BJOP | 12335 | WILEY | Dispatch: 10.7.18 | CE: Kowsalya J
Journal Code | Manuscript No. | | No. of pages: 20 | PE: Pravin Kumar S.

British Journal of Psychology (2018)

The British Psychological Society
www.wileyonlinelibrary.com

Number line estimation in highly math-anxious individuals

María Isabel Núñez-Peña[1,2,3]*, Àngels Colomé[2,4] and David Aguilar-Lleyda[2,4]

[1]Department of Social Psychology and Quantitative Psychology, Faculty of Psychology, University of Barcelona, Spain
[2]Institute of Neurosciences, University of Barcelona, Spain
[3]Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain
[4]Department of Cognition, Development and Educational Psychology, Faculty of Psychology, University of Barcelona, Spain

In this study, we aimed to investigate the difficulties highly math-anxious individuals (HMA) may face when having to estimate a number's position in a number line task. Twenty-four HMA and 24 low math-anxiety (LMA) individuals were presented with four lines with endpoints 0–100, 0–1,000, 0–100,000, and 267–367 on a computer monitor on which they had to mark the correct position of target numbers using the mouse. Although no differences were found between groups in the frequency of their best-fit model, which was linear for all lines, the analysis of slopes and intercepts for the linear model showed that the two groups differed in performance on the less familiar lines (267–367 and 0–100,000). Lower values for the slope and higher values for the intercept were found in the HMA group, suggesting that they tended to overestimate small numbers and underestimate large numbers on these non-familiar lines. Percentage absolute error analyses confirmed that HMA individuals were less accurate than their LMA counterparts on these lines, although no group differences were found in response time. These results indicate that math anxiety is related to worse performance only in the less familiar and more difficult number line tasks. Therefore, our data challenge the idea that HMA individuals might have less precise numerical representations and support the anxiety–complexity effect posited by Ashcraft and colleagues.

Math anxiety, defined as 'feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations' (Richardson & Suinn, 1972, p. 551), is a subject of increasing interest, as shown by the large number of reviews on this topic published in recent years (e.g. Chang & Beilock, 2016; Dowker, Sarkar, & Looi, 2016; Foley *et al.*, 2017; Suárez-Pellicioni, Núñez-Peña, & Colomé, 2016). This interest is fuelled by the fact that math anxiety is a global phenomenon with a high prevalence. According to the 2012 PISA report (Organization for Economic Co-operation and Development (OCDE), 2013), on average 30% of 15-year-old students from OECD countries reported feeling incapable or

*Correspondence should be addressed to María Isabel Núñez-Peña, Department of Social Psychology and Quantitative Psychology, Faculty of Psychology, University of Barcelona, Passeig Vall d'Hebron, 171, 08035 Barcelona, Spain (email: inunez@ub.edu).

DOI:10.1111/bjop.12335

nervous when solving a math problem and 59% reported being worried about the difficulty of math classes. Highly math-anxious (hereinafter, HMA) individuals have lower levels of math performance than their low math-anxiety (hereinafter, LMA) peers (Ashcraft & Krause, 2007). They often avoid mathematical activities and are poorly represented in the science, technology, engineering, and math (STEM) fields, where developed countries require well-prepared citizens.

In this context, an increasing number of studies have been devoted to identifying the cognitive factors that might play a role in the difficulties experienced by HMA when facing math activities. Such knowledge would be useful for designing math-anxiety prevention programmes or interventions that could help HMA individuals to overcome these difficulties. Three main proposals have been put forward to date: namely that HMA might (1) have fewer working memory resources, (2) a less precise representation of magnitude, or (3) an inhibition/attentional-control deficit (Hopko, Mcneil, Gleason, & Rabalais, 2002; Suárez-Pellicioni *et al.*, 2014). Given that this study aims to discriminate between the first two proposals, we will now describe them briefly. First, Ashcraft, Kirk, & Hopko, (2000; Ashcraft & Kirk, 2001; Ashcraft & Krause, 2007) suggested that math anxiety causes a decrease in working memory capacity (WM) when HMA individuals are performing math tasks. Ashcraft and Faust (1994) found that HMA and LMA individuals performed similarly in overlearned simple addition and multiplication tasks but that differences emerged in complex additions. They proposed the *anxiety–complexity effect*, a worsening in HMA individuals' performance when the numerical task becomes more complex (Faust, Ashcraft, & Fleck, 1996). Ashcraft and colleagues suggested that HMA memory resources could be occupied by math anxiety-related ruminations; this would be particularly relevant in complex tasks, in which HMA participants would not have enough WM resources to perform the task properly. Thus, math anxiety would act as a secondary task in a WM dual task, hindering the performance in the main mathematical task the more it required working memory resources.

Second, individuals with high math anxiety might suffer from a low-level numerical deficit, specifically a deficit in their numerical magnitude representation, which would compromise their performance in more complex math tasks. Maloney, Risko, Ansari, and Fugelsang (2010) reported that individuals with high math anxiety performed worse than their LMA peers in a task as simple as enumerating from five to nine objects. A year later, Maloney, Ansari, and Fugelsang (2011) found that HMA individuals showed a larger numerical distance effect in a comparison task on two-one-digit Arabic numbers; that is, HMA were slower than their LMA counterparts as the distance between numbers was reduced (convergent psychophysiological evidence was reported in Núñez-Peña & Suárez-Pellicioni, 2014). Consistent with this second proposal, Lindskog, Winman, and Poom (2017) claimed that HMA individuals have a poorer approximate number system (ANS) or pre-verbal number representation than their LMA peers. However, not all previous studies support this hypothesis. Dietrich, Huber, Moeller, and Klein (2015) failed to find math-anxiety effects in a dot comparison task. Furthermore, despite replicating the larger distance effect for HMA participants in a symbolic comparison, they attributed it to decisional processes rather than the acuity of magnitude representation. Last, Colomé (2018) found no differences between HMA and LMA participants in dot comparison, Arabic digit comparison, or a counting Stroop task. On this basis, further studies are required to confirm the hypothesis that math-anxious individuals have a less precise magnitude representation. The aim of this study was to test this hypothesis in relation to the one proposed by Ashcraft and colleagues.

A task widely used to study the mental representation of number magnitude is the number line estimation task (hereinafter, NLT). In this task, participants are shown a line with the beginning and endpoints marked with numbers (e.g. 0–100) and are asked to indicate the position of a number on this line by marking the appropriate location on it. Number line task performance is a reliable predictor of actual and future numerical competencies (Booth & Siegler, 2008; Link, Nuerk, & Moeller, 2014; Sasanguie, Defever, Van den Bussche, & Reynvoet, 2011; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013) and correlates with performance on other numerical estimation and magnitude comparison tasks (Crollen & Noël, 2015; Laski & Siegler, 2007; Sasanguie, De Smedt, Defever, & Reynvoet, 2012). Therefore, it has been suggested that the observed mapping in the NLT reflects the underlying mental representation of numbers (Booth & Siegler, 2006; Siegler & Booth, 2004; Siegler & Opfer, 2003) and can be used to identify deficits in it (Siegler & Booth, 2005). Developmental studies have shown a representational change from a logarithmic representation of number magnitude to a linear representation (the log-to-linear shift) with increasing age and experience (Booth & Siegler, 2006; Siegler & Booth, 2004; Siegler & Opfer, 2003). Young children initially respond by spacing smaller numbers further apart than larger numbers (logarithmic representation), but, between second (for the 0–100 range) and fourth grade (for the 0–1,000 range), their number placements become increasingly linear, with equal spacing between values (Booth & Siegler, 2006).

