Developmental Science

Developmental Science 17:4 (2014), pp 537–552 DOI: 10.1111/desc.12140

PAPER

Developmental dissociation in the neural responses to simple multiplication and subtraction problems

Jérôme Prado,[1,2] Rachna Mutreja[1] and James R. Booth[1]

1. Department of Communication Sciences and Disorders, Northwestern University, USA
2. Laboratoire Langage, Cerveau et Cognition (L2C2), Centre National de la Recherche Scientifique (CNRS) and Université de Lyon, France

Abstract

Mastering single-digit arithmetic during school years is commonly thought to depend upon an increasing reliance on verbally memorized facts. An alternative model, however, posits that fluency in single-digit arithmetic might also be achieved via the increasing use of efficient calculation procedures. To test between these hypotheses, we used a cross-sectional design to measure the neural activity associated with single-digit subtraction and multiplication in 34 children from 2nd to 7th grade. The neural correlates of language and numerical processing were also identified in each child via localizer scans. Although multiplication and subtraction were undistinguishable in terms of behavior, we found a striking developmental dissociation in their neural correlates. First, we observed grade-related increases of activity for multiplication, but not for subtraction, in a language-related region of the left temporal cortex. Second, we found grade-related increases of activity for subtraction, but not for multiplication, in a region of the right parietal cortex involved in the procedural manipulation of numerical quantities. The present results suggest that fluency in simple arithmetic in children may be achieved by both increasing reliance on verbal retrieval and by greater use of efficient quantity-based procedures, depending on the operation.

Introduction

It is generally assumed that there is a developmental shift from effortful algorithmic procedures to efficient memory-based retrieval over the course of elementary education (Geary, 1996; Siegler, 1996). In support for this *fact-retrieval* hypothesis, a majority of children (up to 3rd and 4th grade) report relying on strategies such as counting (e.g. 8 − 6 = 6 + 1 + 1) or transformation [e.g. 12 − 5 = (12 − 2) − 3] to solve simple subtraction, addition and multiplication problems (Barrouillet, Mignon & Thevenot, 2008; Cooney, Swanson & Ladd, 1988; Robinson, 2001). These strategies seem to disappear in young adults, who report retrieving the answers of the same problems directly from memory (Campbell & Xue, 2001; Geary, Frensch & Wiley, 1993). Increase in the use of retrieval strategies might result from the acquisition of associations between problems and answers during arithmetic practice (Siegler & Shipley, 1995). For example, Siegler's strategy choice model posits that the repeated use of a counting strategy to solve a problem (e.g. 8 − 6) leads to an association between this problem and the answer (e.g. 2) (Siegler & Shrager, 1984). Such increases in the associative strength of a problem with its answer are believed to occur for subtraction (Siegler, 1987) and addition (Geary & Burlingham-Dupree, 1989), but even more so for multiplication which is explicitly learned by verbal rote in school (Dehaene & Cohen, 1995).

A recent behavioral study, however, challenges this developmental hypothesis. Fayol and Thevenot (2012) demonstrated that adult participants who are fluent in arithmetic still predominantly use procedures when solving single-digit subtraction and addition. The study did not make use of self-reports but instead showed that the presentation of an addition or subtraction sign prior a corresponding problem facilitates the resolution of that problem, thereby revealing the automatic activation of

Address for correspondence: Jérôme Prado, Laboratoire Langage, Cerveau et Cognition, 67 Boulevard Pinel, 69675 Bron cedex, France; e-mail: jprado@isc.cnrs.fr

abstract procedures. This effect, however, was not observed with single-digit multiplication, suggesting that multiplication problems are more likely to rely on retrieval only. Critically, this study suggests that procedures can be as efficient as direct retrieval because multiplication problems were not solved faster than subtraction and addition problems. Fayol and Thevenot's findings seem inconsistent with the view that mastering all types of simple arithmetic operations depends upon a shift towards memory-based strategies, as suggested by fact-retrieval models. Instead, they support an alternative *schema-based* view according to which arithmetic fluency might also be achieved via the automatization of algorithmic procedures (Baroody, 1983, 1984, 1994). With practice, these procedures might become so fast and efficient that they may not reach consciousness and cannot be reported as such by the participants (who might mistakenly report these problems as retrieved from memory) (Fayol & Thevenot, 2012).

