graders.

Insert Table 4 about here

Discussion

The present study aimed at testing the bi-directional mapping hypothesis (Castronovo & Seron, 2007; Crollen et al., 2011) from a developmental perspective. Furthermore, it was investigated whether the performance in symbolic and non-symbolic estimations was related and whether it was associated in children with their familiarity with numbers.

Estimation accuracy as a function of task and age group

In line with previous (Crollen et al., 2011) and actual findings for adults (this study), children, too, provided systematically biased estimations, that is, underestimations in the perception task, overestimations in the production task, and intermediate estimations in the reproduction task (see Figure 1). This pattern became most apparent in third graders and indicates that the bi-directional mapping hypothesis applies also to children who are familiar with the tested number range.

It was furthermore revealed that kindergartners and first graders – as well as third graders and adults – significantly underestimated the target magnitudes in the perception task. This finding corresponds with the predictions of the bi-directional mapping hypothesis (Castronovo & Seron, 2007) but challenges the assumption of Mejias and Schiltz (2013) of a general development from over- to underestimation in this task. Their finding that preschoolers produced overestimations in the perception task might be explained instead by calibration processes during the training trials that included feedback on the actual magnitudes.

While the general pattern of over- and underestimations emerged also in kindergartners and first graders, there were two main exceptions that are not in line with the bi-directional mapping hypothesis (Castronovo & Seron, 2007).

First, the general estimation accuracy (i.e., the mean $AERs$) was significantly poorer in the symbolic production task than in the symbolic perception task in kindergartners and first graders, though Crollen et al. (2011) assumed that these were mirrored processes. Moreover, the estimations of kindergartners were clearly less systematic in the production task compared to the other tasks, as reflected by the relatively small number of significant model fits. One might argue at this point that kindergartners did not understand the production task or had problems with handling the potentiometer, which might have caused the poor performance. However, the potentiometer was also used in the reproduction task with a significantly higher precision. Thus, the marked overestimations in the production task should be explained by task-specific requirements, such as the comprehension of two-digit number words, which differs from the perception task that required the production of number words.

Wynn (1995) has already stated that children's comprehension of number words lags behind their production of number words (cf. Condry & Spelke, 2008; Sarnecka, Kamenskaya, Yamana, Ogura, & Yudovina, 2007; Sarnecka & Lee, 2009). This applies for

19

tasks involving multi-digit numbers in particular, where children make systematic biases – for instance when transcoding between number words and symbolic numerals (Camos, 2008; Power & Dal Martello, 1990; Seron & Fayol, 1994). Specific problems with multi-digit numbers arise in languages where tens and units are inversed when spoken (e.g., in German: 23 – "dreiundzwanzig" / "three-and-twenty", for a review see Klein et al., 2013). The principle of inversion led for instance to a poorer performance in tasks, where the larger of two numbers had to be identified (Pixner, Moeller, Hermanova, Nuerk, & Kaufmann, 2011) or in number line estimation tasks (Helmreich, Zuber, Pixner, Kaufmann, Nuerk, & Moeller, 2011). The problem of inversion might arise even in tasks where number words and not digits are involved. In one-digit number words, the first (and only) number word (i.e., the unit) refers to its magnitude. In two-digit numbers of inversed languages, the second number word (i.e., the ten) determines the major part of its magnitude. Thus, even as the German speaking children in the present study did not have to translate between written and spoken number words, inversion might have affected their comprehension of two-digit number words as it follows another principle than that of one-digit number words that they had acquired first. To conclude, the asymmetry between the accuracy of the production and perception task in young children might be assigned to their symbolic number knowledge that is not fully developed yet, with number word comprehension lagging behind number word production, which is potentially reinforced by the principle of inversion (cf. Zuber, Pixner, Moeller, & Nuerk, 2009). Thus, number knowledge – and the access to the meaning of number words, in particular – is an important aspect when considering bi-directional mapping from a developmental perspective (cf. Sasanguie, De Smedt, Defever, & Reynvoet, 2012). Therefore, production and perception in symbolic estimations might be no simply mirrored processes, at least for children with limited numerical skills (see also Meijas, Mussolin, Rousselle, Grégoire, & Noël, 2012).