Nevertheless, recent studies have cast some doubts over whether the log-to-linear shift found in NLT is caused by a developmental change in the representation of numerical magnitude (Barth & Paladino, 2011; Hurst, Leigh Monahan, Heller, & Cordes, 2014; Slusser, Santiago, & Barth, 2013). According to Barth and Paladino (2011), the NLT can mainly be viewed as a proportion estimation task, where reference points can be used. For example, to mark the position of 30 on a 0–100 line, an estimate of the size of 30 relative to the total size of 100 is needed. Thus, the task requires the ability to recall the proper magnitudes associated with the relevant numerals, and its outcome will depend on the biases involved in estimating the part and the whole magnitudes and connecting the two.

Therefore, performance on this task is not just a signature of the underlying representation of number but could also be a measure of the ability to make proportion computations across the range of values presented, as well as using anchor points such as the central value, to facilitate the estimation. To study whether proportional judgements can explain performance in NLT, Barth and Paladino (2011) proposed fitting a proportion estimation model to the data. This was the cyclical power model (CPM) by Hollands and Dyre (2000), adapted from Spence's power models (Figure 1; Spence, 1990). In their experiments, Barth and Paladino fitted two variants of the CPM to children's performance: a *one-cycle power model* that predicts that individuals judge the size of the given numeral comparing the given number to both endpoints and a *two-cycle power model* that predicts that both endpoints and the middle point act as points of reference. They found that the two-cycle proportion judgement model provided the best explanation for their data on 7-year-old children, whereas the one-cycle model provided the better explanation for their 5-year-olds' estimates. Most importantly, it also provided a better fit than that of a linear or logarithmic model. Sullivan, Juhasz, Slattery, and Barth (2011) added further support to this interpretation by reporting that adults show preferential fixation on the mid-point of the line when engaging in a NLT, suggesting that they might create landmarks throughout the line, such as the halfway point. However, whether NLT performance can be understood simply as an indicator of the precision of mental representation of

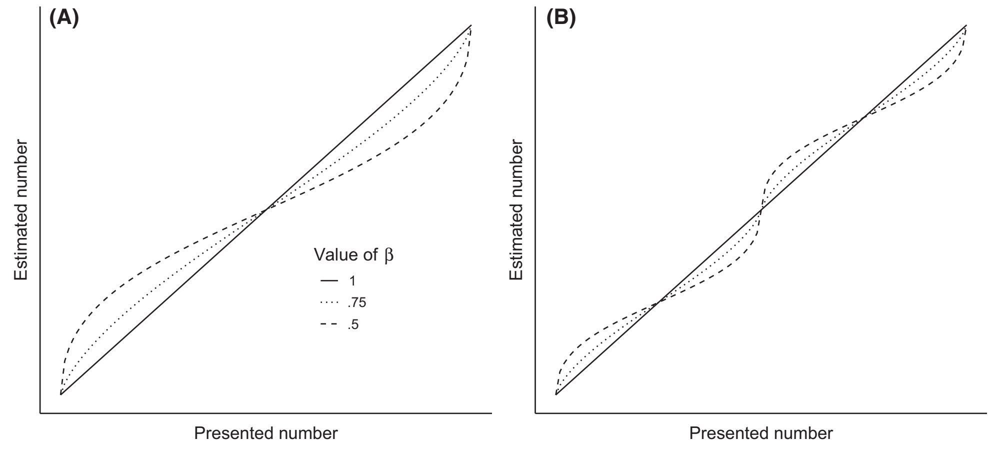

Figure 1. Example of some number line estimations predicted by the proportion estimation models. Lines represent the estimated position as a function of the presented number. (A) Estimation patterns predicted by the one-cycle model (equivalent to the power model of Spence, 1990), where participants would estimate the position of the number by taking both extremes of the line as reference points. The three lines correspond to three different β values, that is the exponent that determines the power function relating the estimated magnitude to the actual magnitude: for $\beta < 1$, the smaller the β, the larger the bias, while when $\beta = 1$, $x = y$. (B) Estimation pattern predicted by the two-cycle model (Hollands & Dyre, 2000), where participants would also use the centre of the line as a reference point. Lines correspond to the same β values used in A.

numerical magnitudes or also stems from the ability to perform proportion judgements remains under debate.[1]

Although traditionally NLT has involved number lines with standard (i.e. multiple of 10) endpoints (e.g. 0–100 or 0–1,000), recent studies have used less familiar ranges in an attempt to investigate which factors determine the participants' response pattern. Hurst *et al.* (2014) compared adults' performance in number line tasks with non-standard endpoints (endpoints 1,639 and 2,897) and with standard endpoints with similar magnitudes (2,000–3,000) or numerical range (0–1,258). All tasks involving standard endpoints resulted in a linear response, but data from the lines with non-standard endpoints were better fit by a logarithmic model. Hurst *et al.* (2014) suggested that performance in the NLT might depend on the fluency with the relative ordering of all the values in the range as well as the facility to identify standard anchors such as the sequence mid-point: Less familiar sequences or sequences with less familiar endpoints might be more cognitively demanding, leading to worse performance. In a similar vein, when using number line tasks with standard endpoints (0–1,000) and non-standard endpoints (a line with endpoints 364 and 1,364), Laski and Dulaney (2015) reported that although the linear function accounted for a greater amount of variance in adults' median estimates than the logarithmic function on both number lines, the logarithmic function fit much better on the non-standard number line than on the standard number line. They also

[1] *In both cases, performance for HMA participants might be impaired: Simms, Clayton, Cragg, Gilmore, and Johnson (2016) showed that proportional reasoning requires both good number knowledge and visuo-spatial skills that allow the participant to judge the scale of the line and divide the space into segments. Recent evidence (e.g. Ferguson, Maloney, Fugelsang, & Risko, 2015) indicates that HMA might perform worse than their LMA peers on tests in small-scale spatial skills.*

reported that estimates were less accurate on the non-standard line. These authors concluded that 'individuals possess multiple representations of numerical magnitude that may be simultaneously activated in estimation tasks and they have a tendency to increase their weighting of the logarithmic representation when confronted with difficult numerical tasks' (Laski & Dulaney, 2015, p. 1040). This would be the case for number lines with non-standard endpoints: They would prevent the use of well-known anchors and require more complex calculations, implying a higher cognitive load.

In this study, we investigated for the first time the ability of individuals with high and low math anxiety to estimate number positions on a line. Number lines of different ranges, from more to less familiar, were used here to determine whether group differences in patterns of estimations depended on the difficulty of the task: (1) two familiar number lines with standard (power of 10) endpoints (0–100 and 0–1,000; hereinafter, 100 and 1,000 lines); (2) a non-familiar number line with standard endpoints (0–100,000; hereinafter 100,000 line)[2]; and (3) a non-familiar number line with non-standard endpoints (267–367; hereinafter 367 line). Using NLT with different endpoints, we wanted to shed light on whether HMA individuals' math difficulties are better explained by the fact that they suffer from a low-level numerical deficit (as proposed by Maloney and colleagues) or that they devote their WM resources to their anxious reaction, not having enough available resources to perform complex math tasks properly (as proposed by Ashcraft and colleagues). A different pattern of results was expected according to each proposal. If HMA individuals suffer from a low-level numerical deficit (i.e. a less precise representation of numerical magnitude; Maloney *et al.*, 2010, 2011; Núñez-Peña & Suárez-Pellicioni, 2014), we would expect them to perform worse than their LMA peers on the four lines, because access to the numerical magnitude representation would be needed in all cases. However, if HMA individuals' anxious reaction depletes their WM resources, leaving insufficient resources available to perform the task properly (Ashcraft & Kirk, 2001; Ashcraft & Krause, 2007; Ashcraft *et al.*, 2000), we would expect them to have more difficulties in non-familiar number line tasks (the anxiety–complexity effect), because unfamiliar lines are expected to be more cognitively demanding (Hurst *et al.*, 2014; Laski & Dulaney, 2015).