Fact-retrieval and schema-based models make different predictions regarding the brain regions that are involved in arithmetic learning in school. On the one hand, fact-retrieval models predict that arithmetic learning should be associated with increasing reliance on language-related regions of the left temporo-parietal cortex, such as the middle temporal gyrus (MTG) and the angular gyrus (AG). Both of these regions are believed to support the representation and storage of math facts in memory according to a verbal code (Dehaene, Piazza, Pinel & Cohen, 2003; Prado, Mutreja, Zhang, Mehta, Desroches, Minas & Booth, 2011). On the other hand, the schema-based hypothesis assumes that children learn single-digit arithmetic by mastering calculation procedures based on the manipulation of numerical quantities. According to this view, arithmetic learning should be associated with developmental increases of activity in parietal regions supporting numerical calculation, such as the intraparietal sulcus (IPS) and the posterior superior parietal lobule (PSPL) (Dehaene *et al.*, 2003). Although this developmental increase might be observed for all problems (Baroody, 1983, 1984, 1994), Fayol and Thevenot's results suggest that this effect might be pronounced for operations that are not explicitly learned by verbal rote in school, such as subtraction or addition.

To date, developmental neuroimaging studies do not provide unequivocal evidence for either fact-retrieval or schema-based models. First, to our knowledge, only three studies have investigated age-related differences of brain activity associated with arithmetic processing in children (Cho, Metcalfe, Young, Ryali, Geary & Menon, 2012; Rivera, Reiss, Eckert & Menon, 2005; Rosenberg-Lee, Barth & Menon, 2011). These studies have found age-related increases of activity in the IPS (Rivera *et al.*, 2005) and PSPL (Cho *et al.*, 2012; Rosenberg-Lee *et al.*, 2011), thereby providing some support for the schema-based hypothesis. This is also supported by the fact that increased IPS activity has been observed in adults compared to children (Kawashima, Taira, Okita, Inoue, Tajima, Yoshida, Sasaki, Suqiura, Watanabe & Fukuda, 2004; Kucian, von Aster, Loenneker, Dietrich & Martin, 2008). Second, to our knowledge, no previous study has shown age-related increases of activity in language-related regions. However, studies have found that the hippocampus, a region involved in memory encoding, plays a role in arithmetic problem-solving in children (Cho *et al.*, 2012; Cho, Ryali, Geary & Menon, 2011; De Smedt, Holloway & Ansari, 2011; Rivera, Menon, White, Glaser & Reiss, 2002). It has been suggested that the role of the hippocampus might be to encode new arithmetic facts before their representations can be transferred to language-related regions of the left temporo-parietal cortex (De Smedt *et al.*, 2011). Although such a transfer has not yet been demonstrated, this proposal is more consistent with the fact-retrieval hypothesis. Overall, it is unclear whether developmental neuroimaging research supports fact-retrieval or schema-based models.

Critically, it remains to be determined (1) whether the regions that show developmental increases of activity in the parietal cortex are related to numerical processing, (2) whether there exists a developmental increase of activity in language-related regions of the left temporo-parietal cortex, and (3) whether any of these effects depend upon the type of operation (e.g. subtraction vs. multiplication). To answer these questions, the present cross-sectional study analyzed functional magnetic resonance imaging (fMRI) data of 34 children from 2nd to 7th grade who evaluated single-digit subtraction and multiplication problems (see Figure 1b). Most previous studies that investigated the neural bases of arithmetic have indirectly inferred the role of activated regions based on anatomy only. This is problematic because regions can be activated by a large number of cognitive processes, and activation of the parietal or temporo-parietal cortex by itself only provides weak evidence for the specific engagement of verbal or numerical processes (Poldrack, 2006). To address this issue and provide a stringent test for fact-retrieval and schema-based views of arithmetic, Regions of Interest (ROIs) involved in language and numerical processing were independently localized in each participant (see Figure 1a). Grade-related variations of brain activity related to multiplication and subtraction were then assessed within each ROI.

