Cognition 133 (2014) 188–200

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Improving arithmetic performance with number sense training: An investigation of underlying mechanism

Joonkoo Park [a,b,*], Elizabeth M. Brannon [b,c]

[a] Department of Psychological and Brain Sciences, University of Massachusetts at Amherst, 135 Hicks Way, Amherst, MA 01003, USA
[b] Center for Cognitive Neuroscience, Duke University, 450 Research Drive, Durham, NC 27708, USA
[c] Department of Psychology and Neuroscience, Duke University, 417 Chapel Drive, Durham, NC 27708, USA

ARTICLE INFO

Article history:
Received 12 September 2013
Revised 12 June 2014
Accepted 13 June 2014

Keywords:
Numerical cognition
Approximate number system
Cognitive training
Math
Approximate arithmetic
Arithmetic

ABSTRACT

A nonverbal primitive number sense allows approximate estimation and mental manipulations on numerical quantities without the use of numerical symbols. In a recent randomized controlled intervention study in adults, we demonstrated that repeated training on a non-symbolic approximate arithmetic task resulted in improved exact symbolic arithmetic performance, suggesting a causal relationship between the primitive number sense and arithmetic competence. Here, we investigate the potential mechanisms underlying this causal relationship. We constructed multiple training conditions designed to isolate distinct cognitive components of the approximate arithmetic task. We then assessed the effectiveness of these training conditions in improving exact symbolic arithmetic in adults. We found that training on approximate arithmetic, but not on numerical comparison, numerical matching, or visuo-spatial short-term memory, improves symbolic arithmetic performance. In addition, a second experiment revealed that our approximate arithmetic task does not require verbal encoding of number, ruling out an alternative explanation that participants use exact symbolic strategies during approximate arithmetic training. Based on these results, we propose that nonverbal numerical quantity manipulation is one key factor that drives the link between the primitive number sense and symbolic arithmetic competence. Future work should investigate whether training young children on approximate arithmetic tasks even before they solidify their symbolic number understanding is fruitful for improving readiness for math education.

1. Introduction

Humans are endowed with an intuitive understanding of number. Without counting or the use of symbols, we are able to estimate, compare, and mentally manipulate large numerical quantities (Feigenson, Dehaene, & Spelke, 2004). This nonverbal primitive number sense, termed the approximate number system (ANS), is shared by a diverse range of animal species (e.g., Merritt, DeWind, & Brannon, 2012) and is already present at birth in humans (e.g., Izard, Sann, Spelke, & Streri, 2009), which suggests that it is an evolutionarily ancient and developmentally rudimentary cognitive system (Brannon, 2006; Dehaene, 1999).

The ANS is characterized by imprecise noisy internal representations of number (Feigenson et al., 2004; Gallistel & Gelman, 2000). The signature of this system is that discrimination of number adheres to Weber's law—that is, the ability to discriminate two values depends on the ratio between the two values and not just their

* Corresponding author at: Department of Psychological and Brain Sciences, University of Massachusetts at Amherst, 135 Hicks Way / Tobin Hall, Amherst, MA 01003, USA. Tel.: +1 413 545 2383; fax: +1 413 545 0996.

E-mail address: joonkoo@umass.edu (J. Park).

http://dx.doi.org/10.1016/j.cognition.2014.06.011
0010-0277/

absolute difference. This characteristic allows us to estimate the precision of the internal representation of number by computing a Weber fraction (w) from data in which participants are asked to make simple non-symbolic numerical comparisons.

The ANS is hypothesized to be a core system and to be foundational for mathematical thinking in human adults (e.g., Feigenson et al., 2004). This hypothesis has been supported by recent studies that show a correlation between individual precision of the ANS, as measured by w, and individual mathematical competence as measured by standardized math tests, often even after controlling for verbal ability and general intelligence measures (DeWind & Brannon, 2012; Gilmore, McCarthy, & Spelke, 2010; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Halberda, Mazzocco, & Feigenson, 2008; Libertus, Feigenson, & Halberda, 2011; Libertus, Odic, & Halberda, 2012; Mazzocco, Feigenson, & Halberda, 2011a; Mazzocco, Feigenson, & Halberda, 2011b; Piazza et al., 2010). It should be noted, however, that many other studies now report negative findings (Castronovo & Gobel, 2012; Fuhs & McNeil, 2013; Gobel, Watson, Lervag, & Hulme, 2014; Holloway & Ansari, 2009; Inglis, Attridge, Batchelor, & Gilmore, 2011; Kolkman, Kroesbergen, & Leseman, 2013; Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013; Price, Palmer, Battista, & Ansari, 2012; Sasanguie, De Smedt, Defever, & Reynvoet, 2012; Sasanguie, Defever, Maertens, & Reynvoet, 2013; Sasanguie, Gobel, Moll, Smets, & Reynvoet, 2013; Tibber et al., 2013; Wei, Yuan, Chen, & Zhou, 2012), creating a controversy over whether the ANS is causally related to symbolic math and if so exactly what mechanisms underlie this relationship.

In a recent study (Park & Brannon, 2013), we tested the hypothesis that the ANS is causally related to math performance and found that training on a non-symbolic approximate arithmetic task leads to improvement in exact symbolic arithmetic. Participants who were trained to roughly add or subtract numerical quantities represented in dot arrays showed a significant improvement in a two- and three-digit arithmetic test. Although this previous study demonstrates that training with non-symbolic approximate arithmetic transfers to improvement in symbolic arithmetic, and provides strong evidence that the ANS may be directly and causally related to math ability, the study leaves open which cognitive component of the approximate arithmetic task was crucial for the obtained transfer effect. The non-symbolic approximate arithmetic task involves a number of different cognitive processes: estimation of numerical quantity from dot arrays, visual short-term memory required to hold numerical quantities and their sum or difference, and mental manipulation to combine or separate numerical quantities.

Here, in Experiment 1, we used a randomized controlled intervention approach to investigate the mechanisms that underlie the relationship between approximate arithmetic and exact symbolic arithmetic. We constructed multiple training conditions aimed at isolating and improving distinct cognitive components of the non-symbolic approximate arithmetic task. We then compared the transfer effects in exact symbolic arithmetic performance across these training conditions. In Experiment 2, a verbal interference approach was used to test the alternative hypothesis that the transfer effect observed in Park and Brannon (2013) and Experiment 1 of this report might be a function of undetected verbal encoding during approximate arithmetic.

2. Experiment 1

2.1. Material and methods

2.1.1. Participants

A total of 88 participants between 18 and 34 years of age participated in Experiment 1. Participants were recruited from the Duke University community and gave written informed consent in accordance with a Duke University Institutional Review Board approved protocol.

2.1.2. Procedure

Seventy-one participants were randomly assigned to four training groups in Experiment 1A: approximate arithmetic ($N = 18$, 7 males, age 21.5 ± 2.55 [mean ± s.d.]), approximate number comparison ($N = 18$, 7 males, age 21.9 ± 3.92), visuo-spatial short-term memory ($N = 18$, 8 males, age 21.4 ± 4.20), and numerical symbol ordering ($N = 17$, 6 males, age 21.5 ± 3.01). An additional 17 participants were recruited for Experiment 1B (9 male, age 21.1 ± 1.79). This group was trained with a task similar to the approximate arithmetic task but without the arithmetic component (henceforth referred to as approximate comparison and matching). See Section 2.1.3 for details about each training condition.

