The interplay between spatial ordinal knowledge, linearity of number-space mapping, and arithmetic skills

PLEASE CITE THE PUBLISHED VERSION

https://doi.org/10.1016/j.cogdev.2020.100915

PUBLISHER

Elsevier

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This paper was accepted for publication in the journal Cognitive Development and the definitive published version is available at https://doi.org/10.1016/j.cogdev.2020.100915.

The interplay between spatial ordinal knowledge, linearity of number-space mapping, and arithmetic skills

Running title: Spatial mapping of numbers

Francesco Sella[1,2], Daniela Lucangeli[3,4], Marco Zorzi[5,6]

[1]*Department of Psychology, University of Sheffield, UK*
[2]*Centre for Mathematical Cognition, Loughborough University, UK*
[3]*Department of Developmental Psychology and Socialisation, University of Padova, Italy*
[4]*Mind4children, University of Padova, Italy*
[5]*Department of General Psychology and Padova Neuroscience Center, University of Padova, Italy*
[6]*Fondazione Ospedale San Camillo IRCCS, Venezia, Italy*

This manuscript is "*in press*" in Cognitive Development.

Citation: Sella, F., Lucangeli, D., & Zorzi, M. (in press). The interplay between spatial ordinal knowledge, linearity of number-space mapping, and arithmetic skills. *Cognitive Development*.

Acknowledgements: This study was supported by a Cariparo Foundation "Excellence Grants 2017" to M.Z. The authors thank the children and their parents for participating in this study as well as Chiara Pozzolo for her help in collecting data.

Abstract

The ability to map numbers onto space has been widely investigated with the number line (NL) task. Accurate (linear) placement is typically preceded by a developmental phase in which children assign more space to small numbers and compress large numbers to the right part of the line, thereby resembling a biased log-like mapping. Here, we exploited a task that separately assesses direction, order and accuracy of spatial mapping (DOS task) to investigate the origin of the biased NL mapping in a sample of first and second graders. Children with reduced ordinal knowledge in the DOS task showed a more biased NL mapping, which was formally assessed using the leading mathematical models of the NL task. However, the log-like biased mapping did not emerge in the DOS task, thereby showing a lack of generalisation across different number-space mapping tasks. Both ordinal knowledge in the DOS task and linearity in the NL task related to arithmetic fluency beyond domain-general cognitive factors. We conclude that the NL task is a meter of children's arithmetic skills rather than an expression of the mental representation of numbers.

Introduction

The number line task (NL) is a widely used paradigm to assess children's ability to place numbers (or other quantities) in space (Opfer & Siegler, 2007; Sella, Berteletti, Lucangeli, & Zorzi, 2015; Siegler & Opfer, 2003; Siegler, Thompson, & Opfer, 2009). Individuals mark the position of several target numbers on a horizontal line entailing a numerical interval (e.g. 0-100). After placing a target number, a new line is presented with the request to place another target number. Children's performance is characterised by a shift from a biased (log-like) to an accurate (linear) positioning of numbers. For instance, 5-year-old preschool children overestimate the position of small numbers and gradually compress the position of large numbers in the numerical interval 0-100 (Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010), whereas seven-year-old children display an accurate mapping (Sella et al., 2015).

The biased positioning observed in young children is an intriguing and highly controversial aspect in the development of numerical skills (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen, Blanc-Goldhammer, & Quinlan, 2018; Cohen & Sarnecka, 2014; Dackermann, Huber, Bahnmueller, Nuerk, & Moeller, 2015; Ebersbach, 2016; Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008; Moeller, Pixner, Kaufmann, & Nuerk, 2009; Opfer, Siegler, & Young, 2011; Opfer, Thompson, & Kim, 2016; Rouder & Geary, 2014; Sella et al., 2015; Slusser, Santiago, & Barth, 2013). In the seminal study of Siegler and Opfer (2003), the finding that a logarithmic function well describes the biased estimates (Booth & Siegler, 2006; Siegler & Booth, 2004; Siegler et al., 2009) was linked to the properties of the Approximate Number System (ANS; Dehaene, 1992, 2003; but see Cohen & Quinlan, 2018). The leading model of the ANS assumes that numerical quantities are represented on a logarithmic scale as Gaussian distributions of activation with fixed standard deviation; accordingly, children would progressively reduce the space separating numbers on the number line as numerical magnitude increases.

The logarithmic to linear shift, however, neglects the role of strategies that individuals implement when solving the task. This limitation prevents from obtaining an unbiased assessment of number representation from the task. Children and adults who are familiar with the presented numerical interval display an accurate linear positioning of numbers associated with the implementation of strategies, such as mentally dividing the line and using anchoring points (Peeters, Degrande, Ebersbach, & Verschaffel, 2016; Peeters, Verschaffel, & Luwel, 2017; Rouder & Geary, 2014; Sella, Sader, Lolliot, & Cohen Kadosh, 2016; Sullivan, Juhasz, Slattery, & Barth, 2011; White & Szűcs, 2012). The use of strategies emerges quite early in development, as first and second graders already use the endpoints and the middle of the line as anchor points (Rouder & Geary, 2014). In this light, it has been suggested that individuals perform a proportional judgement when solving the NL task (Barth & Paladino, 2011). That is, the position of a target number is not estimated in isolation but in relation to the numerical interval, as a part of a whole. Accordingly, instead of logarithmic and linear functions, cyclical power models have been fit to the NL data to describe the pattern of under- and overestimation (Barth & Paladino, 2011; Cohen et al., 2018; Cohen & Quinlan, 2017; Friso-van den Bos et al., 2015; Hollands & Dyre, 2000; Spence, 1990; but see Opfer et al., 2011). For instance, the biased mapping observed in young children can be explained by the inability to implement subtraction and division strategies when placing numbers (Cohen & Sarnecka, 2014).

Another influential account ascribes the biased mapping to the familiarity with the target numbers (Ebersbach, 2016; Ebersbach et al., 2008; Ebersbach, Luwel, & Verschaffel, 2015; Moeller et al., 2009). Children would linearly place small familiar numbers, whereas larger unknown numbers would be mainly placed in the rightmost part of the line. Accordingly, spatial mapping is modelled using a bilinear function: a first straight line with a steep positive slope fits the estimates of small target numbers until a breakpoint, whereas the estimates of the remaining (larger) target numbers are fitted by a second straight line with a slightly positive slope. The effect of familiarity emerges when children have to place numbers on different numerical intervals. Accordingly, the same child displays linear positioning with a familiar and small numerical interval (i.e., 1-20) but biased positioning emerges with a non-familiar and large interval (i.e., 0-100). In the latter scenario, children would unorderly place large unfamiliar target numbers (e.g., between 20 and 100) toward the rightmost part of the line.

The empirical evidence driving the debate on the spatial mapping of numbers is entirely based on the NL task, which does not readily allow to dissect behavioural performance into task components. Three main components should be taken into account when investigating the spatial positioning of numbers: direction, spatial order (ordinality) and spacing (Sella, Lucangeli, & Zorzi, 2018). Numbers can be placed in order according to a preferred direction, namely, from left-to-right (e.g., 1-2-3) or right-to-left (e.g., 3-2-1). Spatial order refers to the ordinal position of numbers on the number line (i.e., 1---2---3; and not 2---1---3), whereas spacing refers to the spatial distance in relation to the numerical magnitude (e.g., 1---2---3; and not 1-2-----3). In the classic version of the NL task, direction is firmly established both in the display (the visual line) and in the experimental instructions (usually from left-to-right; but see Ebersbach, 2013). Moreover, children place each target number on a different line; it is, therefore, difficult to verify whether children know the ordinal relation between target numbers (Sullivan & Barner, 2014), especially in the case of biased and compressed positioning, whereby large numbers are placed close to each other. For instance, a child might place the target number 53 on the line in the position corresponding to number 60 and, in the next trial, the target number 65 in the position of number 55, thereby not respecting the ordinal relation between numbers (i.e., 65----53).

Once spatial order is respected, the assessment of spacing becomes possible by observing whether children respect the distance between target numbers. A precise assessment of the spacing between target numbers becomes particularly relevant in the case of a biased (log-like) positioning. For instance, a compressed mapping implies that the estimated distance between 1 and 5 would roughly match the distance between 5 and 25 (i.e., 1---5---25). What would happen if the distance between 1 and 5 was fixed and children had to place the number 9? In the case of compressed mapping, the target number 9 should be placed near 5 (e.g., 1---5-9), whereas in case of linear spacing it should be placed far from 5 so that the distance matches the distance from 1 to 5 (e.g., 1---5---9). Overall, a detailed investigation of spatial order and spacing of small and large numbers along the visual line would disentangle the nature of the compressed positioning in the NL task.