In addition to the results from NLT, two further sets of data were collected. At the end of the experiment, participants had to rate how well they believed they had performed on each task, in order to obtain a measure of self-perceived level of task difficulty or self-efficacy. Participants also performed a control task to measure their motor precision to rule out the possibility that group differences in performance in the NLT were due to this component.

Methods

Participants

Forty-eight psychology students took part in this experiment and were divided into two equally sized groups of high and low math anxiety. Participants were selected from a larger sample of 581 students from the University of Barcelona who were assessed for math anxiety and trait anxiety (see Materials) within the framework of a longer project.

[2] *In this case, we understand non-familiar in the sense of less frequently encountered or used because of the large magnitude of the numbers involved; Dehaene and Mehler (1992) compared the frequency distribution of numbers in different languages and found a regular decrease in frequency with magnitude.*

Highly math-anxious participants (HMA) scored over the third quartile on the shortened Mathematics Anxiety Rating Scale (sMARS) (Alexander & Martray, 1989), while their low math-anxiety peers (LMA) scored below the first quartile. Despite differing in math anxiety ($t(46) = 22.97$, $p < .001$, $d = 6.63$), both groups were equivalent in age ($t(46) = 1.71$, $p < .095$), trait anxiety ($t(46) = .00$, $p = 1$), and gender distribution ($\chi^2(1) = 2.18, p < .13$). For more detailed information about the two groups, see Table 1.

Material

Screening phase

To form groups, a large sample of undergraduate students was assessed using the following two tests. Data were collected in classroom settings as part of a voluntary activity at the University of Barcelona.

Shortened Mathematics Anxiety Rating Scale (sMARS) (Alexander & Martray, 1989). The sMARS is a 25-item version of the Math Anxiety Rating Scale (MARS) (Richardson & Suinn, 1972). This instrument measures anxiety by presenting 25 situations that may cause math anxiety (e.g. *Thinking about the math exam I will have next week*). The respondent indicates the level of anxiety associated with the item using a five-point Likert scale ranging from 1 (no anxiety) to 5 (high anxiety). The sum of the item scores provides the total score for the instrument, which ranges from 25 to 125. In this study, we used the Spanish version of the sMARS (Núñez-Peña, Suárez-Pellicioni, Guilera, & Mercadé-Carranza, 2013). The scores for the Spanish version of the sMARS have shown strong internal consistency (Cronbach's alpha = .94) and high 7-week test–retest reliability (intra-class correlation coefficient = .72).

State-Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). The STAI is a 40-item scale used to measure state (STAI-S) and trait (STAI-T) anxiety, with 20 items in each. Only the STAI-T subscale, which measures a more general and relatively stable tendency to respond with anxiety, was used in this study. This subscale comprises 20 statements describing different emotions, and for each item, respondents use a four-point Likert scale (ranging from 0: almost never to 3: almost always) to indicate how they feel 'in general'. Good to excellent internal consistency (Cronbach's alpha = .95), adequate 30-day test–retest reliability with high-school

Table 1. Means and standard error of the mean (SEM; in brackets) for age, math anxiety, and trait anxiety, for the LMA and HMA groups. Number of women in each group is also given

	LMA	HMA
Age	19.92 (.25)	20.92 (.53)
Math anxiety	42.38 (1.28)	85.79 (1.39)
Trait anxiety	17.25 (1.71)	17.25 (1.47)
Number of women	16	22

Note. Math anxiety measured using the sMARS (Alexander & Martray, 1986); trait anxiety measured using the trait subscale of the STAI (Spielberger *et al.*, 1983).

students (r = .75), and 20-day test–retest reliability with college students (r = .86) have been reported for the Spanish version of this subscale (Spielberger, Gorsuch, & Lushene, 2008).

Experimental phase

Participants were asked to estimate the position of numbers on four number lines with different endpoints (0–100; 0–1,000; 0–100,000; and 267–367). Lines were centrally displayed in grey on a black screen and took up 90% of the screen width and 4% of its height. We used a 19-inch, square (4:3) CRT monitor of 85 Hz and 1,024 × 768 pixels resolution. Endpoints were identified by the corresponding numbers, which were displayed just below the lines in white, Courier New 18 characters. The number to locate (target) appeared at the left top of the screen (white, Courier New, 25 underlined).

Twenty-four targets were selected for each line (see Appendix). In the case of the 0–1,000 and 0–100,000 lines, the stimuli were the same as in Slusser *et al.* (2013) with the exception of 60,000, which was replaced by 61,305 because the former could be easier to estimate on this line. For the 0–100 line, we used the same targets as Booth and Siegler (2006). Targets for the 267–367 line were created by adding 267 to the 0–100 targets.

The control task used the same line as the NLT. A thin (2% of the screen width) red vertical strip was placed within the line. The distance between the strip and the initial cursor position ranged from 5% to 95% of the line length. Nineteen distances were presented by manipulating the difference between the strip and initial cursor positions at intervals of 5. Each distance was presented twice: In one case, the cursor had to be moved towards the right, and in the other, it had to be moved to the left.

Procedure

Each participant performed the number line estimation task for each of the four line ranges. All trials within the same line range were blocked, and the order of the ranges was counterbalanced across participants. Testing for each line began with four training trials. Targets appearing in these trials were not used in the test phase. After training, two blocks of experimental trials were presented with a half-minute pause in between. Participants had to decide where the target was located on the number line by placing the mouse cursor over the desired position and clicking the left button of the mouse. The initial position of the cursor varied randomly in each trial. The twenty-four targets for each line were randomly presented and appeared once within each block.

Each trial had the following structure. First, an asterisk appeared centred on the screen for 500 ms. After a blank interval of 100 ms, the number line and the target to be located were presented. They remained on the screen until the participant responded or for a maximum of 6,000 ms. A 500-ms interval was left between trials. The latency and the position at which the participant placed the cursor were recorded for each trial.

After the number line estimation blocks, participants were asked to perform a new task to control for their motor ability when using the mouse. In this task, a grey line like the one used in the previous task was displayed and a red vertical strip was placed within it. Participants had to move the mouse cursor and click on the red strip with the left button of the mouse. Each participant performed 38 trials, which were presented in random order. Each trial started with a fixation point that appeared centred on the screen for 500 ms. After 100 ms of blank interstimuli interval, the black line and the red strip were displayed. They disappeared with the participant's response or after a maximum of 4,000 ms if there was no response. Lastly, an interval of 500 ms was left between trials, during which no stimulus was presented. The latency and final position of the cursor were recorded in each trial.

After that, participants were asked to complete a short questionnaire in which they had to rate their own performance for each of the lines on a seven-point Likert scale, with 1 being 'not good at all' and 7 being 'very good'.