A second deviation from the predictions of the bi-directional mapping hypothesis (Castronovo & Seron, 2007) was the finding that kindergartners and first graders overestimated the target magnitudes not only in the symbolic production task but also in the non-symbolic reproduction task, even though to a significantly lesser extent (Figure 1; for similar results see Mejias & Schiltz, 2013). Interestingly, the overestimations in the reproduction task emerged, too, if only the familiar number range of those kindergartners and first graders was considered who were not familiar with the numbers to 100. Crollen et al. (2011), in contrast, predicted relatively accurate estimations in the reproduction task either (a) as no mapping would take place as only the mental representation of non-symbolic magnitudes is assessed or (b) as the reproduction is executed in two steps. First, the perceptual input is transformed into a symbolic magnitude (i.e., perception), which is then translated back into a non-symbolic magnitude (i.e., production). Thus, the biases in perception and production should level each other out if these are mirrored processes. However, the overestimations in the reproduction task of the younger children contradict this assumption. Potentially, they applied a two-step approach. At this point, it is not necessary to assume that in fact a symbolic magnitude is mentally activated as intermediate step. It is also possible that a non-symbolic magnitude is activated by the visual input, which is then reproduced. In each case, the relatively larger overestimations in the production phase compared to the smaller underestimations in the perception phase, as reported earlier, would yield overestimations in the reproduction task. The fact that this pattern emerged only in the familiar number range of kindergartners and first graders suggests furthermore that they might have used this two-step approach in particular for smaller magnitudes, while they applied a direct, non-symbolic reproduction process for magnitudes that exceeded their familiar number range. Interestingly, even Crollen et al. (2011) reported slight but significant overestimations for smaller magnitudes in the reproduction task in adults and their data also suggested that adults

performed poorer in the production task (mean $ER = .52$) compared to the perception task (mean $ER = -.38$). Thus, the imbalance between the two symbolic estimations tasks might evoke even in adults overestimations in the non-symbolic reproduction task.

An additional process might have contributed to the biased performance in the reproduction task especially of young children, that is, their poorer ability to keep their mental representation of non-symbolic magnitudes activated over time (cf. Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009). This is required in particular in the reproduction task that took longer due to the generation of the magnitudes than in the perception task, where only a number word had to be generated. If this activation fades out earlier, the reproduction might become imprecise, too. Even if this process cannot explain the systematical bias that estimations were too large, it should be investigated further.

In adults, the mean ERs and the mean AERs suggested a higher accuracy in the non-symbolic reproduction task than in the two symbolic tasks, as proposed by the bi-directional mapping hypothesis (Crollen et al., 2011). The fact that Mejias, Grégoire, and Noël (2012) found a similar general accuracy in symbolic and non-symbolic estimations of adults, which contradicts the present findings, might be explained by the differing methodology. As mentioned earlier, Mejias et al. (2012) provided also in this study feedback in the training trials, which might have led to an approximation of the performance in both task types.

Comparing the general accuracy in the different age groups revealed no differences between kindergartners and first graders and between third graders and adults. This supports the assumption that similar levels of the familiarity with numbers lead to similar general estimation accuracies. However, one has to keep in mind that the presentation time of the stimuli was longer for adults than for all groups of children. Nevertheless, third graders and adults might perform similarly even with equal presentation times as their number knowledge in the tested range to 100 is consolidated (cf. Booth & Siegler, 2006).

Shape and variability of the estimations across target magnitudes

The signature of scalar variability was revealed in all tasks and age groups with the following exceptions: The variability of the estimations increased proportionally faster than the target magnitudes in adults in the production task and, as a tendency, also in third graders in the reproduction task. Even if this effect was not large, the question might be raised whether the principle of scalar variability should be extended in terms of allowing the variability of the estimations to increase even faster than the target magnitudes. This would not contradict the general assumptions of the principle but could account for findings as ours. Furthermore, the existence of scalar variability can be questioned in kindergartners' estimations in the production task, where the mean estimations increased only marginally with the target magnitudes and the mean *SD*s remained constant. This pattern reflects children's problems with number word comprehension in this task.

The finding of scalar variability in the other age groups and tasks underlines the assumption that it emerges only in those estimation tasks where no upper anchor is provided, such as in our study (cf. Cohen & Blanc-Goldhammer, 2011). If an upper anchor is presented, such as in number line tasks, it can be used to limit and calibrate the estimations, potentially even leading to proportional judgments (cf. Barth & Paladino, 2011) rather than to non-strategic, intuitive estimations with increasing variability.

A power model explained the relationship between the estimations and the target magnitudes quite well – better in older than in younger participants, whose estimations were less systematic. The exponents of the power model were almost always smaller than 1 suggesting that the estimations increased non-proportionally slower than the target magnitudes, which is a signature of the approximate mental representation of magnitudes. Exceptions were the estimations of third graders in the production task with an exponent of 1 and thus linear estimations and the estimations of adults in the production task with an

exponent even larger than 1. The constant as an indicator of the distortion of the power model was also clearly larger in younger than in older participants (Table 3). It can thus be concluded that symbolic and non-symbolic estimations become more systematic, more linear and less distorted with age, even if no upper or other anchors are provided. This conclusion has to be drawn cautiously, of course, as only a cross-sectional design was used.