Here, we used a task to separately assess spatial order and spacing of numbers (DOS, Direction Order and Space task; Sella, Lucangeli, & Zorzi, 2018) while mimicking the structure of the NL task, thereby providing a new source of evidence to the debate on the spatial mapping of numbers. We administered the DOS task and the NL task to a sample of first and second graders using numbers ranging from 0 to 100 to maximise the possibility to observe a range of biased and linear positioning in the NL task. Moreover, we used the same target numbers to match the level of difficulty across the two spatial mapping tasks. We fitted the log-linear, proportional, and bilinear model to the NL estimates to assess the presence of biased log-like mapping and breakpoints separating familiar and unfamiliar numbers. If familiarity exerts an effect on the biased mapping in the NL task, children with a partial ordinal knowledge should show a more evident biased mapping compared with children with perfect ordinal knowledge. Moreover, children with a partial ordinal knowledge should show lower breakpoints in the bilinear function compared to children with perfect ordinal knowledge. Once spatial order is taken into account, we expect to observe children with a biased mapping in the NL task to display a compressed spacing (e.g., 1---5-9) in the DOS task.

Finally, we examined the relation between the performance in the NL task and DOS task and arithmetic skills while controlling for domain-general cognitive factors and the numerical magnitude knowledge as assessed in a number comparison task. In particular, we aimed to understand whether the performance in the NL task still relates to arithmetic skills (Booth &

Siegler, 2006; Link, Nuerk, & Moeller, 2014; Siegler & Booth, 2004) when spatial order and spacing from the DOS task are taken into account. On the one hand, a precise assessment of order and accuracy in placing small and large triplets of numbers might be enough to account for the influence of spatial mapping on arithmetic skills. On the other hand, the NL task might still capture a more global understanding of the relative relation of target numbers within a large numerical interval.

Method

Participants

One hundred sixty-six primary school children from two schools located in northeastern Italy took part in the study after parents or legal guardians gave their informed consent. Seventy-three children attended the first grade, whereas the remaining 93 attended the second grade. Four children missed one of the sessions, and five interrupted the experiment. The final sample was composed of 157 children (73 boys; $M_{age\text{-}months}$=87, SD=7, range=70-100). We calculated the age as the difference in months between the month and the year in which the child was born and the month and the year of the first testing session. Parents or legal guardians filled in a questionnaire regarding the family background and demographic information (e.g., nationality, parents' education). According to parents' and legal guardians' questionnaires, 137 children were born in Italy, 5 in another country whereas this information was not reported for 15 children. Three children among those born in another country arrived in Italy at the age of 1, one at the age of 5, and one at the age of 7. According to the experimenter's judgment, all children demonstrated an understanding of the task instructions. The highest level of education achieved by one of the parents/guardians indicated a middle socio-economic status for most of the families ("Not declared"=11, "primary school"=2, "middle school"=23, "high school"=57, "degree/further education"=64).

Tasks

Intelligence. Children completed the Raven coloured progressive matrices (Raven, 1998). We computed the number of correct responses (max=36).

Memory. Children completed the forward digit span and the forward Corsi block (BVS; Mammarella, Toso, Pazzaglia, & Cornoldi, 2008). In the digit span, there were three trials for each span, starting from the span of three to nine for a total of 21 trials. In the Corsi block, there were three trials for each span, starting from the span of three to eight for a total of 18 trials. The experimenter interrupted the task in case of two errors within the same span. For each child, we calculated the proportion of correct responses as the number of correct responses divided by the total number of possible trials.

Number comparison. Children pressed the corresponding mouse button to indicate the larger between two numbers presented on the left and right side of the computer screen, respectively. There were 12 comparisons (i.e., 1-5, 1-9, 5-9, 11-15, 11-19, 15-19, 41-53, 41-65, 53-65, 74-86, 74-98, 86-98) repeated four times with the larger number on the right side of the screen in half of the trials. The experimenter invited children to respond correctly and as fast as they could. We calculated the proportion of correct responses.

Arithmetic fluency. Children completed the first two columns, additions and subtractions, of the Tempo Test Rekenen (De Vos, 1992). Each column contains forty arithmetic problems and children had two minutes at disposal to solve as many problems as they could. Children first completed the column with additions and, after a brief pause, the column with subtractions. We calculated the total number of correctly solved problems as an index of arithmetic fluency.

Number line task (Siegler & Opfer, 2003). Children saw a horizontal line (length x width: 960 x 3 pixels) in the middle of the screen entailing the numerical interval from 0 to 100 and had to sequentially place twelve randomly represented numbers (i.e., 1, 5, 9, 11, 15, 19, 41, 53, 65, 74, 86, 98) in their correct locations on the line by moving the mouse cursor. The target number appeared in red ink above the line and moved according to the mouse movements. The experimenter said: "*This line goes from zero to one hundred* [pointing at the endpoints of the line]. *Where is the correct position of the number X on the line? Move the mouse and click one of the buttons to place the number*". The children clicked on one of the mouse buttons to place the target number on the selected location. After placing each target number, children had to confirm the response before moving to the next trial. There were two training trials, 0 and 100, in which the experimenter provided feedback in case of incorrect positioning. The feedback on 0 and 100 ensured children were aware of the entire numerical interval, even though it might have promoted the use of the endpoints as anchoring points, thereby increasing the linearity of the estimates (Slusser et al., 2013).

Direction, Order and Space (DOS) task (Sella et al., 2018). In the DOS task, children arranged two sequentially presented numbers spatially on a visual line (Figure 1). Children saw a black horizontal line in the middle of the screen, and a number (hereafter centred number) placed below the middle point of the line. Another number (hereafter target 1) appeared above the line as soon as the child moved the mouse. According to the leftward and rightward movement of the mouse, the target 1 appeared only in two possible locations, one left and one right, equidistant (i.e., 128 pixels) from the centred number. The experimenter said: "*if the number* x [pointing at centred number] *is here, where should the number y* [pointing at target 1] *be placed*?" Children placed the target 1 by clicking one of the mouse buttons. When the child clicked one of the mouse buttons, target 1 appeared in the selected location below the line. Then, the experimenter asked the child whether she was sure about her decision; otherwise, she could repeat the trial and place target 1 again. If the child placed target 1 exactly on the location of the centred number, a warning message appeared informing the child that the position was already occupied and inviting the child to find another location. The position of target 1 determined the direction of the mapping, either left-to-right (e.g., 1-5) or right-to-left (e.g., 5-1). Once target 1 was placed, another number (hereafter target 2) appeared above the centred number, and children placed it on the line. The experimenter said: "*If the number* x [pointing at centred number] *is here and you placed the number* y [pointing at target 1] *here*, *where should the number* z [pointing at target 2] *be placed*?". The placement of target 2 determined whether the child possessed a congruent spatial order of the three presented numbers, regardless of directionality (e.g., 1-5-9 or 9-5-1). Moreover, the distance between the centred number and target 1 represented a fixed interval that acted as a reference to place target 2. Target 2 could be moved along the line using the mouse cursor, whose movement was restrained so that, in case of respected spatial order of the three numbers, there was the same amount of under- or over-estimation (e.g., underestimation: 1---5-9, overestimation: 1---5-----9). Nonetheless, children could move the cursor of the mouse to the same extent on the opposite side of the line, even if locating target 2 in that segment would not respect spatial order. Target 2 appeared in the selected location below the line when the child clicked one of the mouse buttons. Then, the experimenter asked the child whether she was sure about her decision; otherwise, she could place target 2 again. If the child placed target 2 on the location of the centred number or target 1, a warning message appeared inviting the child to find another location because another number already occupied the selected one. There were four triplets 1-5-9, 11-15-19, 41-53-65, and 74-86-98, whose numbers were presented in four different orders (e.g., 5-1-9, 5-9-1, 1-5-9, and 9-5-1) twice for a total of 32 trials. Children took a brief rest after 16 trials. Half of the trials had a two-side arrangement, whereas the other half had a one-side arrangement. In the two-side trials (Figure 1a), the target numbers should be placed one on the left and one on the right side compared to the centred number to achieve correct ordinality; in one-side trials (Figure 1b), both target numbers should be placed on the left or the right side of the centred number to achieve correct ordinality. There was a training trial (i.e., 2-1-3 and 1-2-3) repeated twice to let the child familiarise with the task. For each trial, we recorded whether a child placed numbers in the correct order. In case of a respected order, we also measured the direction of mapping (i.e., left-to-right or right-to-left) and the proportion of absolute error as the difference between the estimated target 2 and the target 2 divided by the maximum of under- and overestimation (i.e., 4 for small triplets and 12 for large triplets). In case of correct ordering, the absolute proportional error could vary between 0 and 1 given the constraint to the movements of the mouse. A video showing examples of a one-side trial and a two-side trial is provided on the Open Science Framework platform (https://osf.io/3sefg/?view_only=b57b3a3d7be14a33a68016f7c860db99).