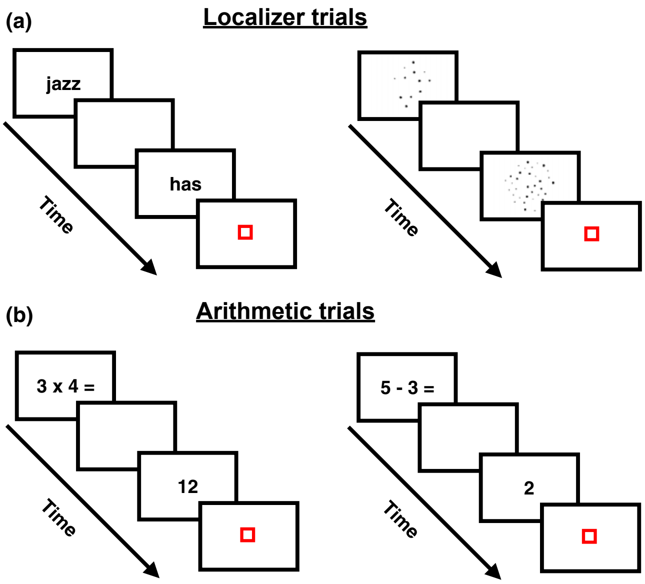

Figure 1 *Localizer and arithmetic trials. (a) In language trials (left), participants decided whether two visually presented words rhymed or not. In numerical trials (right), participants decided which of two dot arrays were composed of the larger number of dots. (b) In the arithmetic trials, participants were asked to evaluate multiplication (left) or subtraction (right) problems. Problems only involved single-digit operands and had two levels of difficulty (see text).*

Materials and methods

Participants

Children from 2nd to 7th grade were recruited from schools in the Chicago metropolitan area to participate in the study. Participants were included in the study if they (1) were native English speakers, (2) were free of past and present neurological or psychiatric disorders, (3) had no history of mathematical, reading, oral language, intelligence or attention deficits, and (4) scored higher than 80 on both performance and verbal IQ. The final sample consisted of 45 participants. Data from 11 participants were excluded because of excessive movement in the scanner (i.e. see criteria below), poor whole-brain coverage (i.e. insufficient coverage of either the temporal or parietal lobe) or low behavioral performance in the scanner (i.e. lower than 11 correct trials in any single arithmetic condition). Therefore, 34 participants from 2nd to 7th grade were included in the analyses (mean age = 11.14 yrs, range = 8.47–13.56 yrs, 13 males). None of the remaining participants had less than 18 correct trials in the smaller multiplication and smaller subtraction condition. Only three participants had less than 14 correct trials in the larger multiplication condition (12, 12 and 11 correct trials, respectively), and only 1 participant had less than 14 correct trials (11 correct trials) in the larger subtraction condition. Two participants were in 2nd grade, six in 3rd grade, four in 4th grade, nine in 5th grade, seven in 6th grade and six in 7^{th} grade. In all analyses, grade was treated as a continuous variable by taking into account the specific date within the grade year in which the participant was scanned. Although grade and age were highly correlated ($r = 0.95$, $p < .00001$), the former measure was chosen because it is arguably a more accurate measure of the level of arithmetic education than the latter. Written consent was obtained from the children and their parents or guardians. All experimental procedures were approved by the Institutional Review Board at Northwestern University.

Standardized measures

Children were administered standardized tests to measure their intellectual, mathematical, and reading abilities and ensure that there was no grade difference with respect to those age-normalized measures. First, IQ was measured with two verbal (vocabulary, similarities) and two performance subtests (block design, matrix reasoning) of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). Second, reading skill was assessed by the Letter-Word Identification and Word Attack subtests of the Woodcock Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew & Mather, 2001). Finally, we assessed mathematical ability with the Math Fluency subtest of the WJ-III (Woodcock *et al.*, 2001). Scores on performance IQ (mean = 113.8, standard deviation [*SD*] = 13.6), verbal IQ (mean = 118.4, *SD* = 17.4), Letter-Word Identification (mean = 115.8, *SD* = 17.0), Word Attack (mean = 109.5, *SD* = 9.7), and Math Fluency (mean = 103.7, *SD* = 15.4) were within or above the normal range and indicated that participants did not show impaired intellectual, mathematical or reading abilities. Most importantly, age-normalized intellectual and academic abilities were comparable across grades. Specifically, there was no significant relationship between grade and performance IQ ($r = -0.24$, $p = .17$), verbal IQ ($r = -0.29$, $p = .09$), Letter-Word Identification ($r = -0.27$, $p = .12$), Word Attack ($r = -0.28$, $p = .11$), or Math Fluency ($r = -0.19$, $p = .27$). Therefore, although the cross-sectional nature of our study does not make it possible to totally exclude the presence of other unaccounted differences between children, participants differed in terms of arithmetic education (and age) but not in terms of age-normalized intellectual, mathematical and reading abilities.