Upon enrollment, all participants were first given a pretest battery that consisted of an exact symbolic arithmetic test, vocabulary test, non-symbolic numerical comparison test, visuo-spatial 2-back test, and numeral order judgment test (see Section 2.1.4 for details). Participants in each training group were trained on six different sessions that took place within a two-week period. These training sessions were followed by a posttest session. The order of the pre- and posttests given was randomized. The average number of days between the first (pretest) and the last (posttest) sessions were 9.2 days (approximate arithmetic), 9.2 days (approximate number comparison), 9.3 days (visuo-spatial short-term memory), 9.1 days (numerical symbol ordering) and 9.0 days (Exp. 1B approximate comparison and matching) for each of the training groups.

2.1.2.1. Approximate arithmetic. The nonsymbolic approximate arithmetic training, identical to the condition used in Park and Brannon (2013), was used to replicate that study and to compare with other training conditions that were designed to isolate particular cognitive components of approximate arithmetic. Identical to the approximate arithmetic training condition used in Park and Brannon (2013), this training condition (Fig. 1A) required participants to add or subtract large quantities of visually presented dot arrays without counting. Participants were cued to mentally add or subtract two numerical quantities, ranging from 9 to 36, represented in dot arrays. Then, they were asked to either compare the sum or the difference

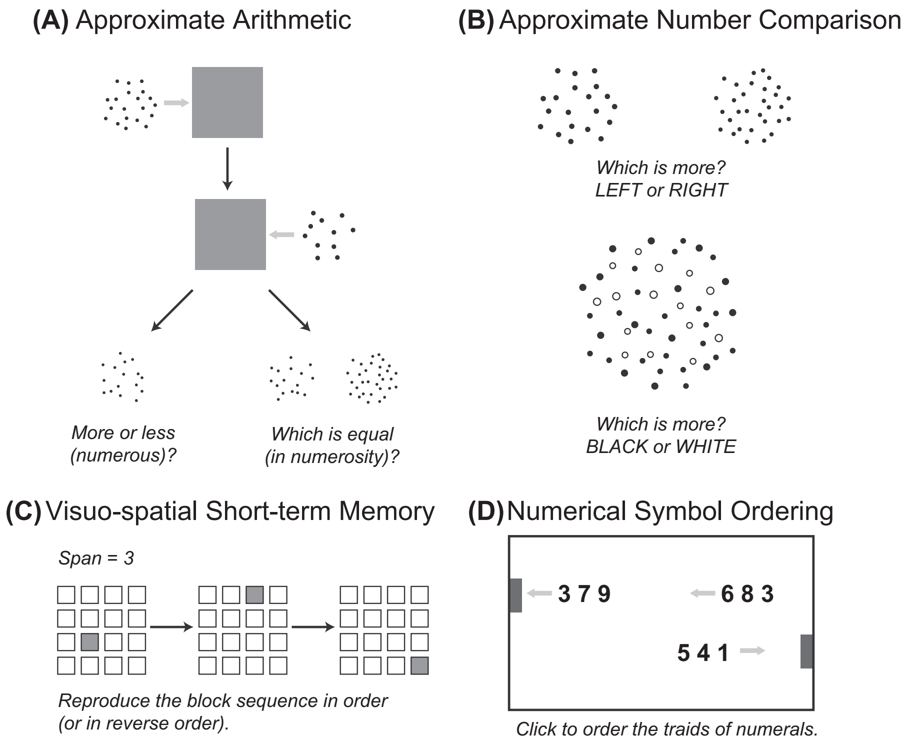

Fig. 1. Illustration of the training conditions used in Experiment 1A.

with a numerical quantity represented in a third dot array (compare trials) or to choose one of two dot arrays that matched the sum or the difference in number (match trials). Two trial types were used to minimize the development of task-specific strategies. Dot arrays were shown briefly (1000 ms for the first two dot arrays and 1500 ms for the dot arrays to be compared or matched) to prevent participants from counting, and the dot size was homogeneous within an array but differed across arrays to prevent participants from relying on total surface area to make judgments. Participants responded with a mouse click on each trial, and feedback was provided after each trial.

The difficulty of the task was manipulated by varying the numerical distance between the correct sum or difference and the alternative option in a log-base2 scale (hence referred to as the log-difference level). In the first training session, participants received ten practice trials with the log-difference level of 1.5. The initial log difference level was 1.5 for both the "compare" and "match" trials. The difficulty level was titrated separately for compare and match trial types; the log-difference levels increased by one of the values randomly chosen from [0.08, 0.09, 0.10, 0.11, 0.12] when the average accuracy of a block of twenty trials was less then 70% and decreased by one of the values randomly chosen from [0.13, 0.14, 0.15, 0.16, 0.17] when the average accuracy was greater than 85%. For example, when the log-difference level was 1.5, participants solved problems where the ratio between the correct answer and the alternative option was 2.83(=2^1.5) to 1. If the log-difference level decreased to 1.4, participants then solved problems with the ratio of 2.64(=2^1.4) to 1. Participants received ten blocks in each session, and each block consisted of twenty trials. The log-difference levels at the end of each session were used as the difficulty level for the beginning of the following session. Each session took approximately 25 min.

2.1.2.2. Approximate number comparison. In order to solve approximate arithmetic problems as described above, participants needed to effectively estimate and compare numerical quantities represented as dot arrays. In order to isolate this cognitive component of the approximate arithmetic task, a non-symbolic approximate number comparison training was devised that was similar to a task conventionally used to assess ANS precision. This training condition (Fig. 1B) required participants to choose one of two dot arrays that contained more dots. On half of the trials, a number of white dots and a number of black dots were flashed intermixed on the center of the gray-background screen for 750 ms. Participants were cued to choose whether there were more white dots or black dots (mixed trials). On the other half of the trials, two dot arrays (consisting either all white or all black dots) were flashed side-by-side also for 750 ms (separate trials). In these cases, participants were cued to choose whether the dot array on the left or right contained more dots. Two trial types were used to minimize the development of task-specific strategies. Orthogonally, on half of the trials, the two dot arrays to be compared matched in average individual dot area, and on the other half, the two dots arrays matched in the total surface area. Individual dot size was heterogeneous within an array, so that at least a subset of dots in each array was equal in dot size across the two arrays. Participants responded with a mouse click on each trial, and feedback was provided after each trial.

As in the approximate arithmetic condition, the difficulty of the task was manipulated by varying the numerical distance between the two dot arrays in a log-base2 scale. Numerosity in one of the two dot arrays ranged from 16 to 32, and the other dot array contained a number that is different by some log-difference ratio. For example, when the log-difference level is r, and if one of the two arrays contained 24 dots, then the other dots could contain 24×2^r or $24/2^r$ dots. In the first training session participants received 15 practice trials with the log difference level of 1.15. The initial log difference level was also 1.15. The difficulty level was titrated separately for the mixed and separate trial types. The log-difference levels increased by one of the values randomly chosen from [0.08, 0.09, 0.10, 0.11, 0.12] when the average accuracy of a block of twenty trials was less then 70% and decreased by one of the values randomly chosen from [0.13, 0.14, 0.15, 0.16, 0.17] when the average accuracy was greater than 85%. Participants received ten blocks in each session, and each block consisted of twenty trials. The log difference levels at the end of each session were used as the difficulty level for the beginning of the following session. Each session took approximately 25 min.