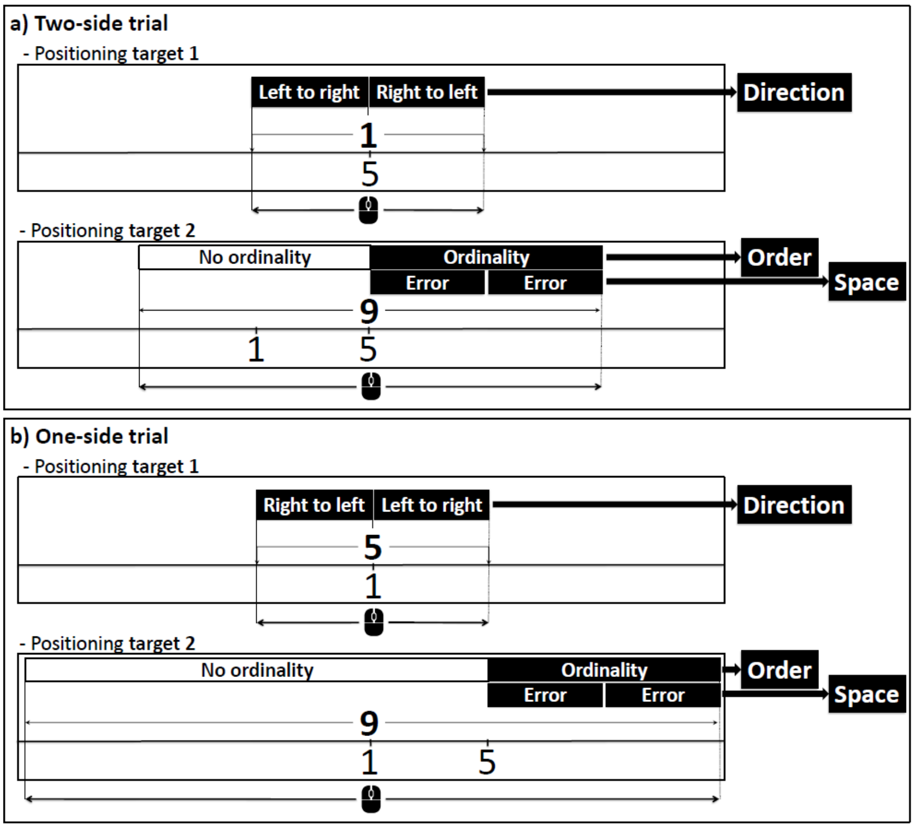

Figure 1. a) An example of a two-side trial for the triplet 1-5-9. The number 5 was presented in the centre of the line (i.e., centred number). Then, the child moved the mouse leftward or rightward to place the number 1 (i.e., Target 1) in one of the two fixed locations. Thereafter, the number 9 (i.e., target 2) appeared above the centred number and the child needed to find its correct location on the line. The positioning of Target 2 determined whether the triplet of numbers was correctly ordered. b) Example of one-side trial for the triplet 1-5-9. The structure was identical to that of the two-side trials. In the one-side trials, both targets 1 and 2 were to be placed on the left or right side with respect to the centred number to achieve a correctly ordered triplet.

Procedure

Children completed the tasks in two sessions, usually one day apart (Mode=1, *M*=1.8, *SD*=1.3, range=1-7). The study was conducted over six months so that some children completed the tasks at the beginning of the school year (i.e., November) and some children at the end (i.e., May). Children completed the NL task and the DOS task in one session (named A) in counterbalanced order and all the remaining tasks (i.e., number comparison, memory tasks, Raven's matrices, and Tempo Test) in another session[1] (named B) following four possible presentation orders: number comparison, Tempo Test, Raven's matrices, digit span, Corsi block; number comparison, Tempo Test, Raven's matrices, Corsi block, digit span; number comparison, Tempo Test, digit span, Corsi block, Raven's matrices; number comparison, Tempo Test, Corsi block, digit span, Raven's matrices. The presentation of session A and B was counterbalanced for a total of 16 different presentation orders. The computerised tasks were programmed using the software E-prime 2.0 (Psychology Software Tools, 2012; screen resolution set to 1280x768) and presented on a laptop (screen size: 15.9 inches) The study was approved by the Ethics Committee for Psychological Research of the University of [*blind for review*].

Results

We ran statistical analyses using the free software R (R Core Team, 2016) along with the BayesFactor package (Morey & Rouder, 2015) using default priors for Bayesian analyses. We reported Bayes factors (BF_{10}) expressing the probability of the data given H1 relative to H0 (i.e., values larger than 1 are in favour of H1 whereas values smaller than 1 are in favour of H0). When comparing regression models, we reported the Bayes factors (BF) as the ratio of BFs_{10} between compared models. If the ratio between BF_{10} of model A and BF_{10} of model B is larger than 1, then there is evidence for model A. Conversely, if the ratio is smaller than one there is evidence for model B. We described the evidence associated with BFs as "anecdotal" ($1/3 < BF < 3$), "moderate" ($BF < 1/3$ or $BF > 3$), "strong" ($BF < 1/10$ or $BF > 10$), "very strong" ($BF < 1/30$ or $BF > 30$), and "extreme" ($BF < 1/100$ or $BF > 100$) (Jeffreys, 1961). The raw data and the R code for the analysis are available on the Open Science Framework platform (https://osf.io/3sefg/?view_only=b57b3a3d7be14a33a68016f7c860db99).

Spatial mapping of numbers in the NL task

We aimed to describe children's compressed (log-like) positioning in the NL task. First, we ran a linear and logarithmic fit on the estimates as a function of target numbers. We retained only those participants with at least one significant model (p<.05) and a positive slope of the linear model. The excluded participants (n=15) mainly placed all the numbers in the middle of the line, at the endpoints, or even showed a reversed mapping (i.e., large numbers placed on the

[1] Twelve children completed the number comparison task in the same session of the NL task and DOS task.

left side of the line and small numbers on the right side). Then, we evaluated the fit of three different statistical models previously used to describe the pattern of NL estimates: i) mixed log-linear model (Anobile, Cicchini, & Burr, 2012; Opfer et al., 2016), ii) proportional (one-cycle unbounded) model (Barth & Paladino, 2011; Spence, 1990), and iii) bilinear model (Ebersbach et al., 2008; Moeller et al., 2009). Models' fittings on individual estimates are reported in the Supplementary Materials.

The mixed log-linear entails the flexible estimation of a lambda coefficient (λ) that gets closer to 0 in case of perfect linear fit and close to 1 in case of a logarithmic fit. Accordingly, we constrained the lambda values to range from 0 to 1. The average of the estimated lambda coefficient was 0.68 (*SD*=0.32, range=[0, 1]) with an average R^2 of 0.83 (*SD*=0.18).

The unbounded one-cycle proportional model includes two parameters, which express the exponent of the power function (β) and the estimated numerical magnitude of the entire line (*W*). The β coefficient approaches 0 when estimates are compressed whereas it approaches 1 in case of linear estimates. With the exclusion of four participants, for whom the model failed to converge, fitting the proportional model returned an average β of 0.47 (*SD*=0.17, range=[0.15, 1.07], *n*=138) and an average *W* of 121.2 (*SD*=32.11, range=[98.01, 330.73] , *n*=138), with an average R^2 of 0.81 (*SD*=0.2). We also fit the bounded one-cycle model that only estimates the exponent β with the numerical magnitude of the line fixed at 100 (i.e., the entire line). The correlation between the βs from the bounded and unbounded models was almost perfect ($r(136)=0.99$, $p<.001$). Therefore, we retained the β values from the unbounded model, knowing that using those of the bounded model would return virtually identical results.

The bilinear model estimates the individual breakpoints separating familiar from non-familiar numbers. We implemented the bilinear fit proposed by Kaps and Lamberson (2004). We selected the bilinear function with the highest adjusted R^2 when the slope of the first regression line was larger compared to the second one, following the log-like pattern of estimates, and the breakpoint was between 0 and 100. Children displayed an average breakpoint of 14.23 (*SD*=10.27, range=[0.42, 71.6]) with an average R^2 of 0.9 (*SD*=0.1). We assessed whether the slopes of the two lines showed a significant difference (i.e., Davies-test; Muggeo, 2014) to guarantee the presence of a discontinuity between familiar and unfamiliar numbers. The difference between the two slopes was significant only for 69 participants.