Data analysis

Several analyses were performed in this study. Firstly, we were interested in determining whether there was a significant difference in the type of number-to-line mapping between anxiety groups (HMA vs. LMA). For this, we calculated the goodness of fit of four different models (linear, logarithmic, one-cycle power, or two-cycle power) on the basis of the Akaike information criterion (Akaike, 1973, 1974) corrected for small samples (AICc, Hurvich & Tsai, 1989). The best model was identified as that with the lowest AICc value among the candidates for each participant and line range. Then, we performed a chi-square test for independence by line range, to determine whether the frequency of participants whose estimations were best explained by each model differed between groups.

Secondly, three other measures were analysed as follows: percentage of absolute error (PAE; i.e. the accuracy of participants' estimation), response times (RT), and self-reported level of efficacy (answers to the questionnaire). ANOVAs were performed for each measure separately, taking *Line* (100, 1,000, 100,000, and 367) as the within-subject factor and *Group* (HMA and LMA) as the between-subjects factors. The Greenhouse–Geisser epsilon (ε) correction for sphericity departures (Geisser & Greenhouse, 1958) was used in ANOVAs whenever necessary. The F value, the uncorrected degrees of freedom, the probability level following correction, the ε value (when appropriate), and the η_p^2 effect size index are presented. *Post-hoc* comparisons were performed by means of t-tests (either for independent or repeated measures depending on the factor analysed), and the Hochberg approach was used to control for the increase in type I error (Keselman, 1998). The t-value, the degrees of freedom, the p-value, and Cohen's d index for effect size were calculated. Only significant effects ($p \leq .05$) are reported.

Finally, differences between groups in RT and accuracy in the control task were studied by means of independent t-tests.

Results

Model adjustment

Figure 2 shows the median estimates as a function of their corresponding unbiased number, pooled across participants within each group, with each panel containing the responses for one of the four lines (100, 1,000, 100,000, and 367). The black line in each panel shows how an unbiased mapping would appear. Deviations from this diagonal line represent estimation errors. This figure shows that both groups are highly accurate and have a high degree of linearity for the familiar lines, but that their estimations are less accurate for the non-familiar ones. The figure for the 367 line also shows differences between groups in the slope and intercept of the linear model.

A chi-square test for independence was carried out by line to study whether the frequency of the best-fitting model was related to math anxiety. The logarithmic model

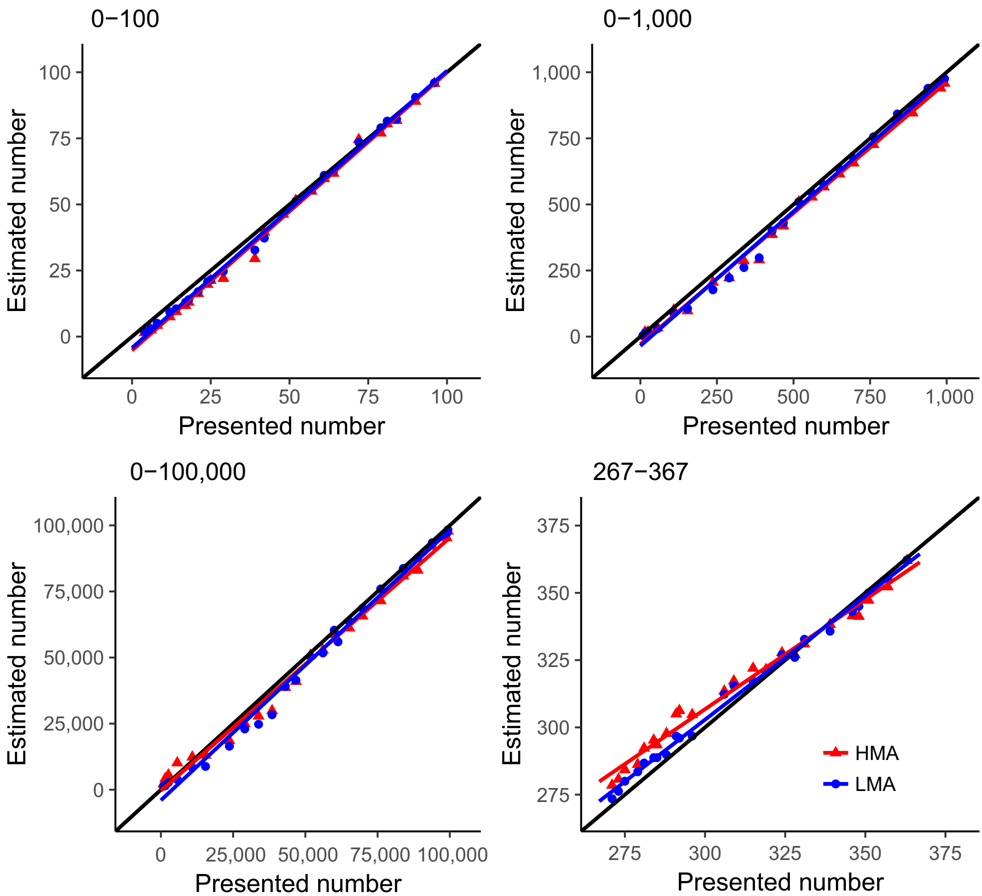

Figure 2. Median estimated position of the numbers presented in the experiment, as a function of their unbiased position. Each of the four panels depicts the data for one of the number lines that were used. Medians are shown for both anxiety groups (HMA and LMA). Each corresponding line shows the fit provided by the best-fitting model, which is linear in all cases for group estimates. Black identity lines show unbiased estimations.

was discarded because none of the participants fitted this model best. The results showed no relation between group and best-fitting model for any line (all $p > .05$). To determine whether there was a model that fit the data best for each line, we performed chi-square tests and found differences between best-fitting model frequencies for the four lines ($\chi^2(2) = 35.37$, $p < .001$; $\chi^2(2) = 27.87$, $p < .001$; $\chi^2(2) = 12.12$, $p = .002$; and $\chi^2(2) = 4.08$, $p = .043$, for the 100, 1,000, 100,000, and 367 lines, respectively). To further investigate these differences, paired comparisons were performed by means of the binomial distribution. Importantly, the linear model was a more frequent best fit than both cycle models for the 100, 1,000, and 100,000 lines (see Table 2). For the 367 line, the linear model fit was also more frequent than the two-cycle model and tended to be more frequent than the one-cycle model. Finally, when both cycle models were compared, no difference was found between the frequencies of best-fit model (all $p > .05$), except for the 367 line, where the one-cycle power model was a more frequent best-fit than the two-cycle power model.

Table 2. Paired comparisons of frequencies for the best-fit model in the four lines. Frequencies (freq) for each model for every line and probability (p) associated with each comparison are given

	100 line		1,000 line		100,000 line		367 line	
Comparison	Freq	p	Freq	p	Freq	p	Freq	p
Linear	35	<.001	33	.001	27	.002	31	.059
One cycle	10		10		8		17	
Linear	35	<.001	33	<.001	27	.038	31	<.001
Two cycle	3		5		13		0	
One cycle	10	.092	10	.302	13	.383	17	<.001
Two cycle	3		5		8		0	

Although the above analysis showed that the linear model was the best fit for all lines, visual inspection of Figure 2 indicated potential differences between groups and lines with regard to their parameters in this model. We therefore decided to study possible differences between groups in terms of their linear fit and compared the slope and intercept values of participants' linear models by means of independent t-tests.[3] There were no differences between groups for the familiar lines; however, differences emerged for the non-familiar ones. For the 367 line, the slope was larger for the LMA than for the HMA group ($t(46) = 2.54$, $p = .015$, $d = .73$) and the intercept was smaller for the LMA than for the HMA group ($t(46) = 2.69$, $p = .010$, $d = .78$). As for the 100,000 line, the slope was larger for the LMA than for the HMA group ($t(46) = 3.93$, $p < .001$, $d = 1.13$). Due to the fact that groups had intercepts with different signs in the 100,000 line, absolute values of deviations from 0 were calculated for each participant and group differences were studied by means of independent t-tests. Groups did not differ in their intercept deviations in the 100,000 line.