Relationship between symbolic and non-symbolic estimations and the impact of children's familiarity with numbers

The performance in the two symbolic tasks was, as expected, negatively correlated in all age groups. The larger the underestimations in the perception task, the larger the overestimations in the production task, which supports assumptions of the bi-directional mapping hypothesis (Castronovo & Seron, 2007). The performances in the symbolic tasks and the non-symbolic reproduction task were associated, too. More specifically, the mean *ER*s in the reproduction task correlated positively with the mean *ER*s in the production task in kindergartners and first graders, suggesting a that production played a specific role also in the non-symbolic reproduction task, which is in line with a two-step approach in the reproduction task. In first graders, there was in addition a negative correlation between mean *ER*s in the reproduction and perception task, which was nevertheless smaller than the correlation between the reproduction and production task. The performance of third graders in the symbolic and non-symbolic tasks did not correlate significantly but in adults there was a significant positive correlation between the mean *ER*s in the reproduction and perception task. This correlation might reflect the impact of the accuracy of the approximate mental number system, as assessed by the reproduction task, in tasks that require a verbal judgment of non-symbolic magnitudes. It remains an open question at this point why this association was not significant in third graders. However, the relationships between the performance in the non-symbolic and symbolic tasks in younger children and even in adults suggest that the accuracy

of the approximate number system plays an important role in symbolic estimations, too. Accordingly, recent research has shown that the accuracy of both the symbolic and non-symbolic magnitude system improves with formal education (Piazza, Pica, Izard, Spelke, & Dehaene, 2013), even if Defever, Sasanguie, Gebuis, and Reynvoet (2011) propose that the overlap of magnitudes on the mental number line is similar for symbolic and non-symbolic magnitudes and stable across age groups ranging from kindergarten to 6th grade, at least.

Kindergartner's familiarity with numbers (i.e., their counting span and their highest decade change) correlated with their performance in both the reproduction task and the production task. The fact that these correlations were not significant in first graders suggest that the development of the symbolic and non-symbolic number system is intertwined in particular in younger children that have not yet received formal instructions on symbolic numbers (see also Mussolin, Nys, Leybaert, & Content, 2012).

Interestingly, kindergartners and first graders who were not familiar with numbers to 100 provided in their familiar number range larger overestimations in both the reproduction and production task than in their unfamiliar number range. This result could be assigned on the one hand to the accuracy measure that refers to the percentage of the deviation between estimation and target. Thus, overestimating a target of 20 by 100%, for instance, has absolutely a smaller impact than overestimating a target of 80 by 100%. On the other hand, this pattern might also suggest that younger children over-represent smaller magnitudes compared to larger magnitudes (see Figure 1; Halberda & Feigenson, 2008) – a pattern that was also reported for number line estimations, and that could be explained by a power model, as in our study, or by a logarithmic model (e.g., Siegler & Opfer, 2003), or even by a 2-phase linear model (Ebersbach et al., 2008). Their performance in the perception task was much less affected by their familiarity with numbers than in the other two tasks, as inferred from lacking correlations between young children's number knowledge and the mean ERs in the perception

25

task as well as from the analyses conducted separately for the familiar and unfamiliar number range. There were also no differences between the three groups of children with regard to their general accuracy in the perception task. These results provide further evidence for the assumption than young children are able to systematically estimate magnitudes even if they exceed their familiar number range (Barth et al., 2009).

Future studies

Future studies research should use a longitudinal design to uncover the antecedents of symbolical and non-symbolical estimation performance in children. In addition, tasks that involve number words, as in our study, should be contrasted with tasks that include Arabic numbers to see whether the effects apply only to number words. This approach could also tap the intensely discussed question of whether there is one universal format-independent mental representation of magnitudes or multiple format-dependent mental representations (see Cohen Kadosh, Lammertyn, & Izard, 2008, for a review). Moreover, the design of the dots as non-symbolic stimuli could be changed by controlling for perceptual variables (cf. Gebuis & Reynvoet, 2012; Sophian & Chu, 2008) to find out whether the overestimations in the reproduction task are a result of increasing surface with number. One might also think of assessing children's number word comprehension and number word production more directly to uncover how both abilities contribute to the estimation accuracy in the different tasks. Finally, the assumption that at least younger children solve reproduction tasks not directly but with an intermediate step that causes their biases in this task and the assumption that younger children are less able to maintain their mental representation of magnitudes activated has to be investigated empirically.

Conclusions

We showed that the bi-directional mapping hypothesis (Castronovo & Seron, 2007; Crollen et al., 2011) is also suited to explain symbolic and non-symbolic estimations of