2.1.2.3. Visuo-spatial short-term memory. The visuo-spatial short-term memory training was designed to enhance visual short-term memory that may be a critical component of approximate arithmetic. It was devised based on the Corsi block memory span task (Corsi, 1972). In this task (Fig. 1C), sixteen white blocks were presented on a 4-by-4 grid. On each trial, a sequence of blocks lit up, and the participant was required to reproduce the sequence by clicking on the blocks with a mouse. Each block lit up for 500 ms, and the duration between blocks was 200 ms. On a random half of the trials, participants were cued to repeat the sequence in order. On the other half, participants were cued to repeat in reverse order. As in the previous two conditions, two trial types were used to minimize the development of task-specific strategies. In the first training session, participants received 8 trials of practice with 3 block spans (i.e. the number of blocks to be memorized). Participants then began the training condition with span of 3 blocks. Task difficulty was titrated based on each subject's performance. After each series of 20 trials, the span increased by 1 if the accuracy was greater than 85% in a series of 20 trials and decreased by 1 if the accuracy was lower than 70%. Each session was designed to last approximately 25 min.

2.1.2.4. Numerical symbol ordering. The numerical symbol ordering training, identical to the condition used in Park and Brannon (2013), was used to replicate that study. The original hypothesis was that triad ordering would increase the automaticity of associations between Arabic numerals and—given the positive correlation between a variant of this task and symbolic arithmetic performance (Lyons & Beilock, 2011)—that such training might enhance symbolic math performance. This training condition required participants to order sets of three Arabic numerals (Fig. 1D). Triads of Arabic numerals moved from one end of the screen (resolution of 1440×900) to the opposite end. A mouse click on a triad changed the position of the three numbers to one of four possible pre-specified sequences. One of the four sequences presented the numerals in ascending (if the triad is moving from right to left) or descending (if the triad is moving from left to right) order. On 90% of the trials, the triads first appeared in a sequence that showed a non-monotonic or reversed order; 10% of the triads first appeared in the correct order. The task was to click on the triads until they were in an ascending (if the triad is moving from right to left) or descending (if the triad is moving from left to right) order before they disappeared off the screen. As in the previously described conditions, two trial types were used to minimize the development of task-specific strategies. A maximum of three triads appeared on the screen at any time. Trial-by-trial feedback as to whether a triad was correctly ordered or not was given on each trial and was indicated by a color change of a gray block into which the triads disappeared.

The difficulty of the task was manipulated by varying the speed (in pixels per second) of the triads traversing the screen (hence referred to as the item speed). In the first training session, participants received about a minute of practice at 125 pixel/s. The initial item speed was also 125 pixel/s. Then, the difficulty level was titrated over each 2.2-min block over the course of training. If the accuracy was less than 80% for a given block, item speed decreased by one of the values randomly chosen from [4, 5, 6, 7, 8] pixel/s. If the accuracy was greater then 90%, item speed increased by one of the values randomly chosen from [10, 11, 12, 13, 14] pixel/s. The final difficulty level for each session was used as the difficulty level for the beginning of the following session. The triads in the very first block were constructed from random numbers chosen from a pool of numbers from 1 to 9. After every two blocks, the pool increased by one, so that new numbers could be introduced over time. Triads never consisted of three consecutive numbers (e.g. 7, 8, 9). Each session took approximately 25 min.

2.1.2.5. Approximate comparison and matching (Experiment 1B). The last training condition, used in Experiment 1B, was designed to be identical to approximate arithmetic but without addition and subtraction components (Fig. 5A). The condition is called Experiment 1B because it was run immediately after the four other conditions rather than as part of the random assignment. Participants briefly saw one dot array and were asked to compare its magnitude with a second dot array or to judge whether it matches one of two dot arrays in number. All other training parameters including the initial difficulty level and the titration procedure were identical to that of the approximate arithmetic training.

2.1.3. Pre- and posttest battery

2.1.3.1. Exact symbolic arithmetic test. Participants solved two- and three-digit addition and subtraction problems on a computer. In each of two 5-min blocks, participants were instructed to solve as many problems as possible using the number pad keys. Each problem consisted of two or

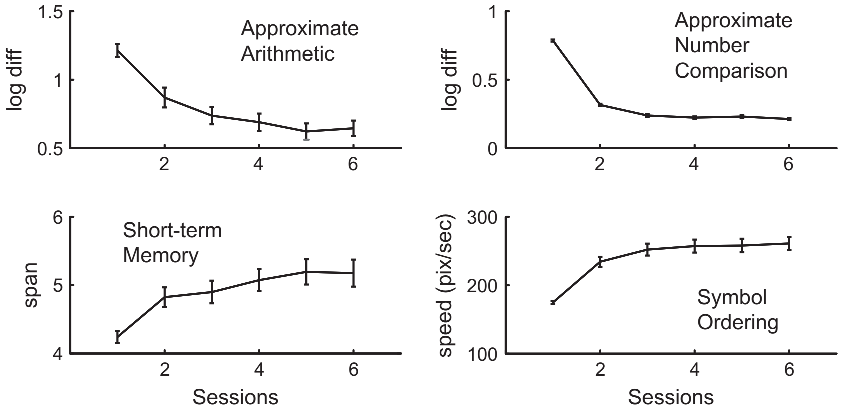

Fig. 2. Performance in the four training conditions in Experiment 1A. The vertical axes for approximate arithmetic and approximate number comparison refers to the numerical distance between the correct answer and the alternative in a log-base2 scale. As sessions progressed, participants were able to solve more difficult problems with same accuracy. The vertical axis for visual-spatial short-term memory refers to the number of blocks in a sequence to be remembered. The vertical axis for numerical symbol ordering refers to the speed at which the triads traversed across the screen. The error bars indicate the standard error of the mean. Standard errors in the approximate number comparison condition were too small to be visible in the graphs. Their numerical values were 0.0071, 0.0087, 0.0113, 0.0093, 0.0105, and 0.0080 for the six sessions.

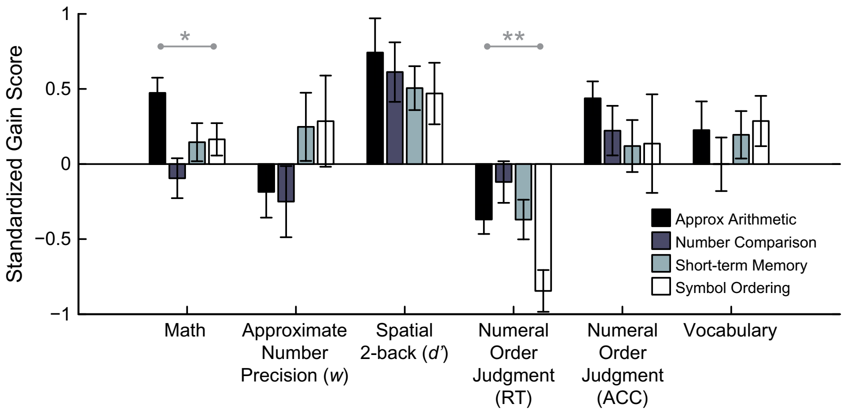

Fig. 3. Performance change of the test battery in a standardized scale. Asterisks indicate significant main effect of condition in an ANOVA: $^{*}p < 0.05$; $^{**}p < 0.01$.

three operands. The operands ranged from 11 to 244, and the correct answers ranged from 11 to 284. Prior to the actual task, participants practiced typing twelve numbers displayed on the screen, and eight arithmetic problems similar to the ones that appeared in the actual task. Two arithmetic problem sets were constructed for pre- and posttest, and the order was counterbalanced across participants. These problem sets included hundreds of arithmetic problems, and on a given trial one problem from the set was chosen randomly without replacement. The overall difficulty of the two sets was roughly equated by balancing the number of problems that involved carrying and borrowing. Performance in each session was quantified as the number of problems each participant solved correctly within the 10-min span. In the Results section of the paper, for conciseness we operationally define math performance as performance on this exact symbolic arithmetic test.