We then performed analyses to examine how much the slope and the intercept deviated from the perfect linear mapping (i.e. slope = 1 and intercept = 0) in each of the groups in the non-familiar lines. Results revealed that the slopes differed significantly from 1 for the 367 line in both groups ($t(23) = 5.09$, $p < .001$, $d = 1.04$ for the LMA group and $t(23) = 7.43$, $p < .001$, $d = 1.52$ for the HMA group) and for the 100,000 line in the HMA group ($t(23) = -4.82$, $p < .001$, $d = .98$). As for the intercepts, the analysis of their deviations from 0 revealed significant differences in the 367 line for the LMA ($t(23) = 5.52$, $p < .001$, $d = 1.13$) and the HMA groups ($t(23) = 7.58$, $p < .001$, $d = 1.55$). As for the 100,000 line, differences were significant for the LMA group ($t(23) = 3.96$, $p = .001$, $d = .81$) and marginal for the HMA group ($t(23) = 1.72$, $p = .098$, $d = .35$). Table 3 shows mean and standard errors of the mean for the slopes and intercepts for each group for the four lines.

Percentage absolute error (PAE)

As we saw above, the fact that data fit a linear model means that participants' estimates were linearly spread along the number line, but this does not necessarily mean that they had answered flawlessly (Simms *et al.*, 2016). Therefore, we decided to calculate the

[3] *This analysis was also performed only with participants whose best fit was linear for each line. The results showed the same pattern, although some differences were only marginally significant (intercept in the 100,000 line and slope in the 367 line).*

Table 3. Means and standard error of the mean (in brackets) for slopes and intercepts for the LMA and HMA groups for the four lines

	100 line		1,000 line		100,000 line		367 line	
	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept
LMA	1.04 (.01)	−4.02 (.49)	1.00 (.01)	−30.65 (4.14)	.99 (.01)	−2,259 (571)	.87 (.02)	39.93 (7.23)
HMA	1.05 (.01)	−5.01 (.63)	.97 (.02)	−21.38 (10.44)	.89 (.02)	2,884 (1,675)	.78 (.03)	72.22 (9.53)

Note. LMA: low math-anxiety group; HMA: high math-anxiety group.

medians of PAE per participant for each line as a measure of estimation accuracy. Medians were used rather than means to minimize the effect of outliers. PAE was calculated using the formula by Siegler and Booth (2004) as the absolute distance between the actual and estimated positions of numbers on the line divided by the scale of the line multiplied by 100:

$$\mathrm{PAE} = \frac{|\text{ estimated position} - \text{actual position }|}{\text{scale of the line}} \times 100$$

For example, if a participant was asked to estimate the position of 39 on a 0–100 number line and placed the mark at the position of 30 on the line, the PAE would be $(|39 - 30|/100) \times 100$ or 9%.

The overall ANOVA revealed that the main effects of *Line* ($F(3,138) = 52.26, p < .001$, $\varepsilon = .53$, $\eta_p^2 = .55$) and *Group* ($F(1,46) = 12.25$, $p = .001$, $\eta_p^2 = .21$), as well as the interaction *Line* × *Group* ($F(3,138) = 5.27, p = .011$, $\varepsilon = .53$, $\eta_p^2 = .10$), were statistically significant. To study this interaction in more detail, two separate analyses were performed. First, groups were compared for each line by means of independent *t*-tests, and second, the effect of *Line* was analysed for each group by means of ANOVA, taking *Line* as the within-subject factor. The first analysis showed that HMA were less accurate than their LMA peers for both the 100,000 line ($t(46) = 2.42, p = .019, d = .70$) and the 367 line ($t(46) = 3.28, p = .002, d = .95$). In the second analysis, the effect of *Line* was significant in both groups: $F(3,69) = 14.70$, $p < .001$, $\varepsilon = .50$, $\eta_p^2 = .39$ for LMA and $F(3,69) = 43.3$, $p < .001$, $\varepsilon = .56$, $\eta_p^2 = .65$ for HMA. Paired contrasts showed that the LMA group was less accurate on the 367 line than the other three lines ($t(23) = 4.29$, $p < .001$, $d = .88$; $t(23) = 4.23, p < .001, d = .86$; and $t(23) = 3.90, p = .001, d = .80$, for the comparisons between the 367 line and the 100, 1,000, and 100,000 lines, respectively). Importantly, for the HMA group, these differences emerged not only in the comparisons between the 367 line and the 100 line ($t(23) = 7.36, p < .001, d = 1.50$), the 1,000 line ($t(23) = 7.82$, $p < .001$, $d = 1.60$), and the 100,000 line ($t(23) = 6.45$, $p = .001, d = 1.32$), showing lower PAE in the former than their LMA counterparts, but also in the comparisons between the 100,000 line and the two familiar lines ($t(23) = 2.8$, $p = .01$, $d = .57$; and $t(23) = 2.45$, $p = .022$, $d = .50$, for the comparisons between the 100,000 line and the 100 and 1,000 lines, respectively). Means of PAE and standard errors of the mean for each group for the four lines are given in Table 4.

Response time (RT)

ANOVA of medians of RT showed a significant effect of *Line* ($F(3,138) = 6.55, p < .001$, $\eta_p^2 = .12$). Neither the effect of *Group* nor the interaction *Line* × *Group* was significant.

Table 4. Means of PAE and standard error of the mean (in brackets) for the LMA and HMA groups for the four lines

	100 line	1,000 line	100,000 line	367 line
LMA	3.2 (.2)	3.1 (.2)	3.4 (.4)	5.5 (.5)
HMA	3.6 (.2)	3.8 (.2)	4.5 (.4)	8.2 (.6)

Note. LMA: low math-anxiety group; HMA: high math-anxiety group.

Paired comparisons between lines showed that RT was slower for the 367 line (mean = 2725.9 ms, SEM = 105.5 ms) than the other three lines: $t(47) = 24.38$, $p < .001$, $d = .72$; $t(47) = 3.22$, $p = .002$, $d = .41$; and $t(47) = 3.47$, $p = .001$, $d = .41$, for the comparisons between the 367line and the 100 (mean = 2483.1 ms, SEM = 102.4 ms), 1,000 (mean = 2517.8 ms, SEM = 120.9 ms), and 100,000 lines (mean = 2504.3 ms, SEM = 109.2 ms), respectively.

Self-reported level of efficacy

The overall ANOVA on the participants' scores in the questionnaire revealed a significant effect of *Line* ($F(3,138) = 95.25$, $p < .001$, $\varepsilon = .87$, $\eta_p^2 = .67$) and *Group* ($F(1,46) = 22.8$, $p < .001$, $\eta_p^2 = .33$). Importantly, the interaction *Line* × *Group* ($F(3,138) = 2.83$, $p = .048$, $\varepsilon = .87$, $\eta_p^2 = .06$) was also significant. To study this interaction in more detail, two separate analyses were performed, like those described in the PAE analysis section. First, *independent t-tests* for each line showed that HMA individuals self-reported a worse performance than their LMA peers for the four lines ($t(46) = 3.17$, $p = .003$, $d = .92$, for the 100 line; $t(46) = 5.75$, $p < .001$, $d = 1.66$ for the 1,000 line; $t(46) = 3.79$, $p < .001$, $d = 1.09$ for the 100,000 line; and $t(46) = 2.27$, $p = .028$, $d = .66$, for the 367 line).