children who were familiar with the tested number range (i.e., third graders in our study). However, estimations of younger children who were not familiar with the whole number range showed systematical deviations from the model predictions. They performed poorly in the symbolic production task in particular, potentially due to their limited comprehension of two-digit number words that lags behind their production of number words (cf. Wynn, 1995). A limited number word comprehension might also explain why the general accuracy of kindergartners and first graders was smaller in the symbolic production task than in the symbolic perception task although the bi-directional mapping hypothesis assumed that these are mirrored processes at least in adults (Crollen et al., 2011). In addition, younger children produced marked overestimations in the reproduction task – perhaps as they solved this task in two steps, that is, by translating the perceptual input into an approximate mental representation of magnitude and by translating it back into a non-symbolic magnitude. Accordingly, overestimations in the reproduction task would result from the fact that overestimations in the production step were markedly larger than underestimations in the perception step. Moreover, contrary to the bi-directional mapping hypothesis, younger children were not per se better in the non-symbolic than in the symbolic estimation tasks as they performed similarly well in the perception and reproduction task and only the direction of the deviations differed (i.e., over- vs. underestimations). This is astonishing given the limited number word knowledge of the two youngest age groups. We demonstrated furthermore that kindergartners' number knowledge had a significant impact in both the production and reproduction task, while the accuracy in the perception task was virtually independent of their number knowledge that played also no role for the accuracy of first graders.

Taken together, it can be concluded that normally developing children perform better in symbolic estimation tasks that involve the perception of magnitudes rather than their

production. Potentially, their intuitive use of number words in the perception task that leads to less biased estimation could be used to establish a more robust comprehension of number words. too, including an advanced place-value understanding (cf. Nuerk & Willmes, 2005). However, the finding that number word production is important for the performance in the non-symbolic reproduction task underlines the close association between children's number knowledge and their estimation accuracy (cf. Kolkman, Kroesbergen & Leseman, 2013). Our results raise the question of whether children with mathematical learning disabilities would show a poorer number word comprehension than number word production, too, or if both competencies are limited to a similar extent (cf. Mejias et al., 2012). Both effects would support the hypothesis that such children have problems in assessing and processing symbolic numbers (cf. De Smedt & Gilmore, 2011; Rousselle & Noël, 2007) rather than a deficient mental representation of magnitudes (Wilson & Dehaene, 2007). We showed that a deficient number word comprehension in particular might lead to poor symbolic (i.e., production task) and even non-symbolic estimations (i.e., reproduction task). To conclude, young children's mapping between symbolic and non-symbolic magnitudes does not seem to be as straightforward as proposed by the bi-directional mapping hypothesis (Castronovo & Seron, 2007). Therefore, its investigation requires taking into account both children's number knowledge as well as potential estimation strategies.

References

Barrouillet, P., Gavens, N., Vergauwe, F., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology, 45*, 477-490. doi: 10.1037/a0014615

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14*, 125-135. doi: 10.1111/j.1467-7687.2010.00962.x

Barth, H., Starr, A., & Sullivan, J. (2009). Children's mappings of large number words to numerosities. *Cognitive Development, 24*, 248-264. doi:10.1016/j.cogdev.2009.04.001

Booth, J. L., & Siegler, R.S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 41*, 189-201. doi:10.1037/0012-1649.41.6.189

Bouwmeester, S., & Verkoeijen, P. P. J. L. (2012). Multiple representations in number line estimation: A developmental shift or classes of representations? *Cognition and Instruction, 30*, 246–260.

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. doi:10.1037/0096-3445.124.4.434

Camos, V. (2008). Low working memory capacity impedes both efficiency and learning of number transcoding in children. *Journal of Experimental Child Psychology, 99*(1), 37–57. doi:10.1016/j.jecp.2007.06.006

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends Cognitive Sciences, 13*, 83–91.

Castronovo, J., & Seron, X. (2007). Numerical estimation in blind subjects: Evidence of the impact of blindness and its following experience. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1089-1106. doi:10.1037/0096-1523.33.5.1089

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review, 18,* 331–338. doi:10.3758/s13423-011-0059-z

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology, 84,* 132–147. doi:10.1016/j.pneurobio.2007.11.001

Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural numbers. *Journal of Experimental Psychology: General, 137,* 22–38. doi:10.1037/0096-3445.137.1.22

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001).Variability signatures distinguish verbal from non-verbal counting for both large and small numbers. *Psychonomic Bulletin and Review, 8,* 698–707.

Crollen, J., Castronovo, J., & Seron, X. (2011). Under- and over-estimation: A bi-directional mapping process between symbolic and non-symbolic representations of number? *Experimental Psychology, 58,* 39-49. doi:10.1027/1618-3169/a000064

Crollen, V., & Seron, X. (2012). Over-estimation in numerosity estimation tasks: More than an attentional bias? *Acta Psychologica, 140,* 246–251. doi: 10.1016/j.actpsy.2012.05.003.

Defever, E., Sasanguie, D., Gebuis, T., & Reynvoet, B. (2011). Children's representation of symbolic and nonsymbolic magnitude examined with the priming paradigm. *Journal of Experimental Child Psychology, 109,* 174–186. doi:10.1016/j.jecp.2011.01.002

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1-42. doi:10.1016/0010-0277(92)90049-N

Dehaene, S. (1997). *The number sense.* New York, Cambridge (UK): Oxford University Press.

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

De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access?: Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108*(2), 278–292.