The bilinear fit yielded the highest average R^2 compared to the other two models. However, it should be noted that the bilinear model has more parameters due to the fitting of two lines. Therefore, we computed the Bayesian Information Criterion (BIC) for each participant and for each model. The BIC accounts for model complexity and provides a measure that penalizes the model in proportion to its number of parameters. The bilinear model yielded the lowest BIC for 61 participants, the mixed log-linear model for 42, and the proportional model for 39. Nevertheless, the model with the lowest BIC (i.e., the best model) could have a BIC close to the model with the second-lowest BIC, making the two models virtually identical in their precision in fitting the data. To overcome this limitation, for each participant, we subtracted the lowest BIC from the BIC of the remaining two models. We considered those models with a difference of BIC below 2 (Raftery, 1995) as essentially comparable to the model with the lowest BIC (i.e., the best model). For instance, for a participant, the bilinear model yielded a BIC of 89, the mixed log-linear of 90, and the proportional model of 95. We subtracted 89 (i.e., the lowest BIC) from 90 and 95, giving 1 and 5, respectively. In this case, the bilinear and the mixed log-linear models were considered the best models, whereas the proportional model was not. When all models had similar BIC (e.g., 88, 89, and 87.5), they all were considered the best models. When one model had a clearly lower BIC (e.g., 77, 90, and 93), it was considered the only best model. After this calculation, the bilinear model was the best model for 91 participants out of 138, the mixed log-linear for 78 and the proportional model for 73. Finally, we also calculated the sum of BICs across participants for the three models: the bilinear model yielded a sum of 12282, the mixed log-linear model of 12416, and the proportional model of 12536.

The relation between the DOS and the NL task

First, we calculated the proportion of correctly ordered trials in the DOS task (*M*=0.94, *SD*=0.13), the proportion of mapping in the left-to-right direction (*M*=0.93, *SD*=0.21) and the proportion of absolute error of correctly ordered trials (*M*=0.30, *SD*=0.20). Then, we separated children with perfect ordinal knowledge (86 knowers) from those who have not achieved it yet (71 partial knowers). There was anecdotal evidence for group difference in the proportion of left-to-right mapping, though both groups mainly placed numbers in the conventional left-to-right direction (BF_{10}=1.85; Partial knowers: *M*=0.89, *SD*=0.23; Knowers: *M*=0.96, *SD*=0.18), and in the proportion of error (BF_{10}=0.61; Partial knowers: *M*=0.33, *SD*=0.2; Knowers: *M*=0.27, *SD*=0.21).

If ordinal knowledge and familiarity with numbers exert an influence on the NL task, we would expect partial knowers and knowers to show a different pattern of estimates when placing numbers on the visual line. In particular, we expected partial knowers to show biased compressed mapping associated with larger λ coefficients of the mixed log-linear model, as well as with smaller β and larger W of the proportional model, compared to knowers. Moreover, partial knowers should be more likely to belong to the group of children having a clear breakpoint as assessed by the bilinear fit, and the breakpoints should be larger for knowers than partial knowers. The number of partial knowers and knowers varied in the following analyses due to the exclusion of those children with a non-numerical mapping in the NL task or failure in fitting the model. Partial knowers had a larger λ of the mixed log-linear model (M=0.77, SD=0.28, n=56) as their mapping in the NL task was more biased compared to knowers (M=0.62, SD=0.33, n=86; BF_{10}=5.13, moderate evidence; Figure 2a). We found anecdotal evidence for a difference in the β between partial knowers (M=0.43, SD=0.14, n=56) and knowers (M=0.49, SD=0.18, n=82; BF_{10}=1), whereas there was moderate evidence for a similar W (Partial knowers: M=120.61, SD=19.62; Knowers: M=121.61, SD=38.5; BF_{10}=0.19; Figure 2b-c). As expected, for 35 out of 56 partial knowers we observed a difference between slopes of the bilinear fit, whereas this was the case only for 34 out of 86 knowers (Contingency table test; BF_{10}=7.42). Contrary to what expected, these 35 partial knowers displayed larger breakpoints (M=16.42, SD=8.5) compared to the 34 knowers (M=12.76, SD=4.82), though the evidence for such difference was only anecdotal (BF_{10}=1.87; Figure 2d).

In summary, children with partial and complete ordinal knowledge differed in their performance in the NL task. The λ mixed log-linear model better captured this difference compared to the parameters extracted from the proportional and bilinear model. The bilinear model provided a better fit to the estimates compared to the mixed log-linear model and the proportional model. However, less than half of the participants displayed a significant difference between slopes, as a valid bilinear fit should entail. Moreover, the extracted breakpoints did not differentiate between children with partial and complete ordinal knowledge, thereby questioning whether the bilinear fit separates between familiar and unfamiliar numbers.

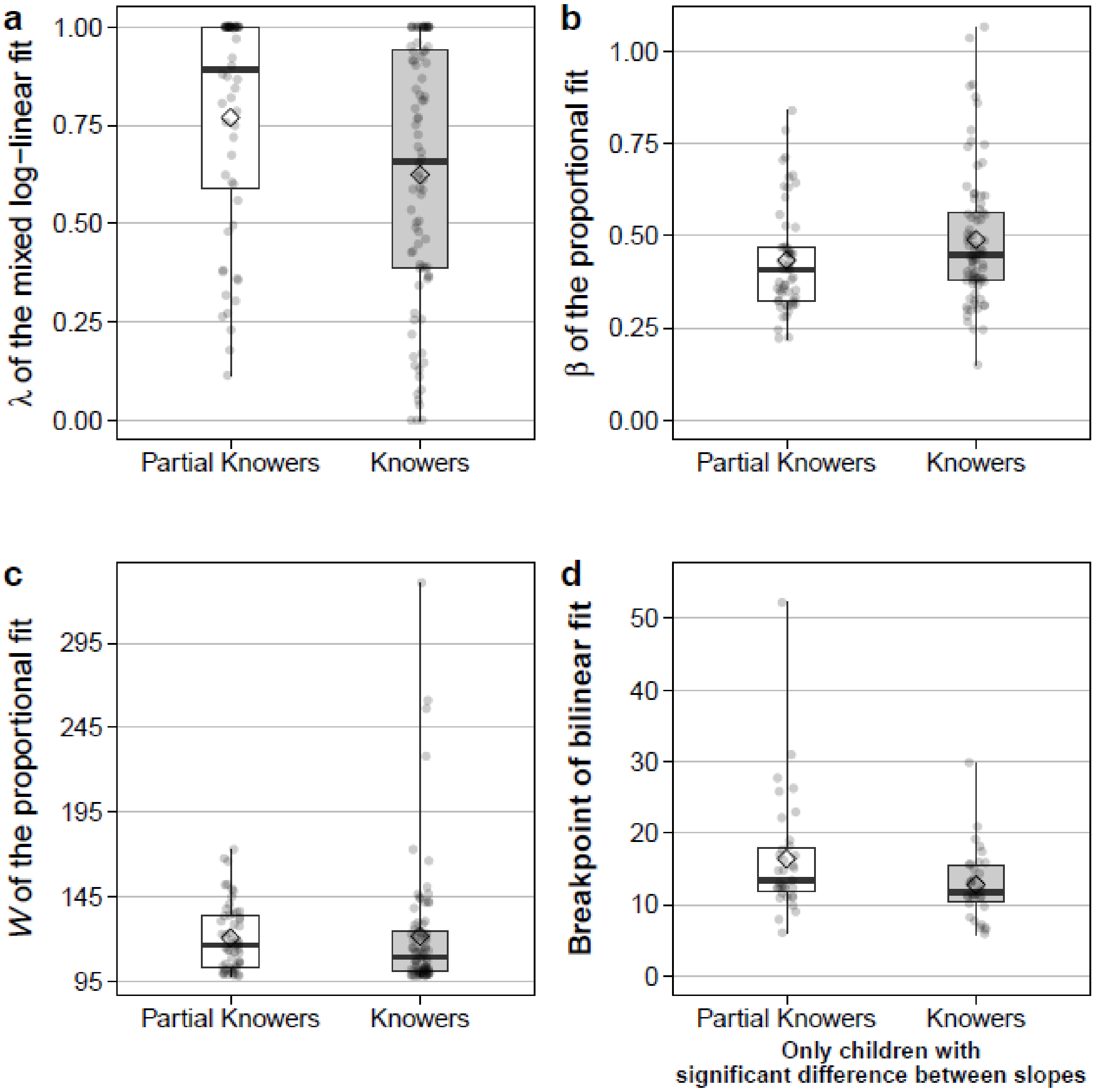

Figure 2. Boxplots representing quartiles of the distributions for (a) λ of the mixed log-linear fit, (b) the β and the (c) W of the proportional fit, and (d) the breakpoints of the bilinear fit separately for children with partial (Partial Knowers) or perfect (Knowers) ordinal knowledge in the DOS task. Transparent dots represent individual values, whereas the diamond represents the mean.