2.1.3.2. Non-symbolic numerical comparison test. On each trial two white dot arrays ranging in numerosity from 8 to 24 were presented side-by-side on a black screen for 750 ms. Participants were asked to judge which side contains more dots. The ratios between the two numerosities were 2:1, 5:4, 7:6, 8:7, 9:8, and 10:9. In order to discourage reliance on other continuous variables when making judgments, the average dot size was equal in the two arrays in one third of the trials, the total surface area was equal in another third of the trials, and finally the total surface area for the array with less dots was greater than that for the array with more dots in the last third of the trial. Furthermore, individual dot size was varied within a dot array so that a subset of dots in each array was equal in dot size across the two arrays. Each participant was presented with a total of 264 trials separated into four blocks with no trial-by-trial feedback. The precision of the approximate number representation was computed individually by estimating a Weber fraction (w) following Pica, Lemer, Izard, and Dehaene (2004). Note that this task was similar but not identical to the approximate number comparison training task.

2.1.3.3. Spatial 2-back test. On each trial of the spatial 2-back test (adapted from Gevins & Cutillo, 1993), a white circle appeared in one of six locations. Participants were asked to judge whether the location of the circle on the current trial matched the location of the circle two trials back in time. A total of 168 trials separated into two blocks were presented with no trial-by-trial feedback. On approximately 20% of the trials, the location of the circle matched the location of the circle two trials back. Performance was quantified as the sensitivity index d′, where the z-score of the false alarm rate was subtracted from the z-score of the hit rate. The sensitivity index of perfect performance was arbitrarily bounded to 5.0.

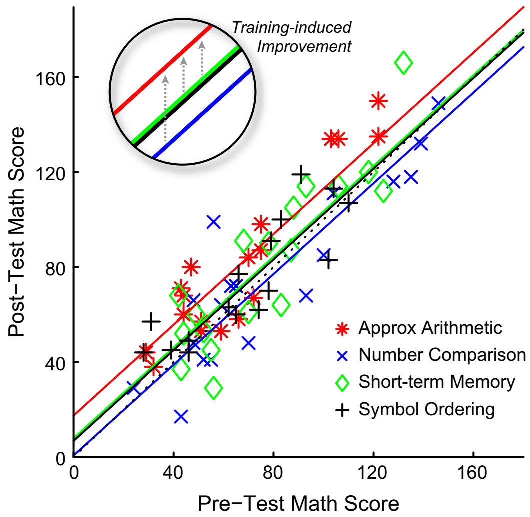

Fig. 4. An ANCOVA approach to assess differences in math improvement across the four training groups in Experiment 1A. Transfer effects in math (that is, any improvement in posttest compared to pretest) is indicated by the degree to which the fitted line departs from the diagonal dotted line. Group differences in the transfer effect are indicated by the degree to which fitted lines are apart from each other. The circle on the upper left corner shows an enlarged view of the fitted lines for the four training groups. Gray arrows graphically depict the results that training-induced improvement in the approximate arithmetic condition exceeds that in all three other conditions.

2.1.3.4. Numeral order judgment test. In this test, modeled after Lyons and Beilock (2011), participants saw a triad of Arabic numerals and judged whether the triad was in an ascending order or not. Half of the trials were presented in an ascending order, and the other half were in either a non-monotonic order or a descending order. The triads were constructed from the numerals 1 through 9. On some trials, the numerical distance between the smallest and the largest number was 4, while the distance between the smallest and the middle number was either 1 or 2. On the other trials, the numerical distance between the smallest and the largest number was 7, while the distance between the smallest and the middle number was either 3 or 4. The trials were presented for 1500 ms, and no feedback was given. Participants performed two blocks of 64 trials each. Performance was quantified as the mean accuracy and the median reaction time across the correct trials in each participant. Note that this task was similar but not identical to the numerical symbol ordering training task.

2.1.3.5. Vocabulary test. Participants were given 5 min to answer 42 multiple choice vocabulary problems on a

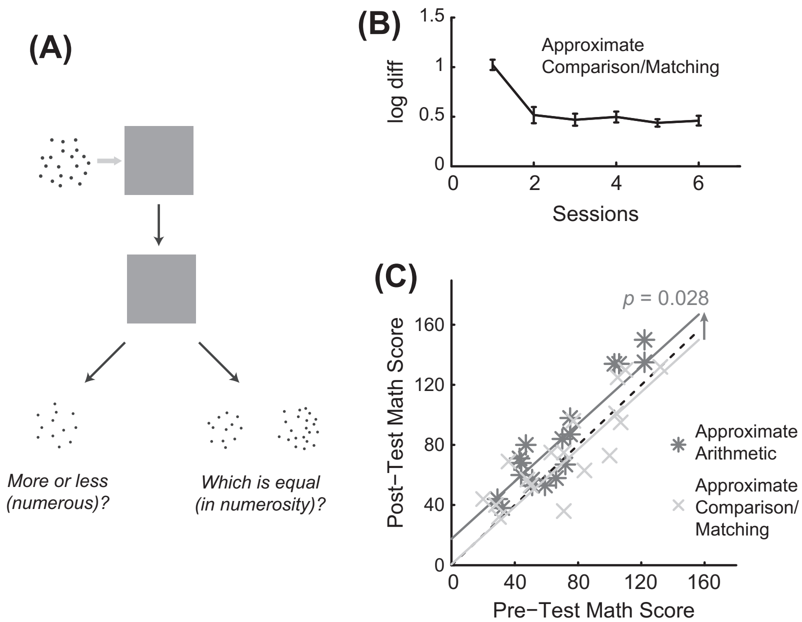

Fig. 5. Design and results of Experiment 1B. (A) A non-symbolic approximate numerical task identical to approximate arithmetic but without the arithmetic component was administered. (B) Training effects of this approximate comparison and matching condition. (C) An ANCOVA revealed a significant group effect, implying that math improvement accompanied by approximate arithmetic exceeded that accompanied by approximate comparison and matching.

computer. The pre and posttest sets were taken from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976). Part 1 of the Extended Range Vocabulary Test and the Advanced Vocabulary Tests I and II made one set, Part 2 of those tests made the other set. The order of the two sets was counterbalanced across participants. Performance was quantified as the number of problems answered correctly minus 1/4 of the number marked incorrectly (participants were instructed about this penalty for guessing).

2.1.4. Analysis of training and transfer effects

Mean performance level on each session was calculated for each participant, and training effects were assessed by evaluating linear and quadratic contrasts in a repeated-measures ANOVA over six sessions of training.

The transfer effects were first assessed by computing gain scores for each measure. The gain score was calculated as the posttest score minus the pretest score for each of the five tests given before and after the training sessions. The gain scores from a few participants in a few of the tests, appeared to be outliers. For example, two participant's pretest score on the non-symbolic numerical comparison test was at or near chance. Another participant's pretest score on the numeral order judgment test was at chance. In addition, performance in the non-symbolic numerical comparison test for some participants did not allow a reliable estimation of the Weber fraction. In order to account for these anomalies, we excluded gain scores when the value was smaller than $Q1 - 2 \times IQR$ or larger than $Q3 + 2 \times IQR$, where Q1 and Q3 are the first and third quartile and IQR is the interquartile range. This procedure removed a total of sixteen data points (out of 528 data points): 8 data points for the measure of the precision of the approximate number representation, 1 data point for the exact symbolic arithmetic test, 2 data points for the spatial two-back d′, 4 data points for the accuracy measure of the numeral order judgment test, and 1 data point for the reaction time measure of the numeral order judgment test. These outliers were equally represented across the five training conditions.

While gain scores can be used to test a significant improvement in performance from pre to posttest due to the training condition, they may alternatively reflect mere practice effects given that participants were taking similar tests on two separate occasions. To rule out such test–retest effects, we conducted a one-way ANOVA on each of the gain scores with training group as a factor.