Ebersbach, M. (in press). Evidence for a spatial-numerical association in kindergartners using a number line task. *Journal of Cognition and Development.* doi: 10.1080/15248372.2013.805134

Ebersbach, M. (2009). Achieving a new dimension: Children integrate three stimulus dimensions in volume estimations. *Developmental Psychology, 45,* 877-883. doi: 10.1037/a0014616

Ebersbach, M., Luwel, K., Frick, A., 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. doi: 10.1016/j.jecp.2007.08.006

Ebersbach, M., Luwel, K., & Verschaffel, L. (2013). Comparing apples and pears in studies on magnitude estimations. *Frontiers in Cognitive Science, 4,* 332. doi: 10.3389/fpsyg.2013.00332

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307–314. doi:10.1016/j.tics.2004.05.002

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43–74. doi:10.1016/0010-0277(92)90050-R

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends in Cognitive Sciences, 4*, 59-65. doi:10.1016/S1364-6613(99)01424-2

Gebuis, T., & Reynvoet, B. (2012). The role of visual information in numerosity estimation. *PLoS ONE, 7*(5), e37426. doi:10.1371/journal.pone.0037426.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465. doi:10.1037/a0012682

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2011). Language effects on children's mental number line: How cross-cultural differences in number word systems affect spatial mappings of numbers in a non-verbal task. *Journal of Cross-Cultural Psychology, 42*, 598-613.

Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': Numerical estimation by 5- to 7-year-olds. *Cognition, 78*, B27-B40. doi:10.1016/S0010-0277(00)00122-0

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*, 1221-1247. doi:10.1016/j.cognition.2007.06.004

Klein, E., Bahnmueller, J., Mann, A., Pixner, S., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2013). Language influences on numerical development - Inversion effects on multi-digit number processing. *Frontiers in Psychology, 4*. doi:10.3389/fpsyg.2013.00480

Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction, 25,* 95–103. doi:10.1016/j.learninstruc.2012.12.001

Krueger, L. E. (1972). Perceived numerosity. *Attention. Perception & Psychophysics, 11,* 5-9. doi:10.3758/BF03212674

Krueger, L. E. (1982). Single judgments of numerosity. *Attention, Perception & Psychophysics, 31,* 175–182. doi:10.3758/BF03206218

Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Attention, Perception & Psychophysics, 35,* 536-542. doi:10.3758/BF03205949

Lipton, J., & Spelke, E. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development, 76,* 978–988. doi: 10.1.1.68.3127

Mejias, S., Grégoire, J., & Noël, M.-P. (2012). Numerical estimation in adults with and without developmental dyscalculia. *Learning and Individual Differences, 22,* 164–170. doi:10.1016/j.lindif.2011.09.013

Mejias, S., Mussolin, C., Rousselle, L., Grégoire, J., & Noël, M.-P. (2012). Numerical and nonnumerical estimation in children with and without mathematical learning disabilities. *Child Neuropsychology, 18*(6), 550–575. doi:10.1080/09297049.2011.625355

Mejias, S., & Schiltz, C. (2013). Estimation abilities of large numerosities in kindergartners. *Frontiers in Psychology, 4.* 1-12. doi:10.3389/fpsyg.2013.00518

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology, 103,* 490-502. doi:10.1016/j.jecp.2009.02.003

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends in Neuroscience and Education, 1*(1), 21–31. doi:10.1016/j.tine.2012.09.003

Nuerk, H.-C., & Willmes, K. (2005). On the magnitude representation of two digit numbers. *Psychology Science, 47*, 52–72.

Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education increases the acuity of the non-verbal approximate number system. *Psychological Science, 24*, 1037-1043.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron, 53.* 293-305. doi:10.1016/j.neuron.2006.11.022

Pixner, S., Moeller, K., Hermanova, V., Nuerk, H.-C., & Kaufmann, L. (2011). Whorf reloaded: Language effects on nonverbal number processing in first grade—A trilingual study. *Journal of Experimental Child Psychology, 108*(2), 371–382. doi:10.1016/j.jecp.2010.09.002

Power, R. J. D., & Dal Martello, M. F. (1990). The dictation of Italian numerals. *Language and Cognitive Processes, 5*(3), 237–254. doi:10.1080/01690969008402106

Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition, 102*(3), 361–395.

Sarnecka, B., Kamenskaya, A. V., Yamana, Y., Ogura, T., & Yudovina, Y. (2007). From grammatical number to exact numbers: Early meanings of 'one', 'two', and 'three' in English, Russian, and Japanese. *Cognitive Psychology, 55*(2), 136–168. doi:10.1016/j.cogpsych.2006.09.001

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology, 103*(3), 325–337. doi:10.1016/j.jecp.2009.02.007

Sasanguie. D., Smedt, B. de, Defever. F., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology, 30*(2), 344–357. doi:10.1111/j.2044-835X.2011.02048.x

Schlottmann, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development, 72,* 103–122.