Arithmetic trials

In each trial, participants were presented with a single-digit multiplication (see Figure 1b, left) or subtraction problem (see Figure 1b, right) followed by an answer. Participants had to decide whether the answer was true or false. Problems were broken down into *smaller* and *larger* single-digit problems. Smaller multiplication problems (12 problems) were characterized by two operands that were smaller than or equal to 5 (e.g. 3×4), whereas larger multiplication problems (12 problems) were characterized by two operands that were larger than 5 (e.g. 6×7). In smaller subtraction problems (12 problems), there was a small difference (i.e. 1, 2 or 3) between the first and second term of the subtraction (e.g. $3 - 2$) (irrespective of the first term size). In larger subtraction problems (12 problems), the first term was relatively large (i.e. 6, 7, 8 or 9), as was the difference between the first and second terms (i.e. 3, 4, 5 or 6) (e.g. $9 - 4$). Each problem was repeated twice with a true answer (e.g. $3 \times 4 = 12$; $8 - 2 = 6$) and once with a false answer. This yielded 72 trials total for each problem type (36 smaller and 36 larger). False multiplication problems included the answer that would be obtained by adding or subtracting 1 to the first operand of the multiplication (e.g. $6 \times 4 = 20$ or $6 \times 4 = 28$). In false subtraction problems, false answers were constructed by adding 1 or 2 to the correct answer (e.g. $8 - 2 = 7$), or by subtracting 1 from the correct answer (e.g. $8 - 5 = 2$). For both multiplication and subtraction, problems involving 0 (e.g. 3×0; $3 - 9$; $3 - 3$), 1 as second operand (e.g. 3×1; $3 - 1$) and ties (e.g. 3×3; $6 - 3$) were not included in the main experiment. These problems were used for practice items. Twenty-four problems with a correct answer and 24 problems with a false answer were included in the practice session.

Localizer trials

Each subject was run on functional localizer scans containing language and numerical trials. In language trials (see Figure 1a, left), participants were presented with two monosyllabic English words and decided whether the words rhymed or not. Orthography and phonology were manipulated independently to ensure that judgments were not based solely on orthographic similarities between words. Specifically, the two words could have similar orthography and similar phonology (e.g. dime–lime; 12 trials), similar orthography but different phonology (e.g. pint–mint; 12 trials), different orthography but similar phonology (e.g. jazz–has; 12 trials) or different orthography and different phonology (e.g. press–list; 12 trials).

In numerical trials (see Figure 1a, right), participants were presented with two dot arrays. They had to detect which array was composed of the larger number of dots (i.e. the larger numerosity). The first dot array was composed of the larger number of dots in half of the trials, while it was composed of the smaller number of dots in the other half. The ratio between the numerosity of each dot array was varied, such that 24 trials had a ratio of 0.33 (i.e. 12 dots vs. 36 dots), 24 trials had a ratio of 0.5 (i.e. 18 dots vs. 36 dots), and 24 trials had a ratio of 0.66 (i.e. 24 dots vs. 36 dots). Six dot sizes were used and stimuli were controlled for differences in cumulative surface areas and distribution of dot sizes (Prado *et al.*, 2011). For both language and numerical trials, 12 trials of each condition were used as practice items (using different sets of stimuli).

Experimental protocol

Subjects participated in a practice session after informed consent was obtained and standardized tests were administered. During this session, they (1) learned to minimize head movement in a mock fMRI scanner and (2) practiced all trials. This was done to ensure that each task was understood and to familiarize participants with the fMRI environment. The actual scanning session took place within a week of the practice session. In the fMRI scanner, participants performed two runs of each type of arithmetic problems. They were also presented with two runs of numerical trials and one run of language trials. The order of the tasks was counterbalanced across participants. The timing and order of trial presentation within each run was optimized for estimation efficiency using optseq2 (http://surfer.nmr.mgh.harvard.edu/optseq/). Behavioral responses were recorded using a keypad below the right hand. In arithmetic trials, participants responded with their index finger if the answer to the problem was correct, and with their middle finger if it was incorrect. In language trials, participants responded with their index finger if the words rhymed and with their middle finger if they did not rhyme. In numerical trials, participants responded with their index finger if the first array was composed of more dots than the second array, and with their middle finger if the second array was composed of more dots than the first array. Stimuli were generated using E-prime software (Psychology Software Tools, Pittsburgh, PA) and projected onto a screen that was viewed by the participants through a mirror attached to the head-coil.