Spatial mapping of numbers in the DOS task

Once ordinality is taken into account, the DOS task can reveal how children space numbers apart, especially if they can respect the equidistance between numbers or display a biased mapping, such as a compressed one. The pattern of error in mapping numbers becomes particularly interesting in the case of the children with partial and perfect ordinal knowledge in the DOS task. We already observed that these children display the same amount of error in the DOS task, but we aimed to verify whether the pattern of errors differed between the two groups.

In correctly ordered trials, the position of target 2 could vary between -4 and +4 from the correct position on the line (i.e., zero error) when the triplets were small (i.e., 1-5-9, 11-15-19) and between -12 and +12 from the correct position when the triplets were large (i.e., 1-5-9, 11-15-19). We rescaled the estimates by dividing them by four or twelve according to whether the triplet was small or large. Therefore, the rescaled estimates could vary between -1 and 1. When the magnitude of target 2 was small compared to target 1 or the centred number (CN), a positive error value represented underestimation, and a negative value represented overestimation (Figure 3a). Conversely, when the relative magnitude of target 2 was large, a positive value represented overestimation, and a negative value represented underestimation (Figure 3b). Crossing under- vs over-estimation of the position of target 2 with relative magnitude (small vs large) leads to four qualitatively different types of mapping, as depicted by the four quadrants on a Cartesian plane (Figure 3c).

Underestimation mapping (Figure 3d) indexes a consistent underestimation of the position of target 2. This mapping might reveal the tendency of positioning target 2 before getting to the correct position as children assume that the task, once numbers are ordered, has been completed and it is time to move to the next trial. *Expansion* mapping (Figure 3e) indexes that the position of target 2 is underestimated when the number is smaller than CN or target 1 while it is overestimated when larger than CN and target 1. This mapping reflects an expanding number line, whereby the spatial distance between numbers gets larger with increasing numerical magnitude. In the *linear* mapping (Figure 3f), the equidistance between numbers is achieved. *Overestimation* mapping (Figure 3g) indexes a consistent overestimation of the position of target 2. This mapping might reveal the tendency of positioning target 2 at the extremes on the line by moving the mouse cursor as far as possible, in a sort of all-to-the-left and all-to-the-right movements. *Compression* mapping (Figure 3h) indexes that the position of target 2 is overestimated when the number is smaller than CN and target 1 while it is underestimated when larger than CN and target 1.

The compression mapping resembles the compressed mapping of the number line, as assessed in the NL task. For instance, in case of a compressed mapping, children would assume that the spatial distance between 1 and 5 is larger compared to the spatial distance between 5 and 9 (i.e., 1---5-9). In the trials where target 2 was smaller than CN and target 1, the distance between 5 and 9 was fixed, and children had to place the number 1. One would expect a child with a compressed mapping to place the number 1 so that the spatial distance between 1 and 5 is larger compared to the spatial distance between 5 and 9 (i.e., 1------5---9). Conversely, in the trials where target 2 was larger than CN and target 1, the distance between 1 and 5 was fixed, and children had to place the number 9. One would expect a child with a compressed mapping to place the number 9 so that the spatial distance between 5 and 9 is smaller compared to the spatial distance between 1 and 5 (i.e., 1---5-9).

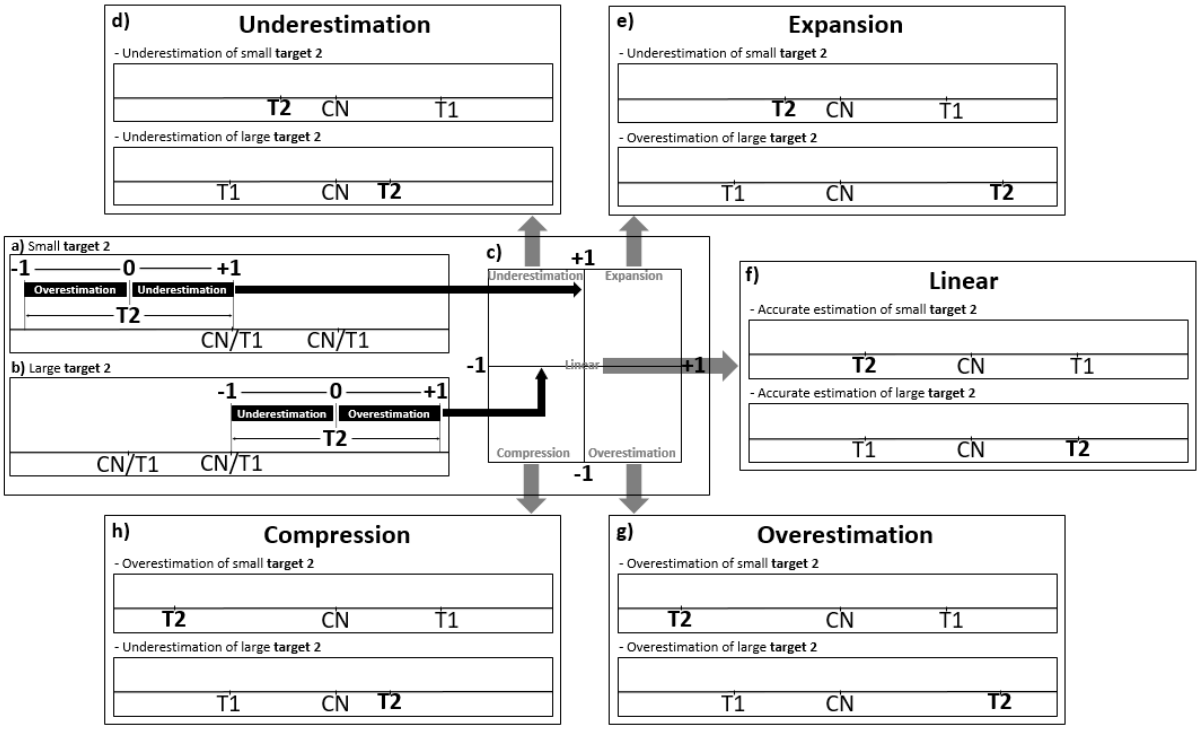

Figure 3. Schematic representation of the positioning of target 2 when it was (a) smaller or larger (b) compared to target 1 or the centred number (CN). When target 2 was smaller compared to target 1 and CN, placing target 2 before getting to the correct position determined underestimation whereas placing it after the correct position determined overestimation. Conversely, when target 2 was larger compared to target 1 and CN, placing target 2 before getting to the correct position determined underestimation whereas placing it after the correct position determined overestimation. (c) Positioning of target 2 when it was smaller (y-axis) or larger (x-axis) compared to target 1 or CN represented on the Cartesian plane. Crossing under- versus overestimation of the position of target 2 with relative magnitude (small vs. large) leads to different types of mapping: (d) underestimation, where target 2 is systematically underestimated regardless of its relative magnitude; (e) expansion, where target 2 is underestimated when its relative magnitude is small, whereas it is overestimated when its relative magnitude is large; (f) linear, where equidistance between target 2, target 1 and the CN is respected; (g) overestimation, where target 2 is systematically overestimated regardless of its relative magnitude; and (h) compression, where target 2 is overestimated when its relative magnitude is small, whereas it is underestimated when its relative magnitude is large.

For each participant, we computed the mean and the 95% Confidence Intervals (CIs) of the rescaled estimates of target 2 separately for the trials in which target 2 was smaller or larger compared to target 1 or the centred number (Figure 4). We assigned each child to a type of mapping according to the location of the rescaled estimates within the four quadrants and whether the 95% CIs overlapped the Cartesian axes (i.e., no error, linear positioning). We classified the mapping as linear when both the 95% CIs overlapped the two Cartesian axes and as partially linear when only one of the 95% CIs overlapped one of the Cartesian axes. When both the 95% CIs did not overlap the two Cartesian axes, we assigned the child to a mapping depending on the location of the mean rescaled estimate within the four quadrants (i.e., underestimation, overestimation, compression, expansion). Table 1 reports the classification of the mapping separately for children with partial and perfect ordinal knowledge in the DOS task. Most children displayed a linear or partially linear mapping, while some children displayed overestimation and only a few showed underestimation mapping. We found no evidence for compressed mapping (e.g., 1---5-9) or expanded mapping (i.e., 1---5------9). Notably, the distribution of mappings was similar between partial knowers and knowers ($\chi^2(3)=0.68$, $p=0.89$, $BF_{10}=0.023$). Children tended to place target 2 close to its correct location (i.e., equidistance, linear mapping) with a tendency to slightly underestimate or clearly overestimate the position of target 2 regardless of its relative magnitude (for similar results, see Sella, Lucangeli, & Zorzi, 2018).