Seron, X., & Fayol, M. (1994). Number transcoding in children. Cross-linguistic and functional analysis. *British Journal of Developmental Psychology. 12,* 281-300.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development. 75,* 428-444.

Siegler, R. S., & Booth, J. L. (2005). Development of numerical estimation: A review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 197-212). Boca Ratan, FL: CRC Press.

Siegler, R. S., & Opfer. J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14,* 237-243.

Sophian, C., & Chu. Y. (2008). How do people apprehend large numerosities? *Cognition, 107,* 460-478.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Non-verbal counting in humans: The psychophysics of number representation. *Psychological Science, 10,* 130–137.

White. S. L. J., & Szucs. 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. doi: 10.1186/1744-9081-8-1

Wilson, A. J., & Dehaene, S. (2007). Number sense and developmental dyscalculia. In D. Coch, G. Dawson, & K. Fischer (Eds.), *Human behavior, learning, and the developing brain: Atypical development* (2nd ed., pp. 212–237). New York, NY: Guilford Press.

Wynn, K. (1995). Origins of numerical knowledge. *Mathematical Cognition, 1,* 35-60.

Zuber, J., Pixner, S., Moeller, K., & Nuerk, H.-C. (2009). On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity. *Journal of Experimental Child Psychology, 102*(1), 60–77. doi:10.1016/j.jecp.2008.04.003

Table 1

Linear regressions relating log Ms, log SDs and CVs to target magnitudes

Age group	Task	Measure	R^2	β	F	df	p
Kindergarten	Reproduction	log M	.25	.50	83.88	1, 251	< .001***
		log SD	.05	.23	14.06	1, 243	< .001***
		CV	.00	-.05	0.65	1, 247	.420
	Perception	log M	.16	.40	48.59	1, 262	< .001***
		log SD	.04	.21	9.69	1, 215	.002**
		CV	.00	-.01	.01	1, 249	.931
	Production	log M	.01	.11	3.15	1, 263	.077°
		log SD	.00	.07	1.30	1, 254	.256
		CV	.00	.06	0.94	1, 259	.334
1st Grade	Reproduction	log M	.29	.54	100.32	1, 239	< .001***
		log SD	.07	.27	17.46	1, 229	< .001***
		CV	.00	.00	0.00	1, 236	.974
	Perception	log M	.25	.50	84.36	1, 250	< .001***
		log SD	.05	.23	11.92	1, 212	.001**
		CV	.00	-.03	0.15	1, 245	.701
	Production	log M	.21	.46	66.09	1, 251	< .001***
		log SD	.04	.20	9.98	1, 240	.002**
		CV	.00	-.01	0.03	1, 244	.858
3rd Grade	Reproduction	log M	.63	.79	425.22	1, 251	< .001***
		log SD	.23	.48	74.48	1, 245	< .001***
		CV	.01	.11	3.15	1, 249	.077°
	Perception	log M	.52	.72	271.63	1, 251	< .001***
		log SD	.08	.30	22.02	1, 231	< .001***
		CV	.00	-.02	0.11	1, 248	.740
	Production	log M	.66	.81	483.20	1, 251	< .001***
		log SD	.15	.39	43.75	1, 241	< .001***
		CV	.00	-.03	0.27	1, 249	.603
Adults	Reproduction	log M	.81	.90	1078.82	1, 251	< .001***
		log SD	.16	.40	46.58	1, 241	< .001***
		CV	.00	-.01	0.04	1, 251	.844
	Perception	log M	.59	.77	354.53	1, 251	< .001***
		log SD	.18	.43	49.45	1, 218	< .001***
		CV	.00	.03	0.15	1, 251	.697
	Production	log M	.80	.90	1004.11	1, 251	< .001***
		log SD	.33	.58	118.27	1, 237	< .001***
		CV	.01	.12	3.93	1, 251	.049*

Note. ***p < .001, **p < .01, *p < 05, °p < .10. Scalar variability can be assumed if the regression model plotting the CVs to the target magnitudes is not significant.

Table 2

Main effects task on the mean error rates (ERs) and mean absolute error rates (AERs) per age group

	Task						
Age group	Reproduction	Perception	Production	F	df	p	n^2
Kindergarten							
mean *ER*	.37 (.48)	-.45 (.29)	1.05 (1.01)	29.76	1.30, 25.97	<.001	.60
mean *AER*	.69 (.37)	.63 (.12)	1.31 (0.85)	14.05	1.25, 24.99	<.001	.41
1st Grade							
mean *ER*	.25 (.42)	-.35 (.32)	.79 (.71)	23.15	1.38, 24.74	<.001	.56
mean *AER*	.58 (.26)	.53 (.09)	.99 (.58)	12.48	1.38, 24.78	.001	.41
3rd Grade							
mean *ER*	-.06 (.20)	-.19 (.20)	.37 (.37)	14.27	2, 19	<.001	.60
mean *AER*	.34 (.08)	.38 (.09)	.51 (.26)	6.79	1.18, 23.50	.012	.25
Adults							
mean *ER*	-.05 (.14)	-.24 (.21)	.31 (.32)	25.31	1.19, 23.84	<.001	.56
mean *AER*	.22 (.05)	.33 (.10)	.41 (.23)	9.79	1.33, 26.53	.002	.33

Note. Standard deviations in parentheses.