For the gain score measures that showed group differences, an ANCOVA was used to assess the relative transfer effects among training groups while accounting for possible systematic differences in pretest scores across different training groups (see Dimitrov & Rumrill, 2003). That is, the posttest scores were compared across the training groups with pretest scores entered as a covariate. Put in another perspective, this analysis allows us to assess the (linear) relationship between pretest and posttest scores for each group separately (see Fig. 4). When there was no evidence of group-by-test score interaction (i.e., when the best fitting linear slopes among the training groups were not significantly different), a model assuming parallel slopes was used to illustrate the results in Figs. 4 and 5C.

2.1.5. Correlation analyses

An orthogonal question our dataset allowed us to address is which pretest measures correlate with symbolic arithmetic performance. To address this, we examined zero-order Pearson correlations between exact symbolic arithmetic performance and each of the other measures at pretest across all participants (see Fig. 7). Data points where performance was below chance, smaller than $Q1 - 2 \times IQR$, or larger than $Q3 + 2 \times IQR$ were excluded in this analysis (7/528 data points).

2.2. Results and discussion

2.2.1. Training effects

Performance improved significantly over six sessions in all four training conditions of Experiment 1A (see Fig. 2, all linear and quadratic contrasts in a repeated-measures analysis of variance [ANOVA], $p < 0.001$). In the approximate arithmetic training condition participants' ability to reliably solve the dot problems improved, on average, from solving a ratio of 2.32:1 to a ratio of 1.56:1. In the approximate number comparison training condition participants improved from reliably comparing a ratio of 1.72:1 to a ratio of 1.16:1. In the short-term memory training condition participants initially remembered an average of 4.2 blocks in session 1 and 5.2 blocks by session 6. Finally, participants who received the numerical symbol ordering training showed substantial improvement as indicated by a significant increase in the item speed from 175 to 261 pixels per second.

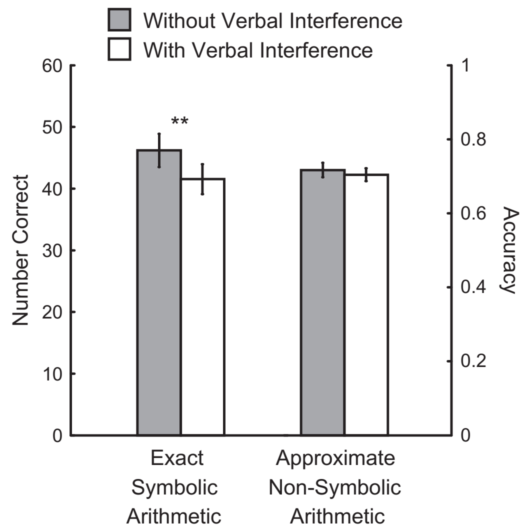

Fig. 6. Results from the verbal interference task in Experiment 2. The number of problems solved correctly in the 10-minutes allotted for the exact symbolic arithmetic test is shown on the left vertical axis, and proportion correct for the approximate non-symbolic arithmetic test is shown on the right vertical axis. $^{**}p < 0.01$.

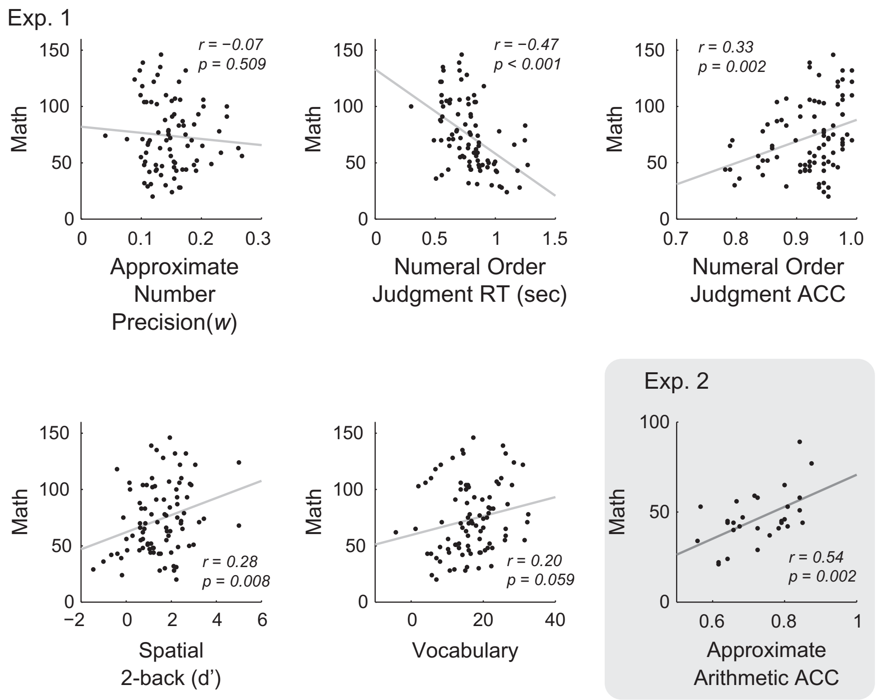

Fig. 7. Correlation between exact symbolic arithmetic (math) performance and other test measures assessed from the pretests in Experiment 1 and from Experiment 2 without the verbal interference condition. Note that a total of 7 min were given to solve math problems in Experiment 2, whereas 10 min were given in Experiment 1. Gray line represents best linear fit.

2.2.2. Transfer effects

More central to our research question was whether and to what degree improvement in each training condition would transfer to improvements in exact symbolic arithmetic (math). Fig. 3 illustrates these transfer effects via a standardized gain score (a measure of performance change computed by subtracting the pretest from the posttest scores for each group and dividing this difference by the standard deviation of the pretest scores across all participants). The ANOVA on these gain scores revealed a significant main effect of condition on math ($F_{3,69} = 3.946$, $p = 0.012$, $\eta_p^2 = 0.132$) and on reaction time for numeral order judgments ($F_{3,70} = 5.526$, $p = 0.002$, $\eta_p^2 = 0.166$). The approximate arithmetic group was the only training condition that resulted in a significant improvement in math scores ($t_{17} = 4.658$, $p < 0.001$); participants in this training condition correctly solved on average 14.4 more problems after training compared to before training (i.e., correctly solved 67.3 problems in a 10-min pretest and 81.7 problems in posttest). In contrast, other training conditions did not yield a significant improvement in math scores (number comparison: $t_{15} = -0.710$, $p = 0.488$; short-term memory: $t_{17} = 1.148$, $p = 0.267$; symbol ordering: $t_{17} = 1.518$, $p = 0.150$), suggesting very little, if any, test–retest effects.

The gain scores for the non-symbolic numerical comparison test (w) did not differ as a function of training condition ($F_{3,62} = 1.352$, $p = 0.266$). However, there was a trend for improvement in the w for the groups that received approximate arithmetic and approximate number comparison training, while the short-term memory and the symbol ordering training groups showed a trend towards a decrement in precision (i.e., increase in w). In order to quantify these patterns, we ran a post hoc contrast analysis comparing the w difference scores for the first two groups (AA and non-symbolic numerical comparison) to the latter two groups (short-term memory and symbol ordering). There was a strong trend in this contrast ($t_{59} = 1.984$, $p = 0.052$), which suggests that training that involves estimation of non-symbolic numerical values led to improvement in the precision of approximate number representations compared to the two other types of training.