The pattern of underestimation might be due to children moving target 2 and placing it before getting to the correct location on the line (i.e., when equidistance is achieved). Some children probably thought that once the numbers were ordered, the task was over, despite the explicit instruction to place the numbers in their correct position on the line. Overestimation, instead, might reflect children's tendency to place target 2 at the end of the line, as in all-to-the-right and all-to-the-left positioning. Such behaviour might reflect a certain degree of carelessness when accomplishing the task as if children overlooked the accuracy in positioning target 2 once the three numbers were placed in order. Crucially, children with a perfect ordinal knowledge still displayed a similar pattern of errors in the DOS task, which is characterised by a tendency for overestimation rather than compression.

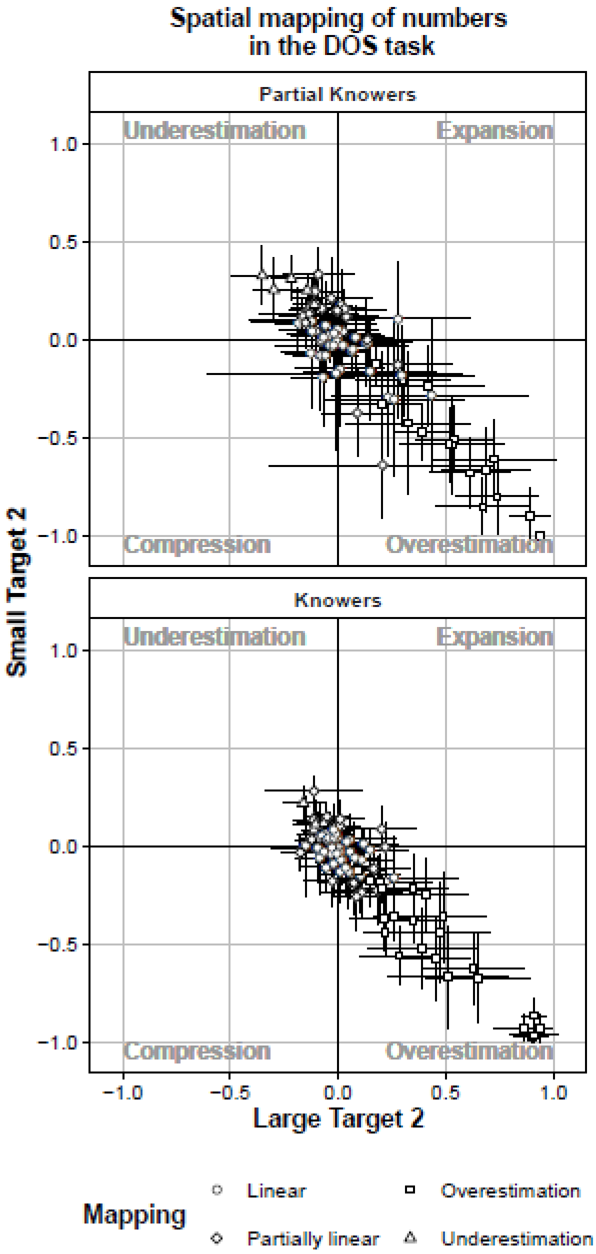

Figure 4. The individual rescaled estimates of target 2 positioning for trials in which target 2 was smaller (y-axis) or larger (x-axis) compared to target 1 or the centred number separately for partial knowers (top panel) and knowers (bottom panel). The error bars represent 95% CIs. Children’s mapping was classified as linear (dots), partially linear (diamonds), overestimation (squares), underestimation (triangles).

	Ordinal knowledge		
Mapping	Partial knowers	Knowers	Total
Linear	32	38	70

Partially linear	16	19	35
Overestimation	16	23	39
Underestimation	7	6	15

Table 1. Classification of children's mapping in the DOS task according to their ordinal knowledge (i.e., partial knowers and knowers). Cell values represent number of children.

Arithmetic fluency and the spatial mapping of numbers

We analysed the relation between arithmetic fluency and performance in the NL task and the DOS task while controlling for number comparison accuracy, domain-general cognitive measures, age, and school grade. We recalculated the latter measure to account for the fact that the study was conducted over an extended period (from November to May in the same school year). We computed the difference in months between the beginning of the study, the end of November, and the date when a child took part. For instance, a child who completed the testing sessions at the beginning of January had a value of 1 because only an entire month passed from the end of November. We calculated the adjusted grade as the difference in months (range:0-5) plus 12 months when children were second graders. The domain-general measures comprised the number of correct responses in Raven's matrices, the proportion of correct responses in the Corsi block and digit span task. We used the absolute error (AE=|estimate-target|) for all target numbers, R^2 of the linear fit (i.e., linearity), the λ coefficients of the mixed log-linear model, the β and W coefficients of the proportional cycle model as indexes of the performance in the NL task. For the DOS task, we used spatial order (i.e., the proportion of correctly ordered triplets irrespective of the direction) and the proportion of absolute error for correctly ordered trials as performance indexes. For the number comparison task, we first removed the responses below 200ms (1 trial) and then we calculated the proportion of correct responses for each child. Finally, the number of correct responses in the Tempo Test constituted the index of arithmetic fluency (i.e., outcome variable). The sample size (N=157) provides 90% power to detect a change of .10 in R^2 considering a model with ten predictors, whereby five predictors are tested (80% power to detect a change of .08 in R^2; Faul, Erdfelder, Buchner, & Lang, 2009). Table 2 reports the descriptive statistics and the zero-order correlations between the variables included in the regression analysis.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Age (months)	87.19	7.19													
2. Grade (adjusted)	9.78	6.54	.87												
3. Raven's matrices	22.63	5.35	.40	.40											
4. Corsi block	0.35	0.13	.44	.41	.37										
5. Digit span	0.30	0.08	.27	.24	.38	.31									
6. NL: AE	18.32	11.43	-.35	-.35	-.42	-.34	-.34								
7. NL: Linearity (R^2 of the linear fit)	0.68	0.28	.42	.44	.40	.40	.38	-.83							
8. NL: λ	0.71	0.32	-.39	-.42	-.27	-.32	-.29	.66	-.71						
9. NL: β	0.45	0.19	.34	.39	.28	.30	.28	-.82	.72	-.87					
10. NL: *W*	121.32	31.33	*-.17*	*-.13*	*-.01*	*-.03*	*-.12*	*.10*	*-.09*	*-.02*	.26				
11. DOS: Order	0.94	0.13	.32	.33	.41	.35	*.21*	-.54	.64	-.34	.41	*-.03*			
12. DOS: Proportion of absolute error	0.30	0.20	*-.23*	*-.20*	*-.14*	*-.20*	*-.00*	*.10*	*-.18*	.26	*-.23*	*-.03*	*-.10*		
13. Number comparison: accuracy	0.95	0.10	.26	.31	.31	.36	*.21*	-.34	.48	-.24	.25	*-.06*	.65	*-.13*	
14.Tempo Test	25.75	12.50	.70	.74	.47	.47	.42	-.51	.63	-.57	.53	*-.04*	.54	*-.22*	.47

Table 2. Descriptive statistics and Pearson's correlations. For all the correlations $BF_{10}>10$, except for those in *Italic* ($1/10<BF_{10}<10$).

First, we entered age (in months), the adjusted grade, the number of correct responses in Raven's matrices, the accuracy in the digit span task and the Corsi block into the model. The model including the adjusted grade and the accuracy in the digit span task had the strongest evidence. Therefore, we considered this model as the domain-general model in the regression analysis (DGM; Table 3).

Then, we compared the DGM model with the models additionally including the order and the proportion of error in the DOS task, individually and simultaneously. Such comparison selected the indexes of the DOS task that strongly related to arithmetic fluency, beyond domain-general measures. The model including spatial order yielded the highest evidence (DGM + DOS). Similarly, the DGM model was compared with the models additionally including the AE, the R^2 of the linear fit, the λ of the mixed log-linear model, the β and the W of the one-cycle unbounded proportional model of the NL task. The model including the R^2 of the linear fit yielded the highest evidence (DGM + NL). Then, we assessed the combined contribution of the spatial order in the DOS task and the R^2 of the linear fit in the NL task. There was strong evidence supporting the model including both variables (DGM + DOS + NL) compared to the model only including the R^2 of the linear fit (DGM + NL; BF=23.27) and compared to the model only including order in the DOS task (DGM + DOS; BF=12.87). Finally, we ensured that the relation between the two mapping tasks and arithmetic fluency was related to the spatial component rather than symbolic magnitude knowledge as measured by the accuracy in the number comparison task (NC), which related to arithmetic fluency above domain-general measures (DGM + NC, BF_{10}=1.31x10^{32}; DGM + NC vs DGM, BF=1762). The model with the domain-general variables, the order in the DOS task and the linearity in the NL task (DGM + DOS + NL) not only provided stronger evidence compared to the model additionally including the accuracy in the number comparison task (DGM + DOS + NL + NC, BF_{10}=6.15x10^{35}) but was also the best model overall (i.e., highest BF_{10} among the models containing all the possible combinations of the assessed variables, see Table 2).