Table 3

Fits of a power model to the estimations, mean model parameter and test if the exponent deviated significantly from 1

Age group	Task	sign. model fits	mean $_{corr}R^2$	mean β	mean constant	t	df
Kindergarten	Reproduction	73%	.37	.57	6.90	-7.58*	15
	Perception	73%	.38	.67	3.28	-3.26*	15
	Production	27%	.21	.37	19.35	-12.99*	5
1st Grade	Reproduction	82%	.39	.54	7.22	-10.47*	17
	Perception	91%	.43	.58	3.35	-8.00*	19
	Production	77%	.39	.60	9.89	-6.05*	16
3rd Grade	Reproduction	100%	.59	.74	2.71	-6.98*	20
	Perception	95%	.59	.62	4.00	-7.77*	19
	Production	100%	.75	.95	1.97	-1.28	20
Adults	Reproduction	100%	.84	.83	1.85	-6.62*	20
	Perception	100%	.79	.66	2.69	-14.91*	20
	Production	100%	.91	1.17	0.85	3.42*	20

Note. The mean model parameter include only to those cases in which the model yielded a significant fit ($p <$ .10). *T*-statistics refers to the test if the exponent β deviated significantly from 1. * $ps < .05$ (Bonferroni)

Table 4

Correlations between the mean error rates (ERs) in each task and children's number knowledge

Age group	Measure	*ER* reproduction	*ER* perception	*ER* production
Kindergarten	*ER* perception	-.06		
	ER production	.57**	-.45*	
	counting span	-.39•	-.28	-.43*
	decade change	-.44*	-.08	-.38°
1st Grade	*ER* perception	-.47*		
	ER production	.63**	-.64**	
	counting span	-.01	.10	-.05
	decade change	-.04	.21	-.2•
3rd Grade	*ER* perception	-.30		
	ER production	.15	-.47*	
Adults	*ER* perception	.57**		
	ER production	-.29	-.53*	

Note. **$p < .01$, *$p < .05$, °$p < .10$. Only non-redundant correlations are shown.
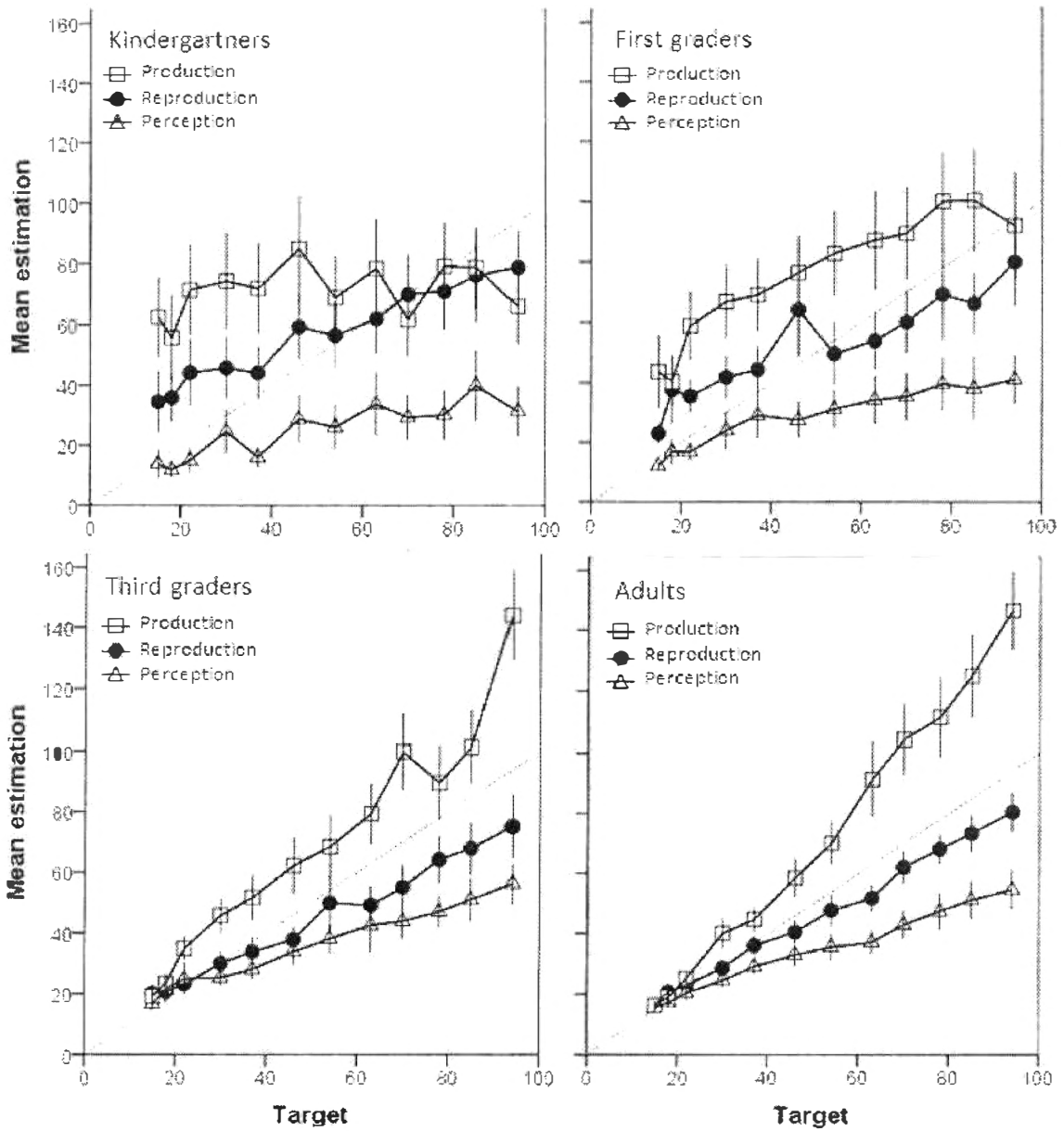
Figure captions