The gain scores in the spatial 2-back test (d') ($F_{3,70} = 0.390$, $p = 0.760$), the numeral order judgment test accuracy (% correct) ($F_{3,68} = 0.459$, $p = 0.712$), or the vocabulary test (number correct) ($F_{3,70} = 0.512$, $p = 0.676$) did not differ as a function of training condition. Participants in all four conditions showed significant improvement in the spatial 2-back performance (i.e. gain score greater than zero, $p < 0.035$ across the four groups), but the gain scores were not different across the groups, suggesting that there was a large overall test–retest effect for this particular test.

The transfer effect between non-symbolic arithmetic training and symbolic math was further assessed using an analysis of covariance (ANCOVA) approach in order to account for possible systematic biases in the pretest scores among different groups. As shown in Fig. 4, a significant effect of group was confirmed in math ($F_{3,62} = 3.737$, $p = 0.016$, $\eta_p^2 = 0.153$) with no group-by-test score interaction ($F_{3,62} = 1.13$, $p = 0.343$). This confirms the gain score analysis and further suggests that the transfer effect in math cannot be explained by systematic differences in the pretest scores across training groups. A post hoc pairwise analysis showed that this overall group effect was driven by a significantly larger transfer effect for symbolic math in the approximate arithmetic training group compared to all the other training groups—versus number comparison ($F_{1,32} = 10.667$, $p = 0.003$, $\eta_p^2 = 0.250$), versus short-term memory ($F_{1,32} = 4.701$, $p = 0.038$, $\eta_p^2 = 0.128$), and versus symbol ordering ($F_{1,30} = 4.397$, $p = 0.045$, $\eta_p^2 = 0.128$) training groups. Also consistent with the gain score analysis, an ANCOVA on the reaction time measure of numeral order judgment revealed a significant group effect ($F_{3,63} = 5.431$, $p = 0.002$, $\eta_p^2 = 0.205$) with a non-significant group-by-test score interaction ($F_{3,63} = 0.960$, $p = 0.417$). A post hoc pairwise analysis showed that the transfer effect for reaction time on the numeral order judgment test was greater in the symbol ordering training condition compared to other training conditions (approximate arithmetic, $F_{1,31} = 5.580$, $p = 0.025$, $\eta_p^2 = 0.153$; number comparison, $F_{1,31} = 13.598$, $p = 0.001$, $\eta_p^2 = 0.305$; short-term memory, $F_{1,31} = 3.535$, $p = 0.070$, $\eta_p^2 = 0.102$).

Orthogonal to the central question regarding the transfer effect, correlations between exact symbolic arithmetic performance and other test measures at pretest were computed as shown in Fig. 7. The reaction time and accuracy in the numerical order judgment test and the spatial 2-back performance measure were significant predictors of the math score ($p < 0.008$).

These results indicate that the primary source of the transfer effect, whereby approximate arithmetic training yields improved symbolic math performance, cannot be attributed solely to training that enhances the precision of approximate number representations, facilitates ordering of numerals, or to improvements in visuo-spatial short-term memory. Thus, the findings lead to the hypothesis that mental manipulation of non-symbolic numerical quantities is the key factor that leads into improvements in exact symbolic arithmetic. To further test this hypothesis we conducted Experiment 1B where we implemented an additional training condition that was as similar as possible to approximate arithmetic but did not involve mental manipulation of non-symbolic quantities. We predicted this new condition would yield less benefit for symbolic math.

2.2.3. Approximate arithmetic versus approximate comparison and matching

In Experiment 1B, participants were trained with a task (memory-based approximate comparison and matching) very similar to the approximate arithmetic training except that no mental manipulation of non-symbolic quantities was required (Fig. 5A). As in the four conditions of Experiment 1A participants in this group showed a significant improvement in the training task over six sessions (Fig. 5B; both linear and quadratic contrasts in a repeated-measures ANOVA, $p < 0.001$). We assessed the transfer effects in this approximate comparison and matching training condition in comparison to the transfer effects in the approximate arithmetic training condition from Experiment 1A. As we hypothesized, approximate arithmetic yielded greater gain scores in symbolic math compared to the memory-based approximate comparison and matching training condition ($F_{1,34} = 4.043$, $p = 0.026$, one-tailed, $\eta_p^2 = 0.098$). Moreover, this approximate comparison and matching training by itself did not yield a test–retest effect on symbolic math performance ($t_{16} = 0.781$, $p = 0.446$). As illustrated in Fig. 5C, the ANCOVA analysis reached the same conclusion showing a significant group effect between the two training conditions ($F_{1,31} = 3.941$, $p = 0.028$, one-tailed, $\eta_p^2 = 0.113$) with no group-by-test score interaction ($F_{1,31} = 2.932$, $p = 0.097$). These results provide further support for the hypothesis that mental manipulation of approximate quantities is indeed the primary source of training-induced improvements in math.

Another possibility is that participants in the approximate arithmetic training condition converted the non-symbolic arrays into verbal, exact, representations of number. For example, in an addition trial of approximate arithmetic, one might roughly estimate a first dot array with 12 dots to be "fourteen" and a second dot array with 17 dots to be "sixteen" and then search for approximately 30. Although we did not detect any verbal strategies during the experiment, and participants did not report using verbal strategies, it is possible that participants unconsciously sub-vocally converted quantities to words. Under this scenario, any benefit for symbolic math would actually be due to subvocal symbolic math practice. We addressed this question in Experiment 2 by using verbal interference to prevent verbalization during the arithmetic tasks.

3. Experiment 2

3.1. Material and methods

3.1.1. Participants and procedure

An independent group of 30 participants between 18 and 27 years of age participated for a single session in Experiment 2. Each participant received two different tasks: an exact symbolic arithmetic task and a non-symbolic approximate arithmetic task. The exact symbolic arithmetic task was similar to the pre- and posttest math test described in Experiment 1, except only one block of 7 min was given. One arithmetic problem set with hundreds of arithmetic problems was created. On a given trial, one problem was randomly selected without replacement. The approximate arithmetic task was similar to the training task used in Experiment 1A, except that the log difference was set to 0.5, 1.0, or 1.5 and there was no trial-by-trial feedback. A total of 120 approximate arithmetic trials were given and were separated into two blocks. The exact symbolic arithmetic and non-symbolic approximate arithmetic tasks were administered twice to each participant (with new problem sets) one with and one without an articulatory suppression concurrent task. The order of the tasks was counterbalanced across participants. The articulatory suppression task required participants to repeat the word "the" out loud at a rate of 1 Hz (see Logie, Gilhooly, & Wynn, 1994; Noel, Desert, Aubrun, &

Seron, 2001). The rate of articulation was facilitated by a metronome beat, and a trained experimenter observed each participant to insure that participants engaged in the articulatory suppression task and the vocal response was also recorded through a microphone.

Data from this Experiment also allow us to examine the association between individual performance in symbolic math and approximate arithmetic. This was not possible in Experiment 1 because approximate arithmetic was only administered in training and involved a titration procedure. To assess this relationship, the zero-order correlation between the exact symbolic arithmetic performance and the non-symbolic approximate arithmetic performance without verbal interference was computed (see Fig. 7).

3.2. Results and discussion

It is widely acknowledged that solving exact symbolic arithmetic problems relies on verbal processes, based on the findings that concurrent articulation of a meaningless word significantly impairs performance (Logie et al., 1994; Noel et al., 2001) and that language contributes to the representation of exact, but not, approximate numbers (Frank, Fedorenko, Lai, Saxe, & Gibson, 2012; Spelke & Tsivkin, 2001). Replicating previous findings (Logie et al., 1994; Noel et al., 2001), verbal interference led to a significant performance decrement in exact symbolic arithmetic ($t_{29} = 2.784$, $p = 0.009$). However, there was no decrement in performance for non-symbolic approximate arithmetic ($t_{29} = 0.730$, $p = 0.471$) (Fig. 6).