For all the regression models reported in Table 3: residuals were normally distributed (Shapiro tests p>.05); multicollinearity was absent (all variance inflation factors were lower than 2 for the models with two or more predictors); heteroscedasticity was absent (Breusch–Pagan tests p>0.5); no influential observations were found (all Cook's distances were below or equal 1).

	Dependent variable: Tempo test (Arithmetic fluency)			
	Domain-general model (DGM)	DGM + DOS	DGM + NL	DGM + DOS + NL
Grade	1.30 (1.10, 1.49)	1.13 (0.95, 1.31)	1.06 (0.87, 1.25)	1.04 (0.86, 1.23)
Digit span	40.47 (24.41, 56.53)	33.92 (19.40, 48.44)	25.33 (10.11, 40.55)	26.87 (12.12, 41.62)
DOS: Order		30.15 (20.71, 39.59)		19.48 (8.22, 30.74)
NL: Linearity			14.73 (10.03, 19.43)	9.11 (3.53, 14.70)
BF_{10}	7.43×10^{28}	2.07×10^{35}	1.14×10^{35}	2.66×10^{36}
R^2	0.61	0.69	0.69	0.71

Table 3. Unstandardized regression coefficients and 95% confidence intervals. BF_{10}=Bayes factor in favour of H1.

Discussion

The biased log-like pattern of estimates in the NL task has been linked to the overlap between noisy representations of numbers (Siegler & Opfer, 2003), the lack of familiarity with the numbers in the presented numerical interval (Ebersbach, 2016; Ebersbach et al., 2008; Moeller et al., 2009), and the implementation of proportional judgements (Barth & Paladino, 2011). We investigated this issue by asking children to complete the NL task with the 0-100 interval as well as a task capable of disentangling direction, order and accuracy of the spatial mapping of numbers on the visual line (DOS task; Sella et al., 2018) while mirroring the structure of the NL task. The DOS task should reveal children's familiarity with the numerical interval and also assess the presence of a biased log-like mapping when ordinal knowledge is taken into account.

First, we fitted the estimates using mixed log-linear, bilinear, and proportional models. The bilinear model provided the best fit of the estimates in the NL task (Ebersbach et al., 2008) compared to the mixed log-linear model (Anobile et al., 2012; Opfer et al., 2016), and the proportional model (Barth & Paladino, 2011; Spence, 1990). However, fewer than half of the participants displayed a significant difference between the two slopes of the bilinear fit. In this light, the two slopes of the bilinear fit might provide more flexibility to fit the estimates, but a real separation between familiar and unfamiliar numbers is not always present.

Second, we explored how ordinal knowledge of numbers in the DOS task related to the compressed (log-like) pattern of estimates in the NL task. If familiarity and knowledge of the numerical interval affect the mapping in the NL task, one would expect children with partial ordinal knowledge (i.e., partial-knowers) to display larger logarithmic component (λ) of the mixed log-linear model, smaller power component (β) of the proportional model, significant differences between regression slopes and smaller breakpoints in the bilinear fit compared to children with perfect ordinal knowledge (i.e., knowers). Partial knowers in the DOS task were more likely to display a significant difference between the slopes of the bilinear fit; however, they had larger, rather than smaller, breakpoints compared to knowers. Partial knowers also displayed a more biased log-like mapping in the NL task, as assessed by the mixed log-linear model (i.e., larger lambda), compared to knowers. Conversely, the proportional model returned anecdotal evidence for a similar power component (β) and moderate evidence for a similar estimated magnitude of the line (W) for partial knowers and knowers.

Third, the observed compressed mapping in the NL task did not transfer to the DOS task (Sella et al., 2018). The presentation of a new visual line for each target number in the NL task makes it challenging to disentangle whether children know the ordinal relation between numbers (Sella et al., 2018; Sullivan & Barner, 2014). The DOS task, instead, assesses ordinal knowledge by asking children to sequentially position triplets of numbers while mirroring the structure of the NL task. In case of inaccurate mapping in the DOS task, estimates were biased toward an overestimation pattern, that is, children placed target 2 at the end of the line, as in all-to-the-right and all-to-the-left positioning. The pattern of overestimation might be due to carelessness when completing the task.

Finally, both order in the DOS task and linearity in the NL task related to arithmetic skills after controlling for education and domain-general cognitive factors. Spatial order in the DOS task might underpin a basic knowledge of the number sequence (Sella et al., 2018), whereas linearity reflects a better understanding of the target numbers within a large numerical interval, which in turn leads to the better arithmetic competence (Kim & Opfer, 2017).

These results provide a complex picture of the interplay between different components related to the spatial mapping of numbers. Though all statistical models of the NL task adequately fitted the biased estimates, systematic examination of the link between models' parameters and children's ordinal knowledge revealed a more complex and varied picture.

The bilinear model was somewhat superior in fitting the NL estimates and children with partial ordinal knowledge were more likely to show a significant difference between the two regression slopes. However, among children who showed a significant difference between slopes, partial knowers tended to have larger breakpoints, which does not match the prediction of the familiarity model. Therefore, the bilinear model seems to benefit from the flexibility associated with fitting two regression lines, and it appears capable of separating children displaying linear vs biased mapping. Still, the estimated breakpoints do not align with the theoretical predictions. One possible caveat is that the present results might be a byproduct of the specific fitting algorithm and the constraints applied to the fit (i.e., first slope larger than the second one, a significant difference between slopes); this issue might be a topic of further investigation.

We observed a strong correlation between the respective model parameters (i.e., lambda coefficient vs power exponent beta). However, the mixed log-linear model did a better job in separating individuals with partial and perfect ordinal knowledge in the DOS task, thereby showing convergent theoretical validity with the DOS task. We suggest that this type of model evaluation is more promising than a strict model comparison in terms of statistical fitness.

An open question regards the reason why the log-like compressed mapping of the NL task did not transfer to the DOS task. The discrepancy between the two tasks might stem from the fact that the DOS task requires to focus on a limited portion of the number line, whereas the NL task entails attending to the entire range of numbers that defines the interval. The NL and the DOS task would prompt children to produce different patterns of responses, as observed in other spatial mapping tasks (Cohen & Blanc-Goldhammer, 2011), neither of which underpins the internal representation of numbers (Cohen & Quinlan, 2018). Moreover, the frequent pattern of overestimation in the case of inaccurate positioning in the DOS task may reflect children's tendency to move the mouse cursor as far as possible, in a sort of all-to-the-left and all-to-the-right movements, signalling carelessness in positioning numbers.

Other additional issues related to the DOS task are worth considering. In the DOS and number comparison task, the larger numbers had both the decade and the unit digits that were larger compared to the smaller numbers (e.g., 41-53-65; 41 vs 53). This congruence might have facilitated the accomplishment of the two tasks as children could achieve a correct response by focusing exclusively on the units or decades without considering the whole number. However, spatial order in the DOS task and accuracy in the number comparison task correlated with arithmetic fluency while controlling for domain-general measures. Therefore, the performance in the two tasks most likely reflect numerical knowledge rather than the implementation of a simple strategy.

Why did the precision in placing numbers (i.e., the proportion of absolute error) in the DOS task not correlate with other numerical skills and arithmetic fluency whereas linearity in the NL task did? A child committed to achieving a precise spacing in the DOS task had to translate the numerical difference between the numbers of the triplet into the spatial distance on the line. The numerical difference was four for small triplets and twelve for large triplets. Therefore, the application of a simple addition and subtraction strategy (e.g., 5-1=9-5) would have been sufficient to achieve a linear spacing (i.e., low absolute error). Conversely, the NL task requires an understanding of the relative position of the target number within the entire numerical interval. Again, the DOS sampled some numbers from different portions of the number line, whereas the NL task entailed a more global understanding of the numerical interval. Moreover, children with and without perfect ordinal knowledge in the DOS task displayed a similar precision in spacing numbers and similar distribution of type of mapping (i.e., mainly partially linear, linear, and overestimation). Therefore, as ordinal knowledge is taken into account, the DOS task failed to differentiate children based on their spacing ability. Conversely, the NL task efficiently highlights the variability in mapping precision also among those children with perfect ordinal knowledge in the DOS task. In this vein, the NL task is a superior tool to differentiate children according to their arithmetic skills.