Figure 1. Mean estimations, separately for each age group and task. *Note.* Dashed line indicates normative solution.

Figure 2. Mean *ERs* and standard errors, separately for each task and the familiar or unfamiliar number range (for kindergartners and first graders not familiar with numbers to 100). *Note.* $**p < .01$, $*p < .05$, $°p < .10$; p-values refer to significant deviations from 0.

Figure 1. Mean estimations, separately for each age group and task. *Note.* Dashed line indicates normative solution.
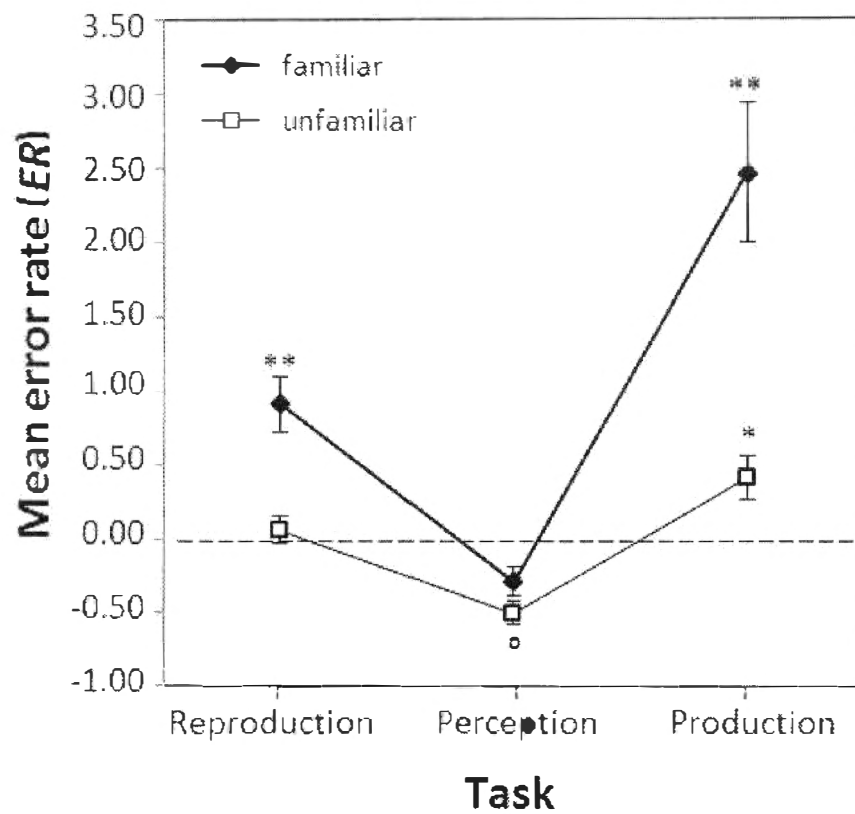
Figure 2. Mean *ER*s and standard errors, separately for each task and the familiar or unfamiliar number range (for kindergartners and first graders not familiar with numbers to 100). *Note.* **p < .01, *p < .05, °*p* < .10; *p*-values refer to significant deviations from 0.

Highlights

- Children and adults estimated quantities either symbolically or non-symbolically.
- Pattern of third graders and adults in line with bi-directional mapping hypothesis.
- Different pattern in kindergartners and first graders.
- Biases in young children due to their limited number knowledge.