Second, when differences between lines in the self-reported level of efficacy were studied separately in each group, the results showed that the *Line* effect was significant in both the LMA ($F(3,69) = 47.44$, $p < .001$, $\eta_p^2 = .67$) and the HMA ($F(3,69) = 50.64$, $p < .001$, $\eta_p^2 = .69$) groups. Paired contrasts in each group showed that for LMA individuals, all the comparisons were significant. They self-reported (1) worse performance for the 367 line than the other three lines ($t(23) = 9.16$, $p < .001$, $d = 1.87$; $t(23) = 8.90$, $p < .001$, $d = 1.81$; and $t(23) = 3.70$, $p = .001$, $d = .76$ for the comparison with 100, 1,000, and 100,000 lines, respectively); (2) worse performance for the 100,000 line than the 1,000 line ($t(23) = 6.07$, $p < .001$, $d = 1.24$) and the 100 line ($t(23) = 5.94$, $p < .001$, $d = 1.21$); and (3) worse performance for the 1,000 line than the 100 line ($t(23) = 2.09$, $p = .047$, $d = .430$). For HMA individuals, the results were similar to those described for their LMA counterparts, with the exception of the fact that no differences were found when the two unfamiliar lines (100,000 and 367) were compared. As for the other comparisons in this group, all of them were significant (all p-values <.001). Means of self-reported level of performance and standard errors of the mean for each group for the four lines are given in Table 5.

Response time and accuracy in the control task

Medians of RT and accuracy in the control task for each individual were calculated, and group differences were studied by means of *independent t-tests*. Accuracy was calculated for each participant as the number of trials in which the mouse click was made on the

Table 5. Means of self-reported level of performance and standard error of the mean (in brackets) for the LMA and HMA groups for the four lines

	100 line	1,000 line	100,000 line	367 line
LMA	5.96 (.14)	5.58 (.14)	4.29 (.26)	3.37 (.31)
HMA	5.25 (.17)	4.00 (.23)	2.92 (.25)	2.50 (.23)

Note. LMA: low math-anxiety group; HMA: high math-anxiety group.

actual red mark shown on the screen. There were no group differences either in RT ($t(46) = .7, p = .49$) or accuracy ($t(46) = .89, p = .41$).

Discussion

In the present study, the estimates made by individuals with high and low math anxiety for several number line tasks were examined to deepen our knowledge on the cognitive factors that might underlie the difficulties HMA individuals face when dealing with numerical tasks. More specifically, we contrasted two proposals: On the one hand, Maloney and collaborators (Maloney *et al.*, 2010, 2011; see also Núñez-Peña & Suárez-Pellicioni, 2014) suggested that HMA individuals suffer from a low-level numerical deficit, specifically a less precise magnitude representation, that compromises their performance in any task requiring access to this representation. On the other hand, Ashcraft and Kirk (2001; Ashcraft & Krause, 2007; Ashcraft *et al.*, 2000) claimed that intrusive thoughts related to their math anxiety would consume necessary working memory resources, preventing HMA individuals from performing numerical tasks properly. The effects of math anxiety would be particularly clear in difficult tasks with high cognitive load. Four number lines differing in their familiarity were selected to investigate whether HMA and LMA individuals' performance depended on the complexity of the task: two familiar number lines with standard endpoints (100 and 1,000 lines), a non-familiar number line with standard endpoints (100,000 line), and a non-familiar number line with non-standard endpoints (367 line). Moreover, different measures were analysed for each line (best-fit model, response time, PAE, and self-reported level of performance) to obtain a greater understanding about possible differences between groups for each number line.

In the present study, best-fit model and PAE analyses showed that HMA and LMA individuals performed similarly in a NLT with familiar lines, but that the former group had difficulties when facing more cognitively demanding tasks (i.e. non-familiar lines). The linear model provided a better explanation of performance than the logarithmic or cycle power models (one or two cycles) in both groups for the 100, 1,000, and 100,000 lines and also a better explanation than the two-cycle power model and tended towards a better explanation than the one-cycle power model in the 367 line. However, a detailed appraisal of the slope and intercept for the linear model showed group differences for the less familiar lines. Although intercepts in both groups differed from the perfect value 0 in the 367 line, suggesting that both groups overestimated all the values (the intercepts were positive in both groups), the intercept value was larger for the HMA group. As for the slopes, highly math-anxious individuals had lower values than their LMA counterparts for the 100,000 and 367 lines. These results suggest that the HMA overestimated the small numbers and underestimated the large numbers on these non-familiar lines compared to their LMA peers.

PAE analyses gave similar results. Again, group differences were only found for the non-familiar lines (100,000 and 367 lines), HMA individuals being less accurate than their LMA peers. Moreover, whereas both groups were less accurate on the non-familiar number line with non-standard endpoints (367 line) than the other three number lines, the HMA individuals were also less accurate on the other non-familiar line (100,000 line) than the familiar ones. These results are consistent with those reported by Laski and Dulaney (2015), who observed less accurate estimates for their 364–1,364 non-standard endpoint line. It is important to highlight that although motor precision is a relevant ability in terms of performing number line tasks accurately, in our study, the absence of differences between groups in both response time and accuracy in the control task allows us to rule out the possibility that our PAE group differences were due to differences in motor skills between groups.

The fact that the two groups did not differ in the familiar NLT challenges the proposal put forward by Maloney *et al.* (2010, 2011). According to these authors, HMA individuals have a less precise magnitude representation; so in the present study, we expected them to have more trouble even with the more familiar lines; it is worth remembering that Maloney *et al.* found math-anxiety effects even in the 1–9 range. However, this pattern was not observed in the present study and HMA and LMA participants showed identical behaviour on the 100 and 1,000 lines. Although their performance differed on the 367 line, these numbers are included within the 0–1,000 line, and so the fact that the HMA group was less precise in their estimations cannot be attributed to a deficit in their magnitude representation. Lastly, even if we consider it more plausible that a complexity effect also explains the performance of HMA participants on the 100,000 line, we cannot entirely rule out the possibility that their failure on this number line was due to a worse representation of the larger magnitudes. Nevertheless, this possibility would also diverge from Maloney *et al.*'s proposal, as they posited a less precise representation of even the smallest numbers for the HMA group.

This raises the question of how to explain Maloney *et al.*'s results. These authors based their proposal on the fact that HMA individuals showed a larger distance effect than their LMA peers in symbolic number comparison tasks. This effect is usually taken as an indicator of the precision of the numerical magnitude representation, with better representations showing smaller distance effects. However, some researchers (Van Opstal, Gevers, De Moor, & Verguts, 2008; Verguts, Fias, & Stevens, 2005) have proposed an alternative explanation for distance effects measured in symbolic number comparison tasks, claiming that these effects may be located at a decisional level. According to these authors, connections between the numerical stimuli and the response (e.g. 'is larger than') increase monotonically. Close stimuli have similar connection weights to the response nodes and will activate the responses 'smaller than' and 'larger than' to a similar degree, causing competition and a delay in the responses. Dietrich *et al.* (2015) used this alternative proposal to explain why they found a larger distance effect for HMA individuals than for their LMA counterparts in a symbolic comparison task, but failed to find group differences in the non-symbolic task (where the ANS is needed).