We speculate that the shift from biased to linear mapping in the NL task reflects the presence of three qualitatively different, but partially overlapping developmental stages. In the first stage, children's knowledge of numbers, especially their ordinality, is limited. This lack of knowledge leads to a biased mapping as children unorderly place unknown large numbers in a section of the line past the segment reserved to small numbers (Moeller et al., 2009). Then, children progressively extend their numerical knowledge to the entire numerical interval but still are not able to rescale the target numbers within a large numerical interval. Finally, children can accurately space numbers on the whole line by implementing appropriate mapping strategies, as an index of advanced arithmetic knowledge (Barth & Paladino, 2011; Link et al., 2014; Rouder & Geary, 2014; Sella et al., 2016; Sullivan et al., 2011).

In conclusion, we exploited a spatial mapping task that mirrors the structure of the NL task to show that children with a biased (log-like) mapping have reduced ordinal knowledge. The linearity in the NL task reliably predicted arithmetic performance even when controlling for both cardinal (number comparison) and ordinal knowledge (DOS) of numbers. These results further highlight the value of the NL task as a simple and easy-to-administer assessment tool (for a review, Schneider et al., 2018).

References

Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*(3), 454–459. https://doi.org/10.1016/j.cognition.2011.11.006

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Developmental Science*, *14*(1), 125–135. https://doi.org/10.1111/j.1467-7687.2010.00962.x

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*(2), 545–551. https://doi.org/10.1037/a0017887

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*(1), 189–201. https://doi.org/10.1037/0012-1649.41.6.189

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

Cohen, D. J., Blanc-Goldhammer, D., & Quinlan, P. T. (2018). A Mathematical Model of How People Solve Most Variants of the Number-Line Task. *Cognitive Science*, 1–27. https://doi.org/10.1111/cogs.12698

Cohen, D. J., & Quinlan, P. T. (2017). The log–linear response function of the bounded number-line task is unrelated to the psychological representation of quantity. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1290-z

Cohen, D. J., & Quinlan, P. T. (2018). The log–linear response function of the bounded number-line task is unrelated to the psychological representation of quantity. *Psychonomic Bulletin and Review*, *25*(1), 447–454. https://doi.org/10.3758/s13423-017-1290-z

Cohen, D. J., & Sarnecka, B. W. (2014). Children's Number-Line Estimation Shows Development of Measurement Skills (Not Number Representations). *Developmental Psychology*. https://doi.org/10.1037/a0035901

Dackermann, T., Huber, S., Bahnmueller, J., Nuerk, H.-C., & Moeller, K. (2015). An integration of competing accounts on children's number line estimation. *Frontiers in Psychology*, *6*(July), 1–7. https://doi.org/10.3389/fpsyg.2015.00884

De Vos, T. (1992). *Tempo-Test-Rekenen. Handleiding.[Tempo Test Arithmetic. Manual]*. Nijmegen: Berkhout.

Dehaene, S. (1992). Varieties Of Numerical Abilities. *Cognition*, *44*, 1–42.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12691758

Ebersbach, M. (2013). Evidence for a Spatial-Numerical Association in Kindergartners Using a Number Line Task. *Journal of Cognition and Development*, (March), null-null. https://doi.org/10.1080/15248372.2013.805134

Ebersbach, M. (2016). Development of Children's Estimation Skills: The Ambiguous Role of Their Familiarity With Numerals. *Child Development Perspectives*, *10*(2), 116–121. https://doi.org/10.1111/cdep.12172

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), 1–17. https://doi.org/10.1016/j.jecp.2007.08.006

Ebersbach, M., Luwel, K., & Verschaffel, L. (2015). The Relationship Between Children ' s Familiarity with Numbers and Their Performance in Bounded and Unbounded Number Line Estimations The Relationship Between Children ' s Familiarity with Numbers and Their Performance in Bounded and. *Mathematical Thinking and Learning*, *17*(2–3), 136–154. https://doi.org/10.1080/10986065.2015.1016813

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., & Van Lieshout, E. C. D. M. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology*, *134*, 12–29. https://doi.org/10.1016/j.jecp.2015.02.002

Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: the cyclical power model.

Psychological Review, *107*(3), 500–524. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10941278

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Kaps, M., & Lamberson, W. R. (2004). Biostatistics for animal science. CABI Pub.

Kim, D., & Opfer, J. E. (2017). A unified framework for bounded and unbounded numerical estimation. Developmental Psychology, 53(6), 1088.

Link, T., Nuerk, H. C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *Quarterly Journal of Experimental Psychology*, *67*(8), 1597–1613. https://doi.org/10.1080/17470218.2014.892517

Mammarella, I. C., Toso, C., Pazzaglia, F., & Cornoldi, C. (2008). *BVS-Corsi. Batteria per la valutazione della memoria visiva e spaziale.Edizioni Erickson.* Trento: Erickson.

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2009). Children's early mental number line: logarithmic or decomposed linear? *Journal of Experimental Child Psychology*, *103*(4), 503–515. https://doi.org/10.1016/j.jecp.2009.02.006

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-2.

Muggeo, V. M. R. (2014). segmented : An R Package to Fit Regression Models with Broken-Line Relationships, (September).

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, *55*(3), 169–195. https://doi.org/10.1016/j.cogpsych.2006.09.002

Opfer, J. E., Siegler, R. S., & Young, C. J. (2011). The powers of noise-fitting: reply to Barth and Paladino. *Developmental Science*, *14*(5), 1194–1204; discussion 1205-6. https://doi.org/10.1111/j.1467-7687.2011.01070.x

Opfer, J. E., Thompson, C. A., & Kim, D. (2016). Free versus anchored numerical estimation: A unified approach. *Cognition*, *149*, 11–17. https://doi.org/10.1016/j.cognition.2015.11.015

Peeters, D., Degrande, T., Ebersbach, M., & Verschaffel, L. (2016). Children ' s use of number line estimation strategies, 117–134. https://doi.org/10.1007/s10212-015-0251-z

Peeters, D., Verschaffel, L., & Luwel, K. (2017). Benchmark-based strategies in whole number line estimation, 668–686. https://doi.org/10.1111/bjop.12233

Psychology Software Tools, I. (2012). E-Prime 2.0. Pittsburgh.

R Core Team. (2016). R Development Core Team. *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.r-project.org/

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163.

Raven, J. C. (1998). *Raven's Coloured Progressive Matrices.* London: Pearson Education.

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science*, *17*(4), 525–536. https://doi.org/10.1111/desc.12166

Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of Number Line Estimation With Mathematical Competence: A Meta-analysis. *Child Development*, *00*(0), 1–18. https://doi.org/10.1111/cdev.13068

Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2015). Varieties of Quantity Estimation in Children. *Developmental Psychology*, *51*(6), 758–770. https://doi.org/10.1037/a0039183

Sella, F., Lucangeli, D., & Zorzi, M. (2018). Spatial order relates to the exact numerical magnitude of digits in young children. *Journal of Experimental Child Psychology*, 1–20. https://doi.org/10.1016/j.jecp.2018.09.001

Sella, F., Sader, E., Lolliot, S., & Cohen Kadosh, R. (2016). Basic and advanced numerical performances relate to mathematical expertise but are fully mediated by visuospatial skills. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(9). https://doi.org/10.1037/xlm0000249

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*(2), 428–444. https://doi.org/10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237–243.

Siegler, R. S., Thompson, C. a., & Opfer, J. E. (2009). The Logarithmic-To-Linear Shift: One Learning Sequence, Many Tasks, Many Time Scales. *Mind, Brain, and Education*, *3*(3), 143–150. https://doi.org/10.1111/j.1751-228X.2009.01064.x

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology. General*, *142*(1), 193–208. https://doi.org/10.1037/a0028560

Spence, I. (1990). Visual Psychophysics of Simple Graphical Elements. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(4), 683–692.

Sullivan, J., & Barner, D. (2014). The development of structural analogy in number-line estimation. *Journal of Experimental Child Psychology*, *128*, 171–189. https://doi.org/10.1016/j.jecp.2014.07.004

Sullivan, J., Juhasz, B. J., Slattery, T. J., & Barth, H. C. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review*, 557–563. https://doi.org/10.3758/s13423-011-0081-1

White, S. L. J., & 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 : BBF*, *8*(1), 1. https://doi.org/10.1186/1744-9081-8-1

Supplementary Materials

[Insert Figure S1 here]

Figure S1. Individual estimates as a function of target numbers with the mixed log-linear (solid line), proportional one-cycle unbounded (dotted line), and the bilinear (long dashed line) fit. The label at the bottom of each plot indicates the participant's grade (i.e., 1 or 2), whether they had partial or perfect ordinal knowledge in the DOS task (i.e., "Partial knower", "Knower"), and the fit that provided the lowest BIC or whether the mapping was non-numerical ("None"; i.e., no significant linear or logarithmic model or negative linear slope).