Average reaction time across participants for exact symbolic arithmetic with verbal interference (6.57 s) was slower than that without verbal interference (6.09 s), although this difference was not significant ($t_{29} = 1.664$, $p = 0.107$). The opposite pattern was observed in the non-symbolic approximate arithmetic task, with verbal interference (0.879 s) non-significantly faster than that without verbal interference (0.913 s) ($t_{29} = -1.444$, $p = 0.159$). Besides these effects of verbal interference, there was a positive correlation between exact symbolic arithmetic and non-symbolic approximate arithmetic performance measures in no interference conditions ($p = 0.002$) (Fig. 7).

These results indicate that it is unlikely that the participants in the approximate arithmetic training condition in Experiment 1A used a verbal encoding strategy to solve the task, and instead support the conclusion that the improvement in symbolic math observed in the approximate arithmetic training condition was due to practice with nonverbal processes.

4. General discussion

The first finding from the current study is that non-symbolic approximate arithmetic training leads to a significant improvement in exact symbolic arithmetic. These results replicate our initial study with the same paradigm (Park & Brannon, 2013). Novel to this study, we present evidence suggesting that mental manipulation of numerical quantities is the critical cognitive component shared by approximate arithmetic and exact symbolic arithmetic. Six sessions of training improved performance on each of the five training tasks. However, training that involved (a) comparing the relative magnitude of dot arrays, (b) short-term memory, (c) ordering numerals, or (d) matching and comparing dot arrays without mental addition or subtraction, did not yield any benefits for symbolic arithmetic as measured by the change in pre to post-test scores. In contrast, mentally manipulating non-symbolic arrays in processes akin to addition and subtraction yielded a clear benefit for symbolic arithmetic.

Weber fraction (w) as modeled from performance on numerical comparison tasks is widely interpreted as a measure of individual ANS competence. Numerous studies have found a correlation between w and math ability (DeWind & Brannon, 2012; Gilmore et al., 2010; Halberda et al., 2012; Halberda et al., 2008; Libertus et al., 2011; Libertus et al., 2012; Mazzocco et al., 2011a; Mazzocco et al., 2011b; Piazza et al., 2010), which has led to the hypothesis that ANS competence is a key foundation of school-learned math. However two main aspects of our results suggest that the relationship between non-symbolic numerical comparison and symbolic mathematics may be more nuanced.

First, we found no transfer from numerical comparison training to exact symbolic arithmetic. When subjects were trained with a numerical comparison task, similar to what is used to measure w, performance improved on this task over the course of training consistent with other reports (DeWind & Brannon, 2012). Additionally, compared to the non-ANS based training conditions (short-term memory and symbol ordering), numerical comparison training showed some evidence of improved w measured on a variant of this task, suggesting a near transfer effect to a different type of number comparison task (using different ranges or number, ratio, and visual properties of the dot arrays). However, despite improvement over training and some evidence of transfer to improved w, training on this approximate number comparison task did not yield any transfer for exact symbolic arithmetic. These results suggest that the correlation between w as assessed by approximate number comparison and symbolic math may be driven by a third factor, or a common source, that has causal influence on both.

Second, our results indicate that the correlation between w and math ability may not be as robust as some reports have suggested. Orthogonal to our central hypothesis, our dataset allowed us to assess potential correlations between symbolic arithmetic and a variety of other performance measures used in this study. As illustrated in Fig. 7, we found that performance on approximate arithmetic, numerical order judgment, spatial 2-back, and even vocabulary were better predictors of exact symbolic arithmetic performance than w. In fact, w was not correlated with symbolic arithmetic performance. This finding is consistent with recent studies that report small or negligible correlation between performance in non-symbolic number comparison performance and math (Castronovo & Gobel, 2012; Fuhs & McNeil, 2013; Gobel et al., 2014; Holloway & Ansari, 2009; Inglis et al., 2011; Kolkman et al., 2013; Nosworthy et al., 2013; Price et al., 2012; Sasanguie et al., 2012; Sasanguie et al., 2013; Sasanguie et al., 2013; Tibber et al., 2013; Wei et al., 2012). Moreover, many of the studies that show a reliable correlation between performance on a non-symbolic number comparison task and math did not adequately control for other non-numerical fluid processing performance measures (such as attention, memory, and executive function) that may be important for math ability (Anobile, Stievano, & Burr, 2013; Blair & Razza, 2007; Bull, Espy, & Wiebe, 2008; Clark, Pritchard, & Woodward, 2010; DeStefano & LeFevre, 2004; Geary, 2011; St Clair-Thompson & Gathercole, 2006). Some of the inconsistent findings in the relationship between non-symbolic number comparison performance and math across studies may be driven by differences in the tasks (both in the design of the non-symbolic number comparison and the type of tests used to measure math ability) and participant demographics across studies. Nevertheless, another possibility is that the variability in the literature may reflect the fact that individual precision of the ANS, as measured by w, may not be causally related to symbolic math performance in adults. This raises the possibility that w as measured by non-symbolic numerical comparison may not be a valid and reliable assay of the ANS. Consistent with this notion, a previous study found that different tasks used to measure individual ANS competence show little correlation with one another (Inglis et al., 2011). Collectively, the current study highlights that the ANS may not be a unitary system and suggests that using a single measure to describe individual ANS acuity may be perilous.

Our results also challenge a proposal that emphasizes the automaticity of symbolic number processing as a driving force for symbolic arithmetic performance. A few recent studies have suggested that automaticity in numerical symbol understanding, is a more reliable predictor of math in children and adults than performance on an approximate numerical comparison task (Castronovo & Gobel, 2012; Durand, Hulme, Larkin, & Snowling, 2005; Holloway & Ansari, 2009; Kolkman et al., 2013; Lyons & Beilock, 2011; Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Nosworthy et al., 2013; Sasanguie et al., 2012; Sasanguie et al., 2013; Sasanguie et al., 2013). Consistent with these findings, we found a strong positive correlation between performance in the exact symbolic arithmetic test and the numeral order judgment test (see Fig. 7). Despite this correlation, however, training to order triads of Arabic numerals showed little, if any, transfer to symbolic arithmetic performance in this study nor in our previous study (Park & Brannon, 2013). These findings again suggest caution and indicate that the correlation between symbol ordering and mental arithmetic may be driven by a third common factor. Alternatively, one potential explanation for the lack of the transfer effects in the numerical symbol ordering training condition is that numeral order judgment performance may only serve as a measure of a deeper cognitive construct for numerical symbol automaticity (Lyons & Beilock, personal communications). Under this scenario, numerical symbol automaticity more broadly may be causally related to both mental arithmetic and numeral order judgments, but a direct causal connection between the numeral order judgment task and mental arithmetic may be lacking. An important step to test this hypothesis would be to find a task that taps this deeper cognitive construct for numerical symbol automaticity and conduct a training study with such a task. This is an interesting avenue for future research.

Although the approximate arithmetic condition was the only condition to show transfer to exact symbolic arithmetic, it is conceivable that had we used a different number of training sessions, one or more of the other training tasks might have been effective. For instance, numerical symbol ordering or short-term memory training might require longer training before they yield benefits for symbolic arithmetic. Nonetheless, our results are meaningful because they demonstrate, that given similar amounts of training approximate arithmetic training is much more effective than the other types of training and suggests that there is a significant cognitive overlap between symbolic arithmetic and non-symbolic approximate arithmetic.