Although Lindskog *et al.* (2017) found an interaction between math anxiety and distance effect in a non-symbolic dot comparison task, it is worth remembering that in their experiment, the dot sets to be compared were presented in an intermixed way. This apparently small difference in the experimental design means an increase in processing demands (Price, Palmer, Battista & Ansari, 2012) that might have particularly hampered the performance of HMA participants. Although this possibility remains a hypothesis, it would support the anxiety–complexity effect (Faust *et al.*, 1996).

In general, the above-described results show that the difference between the performance of HMA and LMA participants increases when the task becomes more difficult or complex. These results are consistent with *the anxiety–complexity effect* reported by Ashcraft and Faust (1994), who found that HMA and LMA individuals performed similarly in simple addition and multiplication tasks but that differences between groups emerged in complex additions and mixed arithmetic operations. In the present study, dealing with a line with non-standard points (367 line) proved to be difficult for all participants, who had to calculate proportions from numbers not ending in 0 and hence, that were less easy to manipulate. However, HMAs' performance was particularly impaired. Ashcraft and Kirk (2001) claimed that the effects of math anxiety arise due to the WM load imposed by the intrusive thoughts. Given that Hurst *et al.* (2014) suggested that less familiar endpoints would be more cognitively demanding, our results fit neatly with Ashcraft *et al.*'s prediction that the burden of WM resources would be particularly detrimental in the more complex tasks. We will come back to this hypothesis below and discuss some potential limitations of our study.

As for the 100,000 line, the difficulty came from the unfamiliarity with the largest numbers. Hurst *et al.* (2014) also claimed that properly responding depended upon the fluency with the values in that range. As larger numbers are less frequently encountered than the smallest ones (Dehaene & Mehler, 1992), they were probably less easy to manipulate. This lack of experience was probably increased in HMA participants, who tend to avoid numerical situations because of their condition.

Other results of the present study are worth discussing. First, we found that HMA individuals self-reported a worse performance than their LMA peers for all NLTs. This result is consistent with other studies that have reported that HMA individuals have a low perceived math self-efficacy and distrust their potential to do math tasks successfully (Hembree, 1990; Meece, Wigfield, & Eccles, 1990). This low assessment of their math self-efficacy is a factor that plays an important role in their avoidance of math-related situations and math courses (Ashcraft, Krause, & Hopko, 2007). It probably affects learning motivations and attitudes about math, interfering with the acquisition of mathematics-related competence. However, crucially, the present study has shown that HMA individuals' perception of their self-efficacy does not correspond totally with the reality, because, in fact, they performed as well as their LMA peers on the more familiar lines (100 and 1,000).

Second, we return our attention to the best-fit model analyses. They revealed that there was no relation between the model with better fit to participants' estimates (linear, logarithmic, one-cycle power, and two-cycle power) and group. Moreover, the linear model provided a better explanation of performance in HMA and LMA individuals than the logarithmic or cycle power models (one or two cycles) for the 100, 1,000, and 100,000 lines (for the 367 line, the linear model was also a better fit than the two-cycle model and tended to be better than the one-cycle model). A more linear fit on NLT correlates positively with math achievement (Ashcraft & Moore, 2012; Booth & Siegler, 2006), and it is a reliable predictor of actual and future numerical competencies (Booth & Siegler, 2008; Link *et al.*, 2014; Sasanguie *et al.*, 2011, 2013). Furthermore, it is usually considered an indicator of a more sophisticated way of processing to estimate positions in NLT and more support for accurate estimations (e.g. Cohen & Sarnecka, 2014; Slusser *et al.*, 2013). However, the present study showed that linear estimation of magnitudes does not necessarily mean a perfect match between the value being judged and the estimate of its value. This will only happen when the slope equals one and the intercept equals zero. Slopes and intercepts were analysed and the results showed differences between groups, the HMA group showing lower values for the linear models' slopes and higher values for the linear models' intercepts than their LMA counterparts for the non-familiar lines. Furthermore, differences from the perfect linear mapping for slopes and intercepts were also found. We propose that linear adjustments for estimates in NLT should be interpreted carefully and should be accompanied by slope and intercept analyses.

To conclude, we should mention a few limitations of our study. First of all, we have suggested that the worse performance of HMA individuals is better explained by the hypothesis that their WM resources are reduced by the presence of intrusive thoughts. However, testing the effects of math anxiety and WM load on cognitive reflection questions involving numbers, Morsanyi, Busdraghi, and Primi (2014) found that both were associated with poorer performance, but that their effects did not interact. More importantly, math anxiety, but not WM, was also related to reduced latencies and self-confidence in one's own efficiency. Morsanyi *et al.* (2014) concluded that even if math anxiety might reduce cognitive reflection by diminishing the working memory resources available, WM load alone could not explain their results and proposed that faster responses and low confidence might also have contributed. We also found lower self-confidence in our participants, although it was generalized to all number lines. Furthermore, even if our participants' latencies were not faster than those of their LMA peers, the fact that they tended to answer as fast as their counterparts at the cost of providing less precise responses might be interpreted as an attempt to escape the mathematical context (*local avoidance effect*, Ashcraft & Ridley, 2005), which might have hindered their performance.

Second, we did not control our participants' arithmetical abilities. Given that it is currently under debate whether NLT performance only indicates the precision of mental representation of numerical magnitudes, or also stems from the ability to perform proportion judgements, we cannot completely rule out the possibility that group differences were partly caused by HMAs' poorer capacity for performing the proportion calculations required.

Last, as we did not measure math anxiety during the task, we cannot be sure that positioning numbers in a number line caused math anxiety in our participants. Nevertheless, given that math anxiety has been found to affect simpler tasks such as one-digit comparison (Maloney *et al.*, 2011) and that HMA react differently even in front of math-related words (Suárez-Pellicioni, Núñez-Peña, & Colomé, 2015), it seems improbable that the current task did not trigger anxiety.

Summarizing, further research must be conducted before we can conclusively attribute the worse performance of highly math-anxious individuals in number line tasks to the burden of WM caused by their ruminations. However, our findings clearly rule out the possibility that HMA individuals have a less precise representation of even the smallest magnitudes. Therefore, our results help to broaden our understanding of the cognitive correlates of math anxiety and open the door to other studies that might expand our knowledge of the reasons why math-anxious individuals are particularly impaired in complex tasks.

Acknowledgements

We thank Francisco Javier Diaz Santaella for writing the E-prime program for task presentation and JuliàAgramunt Moreno for his help with data collection. This research was supported by the Spanish Ministry of Economy and Competitiveness and the European Regional

Development Fund, FEDER, under grant PSI2015-69915-R (MINECO/FEDER) and by the Generalitat de Catalunya under grants SGR2017-974 and SGR2017-48.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrow & F. Csáki (Eds.), *Information theory: Proceedings of the 2nd International Symposium* (pp. 267–281). Budapest: Akadémiai Kiadó.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Alexander, L., & Martray, C. (1989). The development of an abbreviated version of the Mathematics Anxiety Rating Scale. *Measurement and Evaluation in Counseling and Development*, *22*, 143–150.

Ashcraft, M. H., & Faust, M. W. (1994). Mathematics anxiety and mental arithmetic performance: An exploratory investigation. *Cognition & Emotion*, *8*(2), 97–125.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology. General*, *130*(2), 224–237.

Ashcraft, M. H., Kirk, E. P., & Hopko, D. (2000). On the cognitive consequences of mathematics anxiety. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 175–196). New York: Psychology Press.

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*, *14*(2), 243–248.

Ashcraft, M. H., Krause, J. A., & Hopko, D. R. (2007). Is math anxiety a mathematical learning disability? In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is math so hard for some children?* (pp. 329–348). Baltimore, MD: Brookes.

Ashcraft, M. H., & Ridley, K. (2005). Math anxiety and its cognitive consequences: A tutorial review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 315–325). New York: Psychology Press.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*(1), 125–135.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*(1), 189–201.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude presentations influence arithmetic learning. *Child Development*, *79*(4), 1016–1031.

Brunyé, T. T., Mahoney, C. R., Giles, G. E., Rapp, D. N., Taylor, H. A., & Kanarek, R. B. (2013). Learning to relax: Evaluating four brief interventions for overcoming the negative emotions accompanying math anxiety. *Learning and Individual Differences*, *27*, 1–7. 10

Chang, H., & Beilock, S. L. (2016). The math anxiety-math performance link and its relation to individual and environmental factors: A review of current behavioral and psychophysiological research. *Current Opinion in Behavioral Sciences*, *10*, 33–38.

Cohen, D. J., & Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology*, *50*(6), 1640–1652.

Colomé, À. (2018). Representation of numerical magnitude in math-anxious individuals. *Quarterly Journal of Experimental Psychology*, *???*, 1–12. https://doi.org/10.1177/1747021817752094 11

Crollen, V., & Noël, M. P. (2015). Spatial and numerical processing in children with high and low visuospatial abilities. *Journal of Experimental Child Psychology*, *132*, 84–98.

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

Dietrich, J. F., Huber, S., Moeller, K., & Klein, E. (2015). The influence of math anxiety on symbolic and non-symbolic magnitude processing. *Frontiers in Psychology*, *27*(6), 1621.

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, *7*, 508.

Faust, M. W., Ashcraft, M. H., & Fleck, D. E. (1996). Mathematics anxiety effects in simple and complex addition. *Mathematical Cognition*, *2*(1), 25–62.

Ferguson, A. M., Maloney, E. A., Fugelsang, J., & Risko, E. F. (2015). On the relation between math and spatial ability: The case of math anxiety. *Learning and Individual Differences*, *39*, 1–12.

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The math anxiety-performance link. *Current Directions in Psychological Science*, *26*(1), 52–58.

Geisser, S., & Greenhouse, S. W. (1958). An extension of Box's result on the use of F distribution in multivariate analysis. *Annals of Mathematical Statistics*, *29*, 885–891.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, *21*(1), 33–46.

Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review*, *107*(3), 500–524.

Hopko, D. R., Mcneil, D. W., Gleason, P. J., & Rabalais, A. E. (2002). The emotional stroop paradigm: Performance as a function of stimulus properties and self-reported mathematics anxiety. *Cognitive Therapy and Research*, *26*(2), 157–166.

Hurst, M., Leigh Monahan, K., Heller, E., & Cordes, S. (2014). 123s and ABCs: Developmental shifts in logarithmic-to-linear responding reflect fluency with sequence values. *Developmental Science*, *6*, 892–904.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307.

Keselman, H. J. (1998). Testing treatment effects in repeated measures designs: An update for psychophysiological researchers. *Psychophysiology*, *35*(4), 470–478.

Laski, E. V., & Dulaney, A. (2015). When prior knowledge interferes, inhibitory control matters for learning: The case of numerical magnitude representations. *Journal of Educational Psychology*, *107*(4), 1035–1050.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, *78*(6), 1723–1743.

Lindskog, M., Winman, A., & Poom, L. (2017). Individual differences in nonverbal number skills predict math anxiety. *Cognition*, *159*, 156–162.

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.

Maloney, E. A., Ansari, D., & Fugelsang, J. A. (2011). The effect of mathematics anxiety on the processing of numerical magnitude. *Quarterly Journal of Experimental Psychology*, *64*(1), 10–16.

Maloney, E. A., Risko, E. F., Ansari, D., & Fugelsang, J. A. (2010). Mathematics anxiety affects counting but not subitizing during visual enumeration. *Cognition*, *114*(2), 293–297.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology*, *82*(1), 60–70.

Morsanyi, K., Busdraghi, C., & Primi, C. (2014). Mathematical anxiety is linked to reduced cognitive reflection: A potential road from discomfort in the mathematics classroom to susceptibility to biases. *Behavioral and Brain Functions*, *10*, 31.

Núñez-Peña, M. I., & Suárez-Pellicioni, M. (2014). Less precise representation of numerical magnitude in high math-anxious individuals: An ERP study of the size and distance effects. *Biological Psychology*, *103*, 176–183.

Núñez-Peña, M. I., Suárez-Pellicioni, M., Guilera, G., & Mercadé-Carranza, C. (2013). A Spanish version of the short Mathematics Anxiety Rating Scale (sMARS). *Learning and Individual Differences*, *24*, 204–210.

OCDE. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading science,*
12 *problem solving and financial literacy*. Paris, France: ???.

Park, D., Ramirez, G., & Beilock, S. L. (2014). The role of expressive writing in math anxiety. *Journal of Experimental Psychology, Applied*, *20*, 103–111.

Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counceling Psychology*, *19*(6), 551–554.

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, *30*(2), 344–357.

Sasanguie, D., Defever, E., Van den Bussche, E., & Reynvoet, B. (2011). The reliability of and the relation between non-symbolic numerical distance effects in comparison, same-different judgments and priming. *Acta Psychologica*, *136*(1), 73–80.

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, *114*(3), 418–431.

Siegler, R., & Booth, J. (2004). Development of numerical estimation in young children. *Child Development*, *75*(2), 428–444.

Siegler, R., & Booth, J. (2005). Development of numerical estimation: A review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 197–212). New York: Psychology Press.

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.

Simms, V., Clayton, S., Cragg, L., Gilmore, C., & Johnson, S. (2016). Explaining the relationship between number line estimation and mathematical achievement: The role of visuo-motor integration and visuo-spatial skills. *Journal of Experimental Child Psychology*, *145*, 22–33.

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

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 683–692.

Spielberger, C. D., Gorsuch, R., & Lushene, R. (2008). *Cuestionario de ansiedad Estado-Rasgo, STAI*. Madrid: TEAEdiciones.

Spielberger, C. D., Gorsuch, R., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait anxiety inventory*. Palo Alto, CA: Consulting Psychology Press.

Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, À. (2015). Attentional bias in high math-anxious individuals: Evidence from an emotional Stroop task. *Frontiers in Psychology*, *6*, 1577.

Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, À. (2016). Math anxiety: A review of its cognitive consequences, psychophysiological correlates, and brain bases. *Cognitive, Affective & Behavioural Neuroscience*, *16*(1), 3–22.

Sullivan, J. L., Juhasz, B. J., Slattery, T. J., & Barth, H. C. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review*, *18*(3), 557–563.

Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review*, *15*(2), 419–425.

Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, *12*, 66–80.

Received 22 December 2017; revised version received 29 June 2018

Appendix : Target numbers presented for the four lines

100 line	1,000 line	100,000 line	367 line
4	8	870	271
6	15	1,522	273
8	25	2,609	275
12	56	5,652	279
14	109	10,870	281
17	154	15,435	284
18	237	23,696	285
21	290	29,022	288
24	338	33,805	291
25	388	38,478	292
29	430	43,043	296
39	467	46,739	306
42	517	51,739	309
48	560	56,087	315
52	599	61,305	319
57	650	65,217	324
61	696	69,783	328
64	761	76,087	331
72	839	83,913	339
79	889	88,913	346
81	939	93,913	348
84	980	98,043	351
90	989	98,913	357
96	993	99,348	363