What exactly comprises this cognitive overlap, then? Conventional mathematics learned in school requires the interplay of complex cognitive systems. Even one of the simplest branches of mathematics such as arithmetic used in this study involves at least three cognitive processes (for similar ideas see Dehaene, Piazza, Pinel, & Cohen, 2003; LeFevre et al., 2010). First, arithmetic—as an abstract study of number and quantity—by definition relies on numerical quantity processing. As described in the Introduction and the beginning of this Discussion, numerous studies have now investigated and demonstrated how nonverbal numerical quantity processing may be directly or indirectly related to math ability. Second, arithmetic also heavily relies on linguistic processes. In particular, language is critical in the representation of large exact numerical value (e.g., Spelke & Tsivkin, 2001), retrieval of learned arithmetic facts (e.g., Dehaene & Cohen, 1995), and storage of temporary answers (e.g., Logie et al., 1994). Third, arithmetic as well as other mathematical skills relies on spatial attention and working memory. Developmental studies have long posited spatial working memory as an important contributor of math learning in children (Bull et al., 2008; Geary, 2011; Raghubar, Barnes, & Hecht, 2010). More specific to arithmetic, it has recently been found that the process of addition and subtraction interacts with spatial attention (Knops, Viarouge, & Dehaene, 2009; McCrink, Dehaene, & Dehaene-Lambertz, 2007).

Approximate arithmetic, used in the current study, contains two of these three important subsystems underpinning exact symbolic arithmetic. Namely, it requires nonverbal numerical quantity representation and spatial working memory. However, unlike the tasks simply involving approximate comparison and matching of numerosities, approximate arithmetic necessitates not only maintenance but also nonverbal manipulation of visual items in working memory. Thus, the finding that exact symbolic arithmetic is most effectively improved by approximate arithmetic, and not by approximate comparison or memory-based comparison and matching, suggests that the more active process of *manipulation* of mental representations is the critical mechanism underlying the observed transfer effect. That is, repeated training in approximate arithmetic may have facilitated the effective use of numerical quantity representations and their manipulation in the mental workspace, which in turn may have resulted in effective cognitive processing for exact symbolic arithmetic later during posttest.

From a practical perspective, the fact that approximate arithmetic training transfers to school-learned math leads to the exciting hypothesis that approximate arithmetic training might benefit young children who have yet to master the meaning of exact number or numerical symbols. Indeed, in a recent study Hyde, Khanum, and Spelke (2014) showed that children who solved a set of non-symbolic numerical tasks performed better on subsequent exact symbolic arithmetic problems compared to children who received other non-numerical tasks. The lack of a pre- and posttest design and short practice followed by immediate testing make direct comparisons to the current results difficult. Nevertheless, Hyde, Khanum, and Spelke (2014) study is promising in that it raises the possibility that young children who are just beginning to learn symbolic arithmetic may benefit from training that taps the ANS through approximate arithmetic.

In sum, our study demonstrates that providing adult participants with multiple sessions of approximate arithmetic training improves exact symbolic arithmetic. Results from Experiment 1 ruled out the possibility that the benefits yielded by approximate arithmetic training were a result of more general visuo-spatial short-term memory training or increases in the precision of approximate number representations as measured by w. Data from Experiment 2 ruled out the possibility that the transfer effect between approximate dot arithmetic and symbolic arithmetic was due to verbal coding. Based on these results, we propose that nonverbal numerical quantity manipulation is the key factor that drives the observed transfer of skills between approximate arithmetic and symbolic math.

Acknowledgements

We thank Deborah Moon, Rachel Roberts, Lillian Zerihun, Chandra Swanson, and Vanessa Bermudez for their assistance in data collection. This study was supported by the Duke Fundamental & Translational Neuroscience Postdoctoral Fellowship to JP and the James McDonnell Scholar Award and NIH RO1 HD059108 to EMB.

References

Anobile, G., Stievano, P., & Burr, D. C. (2013). Visual sustained attention and numerosity sensitivity correlate with math achievement in children. *Journal of Experimental Child Psychology, 116*(2), 380–391.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647–663.

Brannon, E. M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology, 16*(2), 222–229.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205–228.

Castronovo, J., & Gobel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS ONE, 7*(4), e33832.

Clark, C. A., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology, 46*(5), 1176–1191.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. Montreal: McGill University.

Dehaene, S. (1999). *The number sense: How the mind creates mathematics*. Oxford University Press.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition, 1*(1), 83–120.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*(3), 487–506.

DeStefano, D., & LeFevre, J. A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*(3), 353–386.

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience, 6*.

Dimitrov, D. M., & Rumrill, J. P. D. (2003). Pretest-posttest designs and measurement of change. *Work: A Journal of Prevention, Assessment and Rehabilitation, 20*(2), 159–165.

Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *Journal of Experimental Child Psychology, 91*(2), 113–136.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Education Testing Service: Princeton N.J.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.

Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology, 64*(1–2), 74–92.

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science, 16*(1), 136–148.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4*(2), 59–65.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology, 47*(6), 1539–1552.

Gevins, A., & Cutillo, B. (1993). Spatiotemporal dynamics of component processes in human working memory. *Electroencephalography and Clinical Neurophysiology, 87*(3), 128–143.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115*(3), 394–406.

Gobel, S. M., Watson, S. E., Lervag, A., & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate number sense, that counts. *Psychological Science, 25*(3), 789–798.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences*.

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665–668.

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*(1), 17–29.

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition, 131*(1), 92–107.

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin and Review, 18*(6), 1222–1229.

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences of the United States of America, 106*(25), 10382–10385.

Knops, A., Viarouge, A., & Dehaene, S. (2009). Dynamic representations underlying symbolic and nonsymbolic calculation: Evidence from the operational momentum effect. *Attention, Perception and Psychophysics, 71*(4), 803–821.

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

LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development, 81*(6), 1753–1767.

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*(6), 1292–1300.

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica, 141*(3), 373–379.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory and Cognition, 22*(4), 395–410.

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121*(2), 256–261.

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science*. n/a–n/a.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*(4), 1224–1237.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011b). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE, 6*(9), 1–8.

McCrink, K., Dehaene, S., & Dehaene-Lambertz, G. (2007). Moving along the number line: Operational momentum in nonsymbolic arithmetic. *Attention, Perception and Psychophysics (pre-2011), 69*(8), 1324–1333.

Merritt, D. J, DeWind, N. K, & Brannon, E. M (2012). Comparative cognition of number representation. In T. R. Zentall & E. A. Wasserman (Eds.), *The oxford handbook of comparative cognition* (2nd ed., pp. 451–476). The Oxford Handbook of Comparative Cognition.

Noel, M.-P., Desert, M., Aubrun, A., & Seron, X. (2001). Involvement of short-term memory in complex mental calculation. *Memory and Cognition, 29*(1), 34–42.

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS ONE, 8*(7), e67918.

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*(1), 33–41.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306*(5695), 499–503.

Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica, 140*(1), 50–57.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*(2), 110–122.

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

Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2013). The approximate number system is not predictive for symbolic number processing in kindergartners. *The Quarterly Journal of Experimental Psychology*, 1–26.

Sasanguie, D., Gobel, 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.

Spelke, E. S., & Tsivkin, S. (2001). Language and number: A bilingual training study. *Cognition, 78*(1), 45–88.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59*(4), 745–759.

Tibber, M. S., Manasseh, G. S., Clarke, R. C., Gagin, G., Swanbeck, S. N., Butterworth, B., et al. (2013). Sensitivity to numerosity is not a unique visuospatial psychophysical predictor of mathematical ability. *Vision Research, 89*, 1–9.

Wei, W., Yuan, H., Chen, C., & Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. *The British Journal of Educational Psychology, 82*(Pt 1), 157–181.