Stimulus timing

Stimulus timing was identical for all tasks. A trial started with the presentation of a first stimulus for 800 ms, followed by a blank screen for 200 ms. A second stimulus was then presented for 800 ms. This second stimulus was followed by a red fixation square (duration: 200 ms). Variable periods of passive visual fixation (ranging from 2600 ms to 3400 ms) were added between each trial. Furthermore, each run ended with 22 s of passive visual fixation. Fixation periods (between trials and at the end of the run) constituted the baseline. In addition, 12 null trials were included in each run. In these trials, participants were asked to press a button upon the appearance of a red square.

Behavioral analyses

Mean correct RT and mean error rate for arithmetic trials were analyzed using General Linear Model (GLM) analyses with the continuous predictor Grade and the within-subject factors Operation (Multiplication, Subtraction) and Problem size (Smaller, Larger). Mean RT and mean error rate for localizer trials were analyzed using GLM analyses with the continuous predictor Grade and the within-subject factor Trial type (Language, Numerical).

fMRI data acquisition

fMRI data were collected using a Siemens 3T TIM Trio MRI scanner (Siemens Healthcare, Erlangen, Germany). The fMRI blood oxygenation level dependent (BOLD) signal was measured with a susceptibility weighted single-shot echo planar imaging (EPI) sequence. The following parameters were used: TE = 20 ms, flip angle = 80°, matrix size = 128 × 120, field of view = 220 × 206.25 mm, slice thickness = 3 mm (0.48 mm gap), number of slices = 32, TR = 2000 ms. A high resolution T1 weighted 3D structural image was also acquired for each subject (TR = 1570 ms, TE = 3.36 ms, matrix size = 256 × 256, field of view = 240 mm, slice thickness = 1 mm, number of slices = 160).

fMRI preprocessing

Data analysis was performed using SPM8 (www.fil.ion.ucl.ac.uk/spm). After discarding the first six images of each run, functional images were corrected for slice acquisition delays, realigned to the first image of the first run, and spatially smoothed with a Gaussian filter equal to twice the voxel size ($4 \times 4 \times 8$ mm^3 full width at half maximum). Prior to normalizing images with SPM8, we used ArtRepair (http://cibsr.stanford.edu/tools/ArtRepair/ArtRepair.htm) to (1) suppress residual fluctuations due to large head motion and (2) identify volumes with significant artifact and outliers relative to the global mean signal (4% from the global mean). Volumes showing rapid scan-to-scan movements of greater than 1.5 mm were excluded via interpolation of the two nearest nonrepaired volumes. Interpolated volumes were then partially deweighted when first-level models were calculated on the repaired images. Finally, all individual brains were normalized into the same stereotactic space. Although brain anatomy changes over development (Wilke, Schmithorst & Holland, 2002), anatomical differences between children older than 7–8-year-olds and adults are small enough that they are beyond the resolution of fMRI experiments (Burgund, Kang, Kelly, Buckner, Snyder, Petersen & Schlaggar, 2002; Kang, Burgund, Lugar, Petersen & Schlaggar, 2003). Therefore, considering the age of our participants and the resolution of our data, we normalized all individual brains into the standard adult MNI space. This normalization was performed in two steps. First, after co-registration with the functional data, the structural image was segmented into gray matter, white matter and cerebrospinal fluid by using a unified segmentation algorithm (Ashburner & Friston, 2005). Second, the functional data were normalized to the MNI space using the normalization parameters estimated during unified segmentation (normalized voxel size, $2 \times 2 \times 4$ mm^3). The quality of the normalization was verified in each participant by visually checking the registration and ensuring an adequate correspondence between each individual's brain and the MNI template. All participants included in the analysis have less than 3% of the total number of volumes replaced in a single run and the number of volumes replaced did not differ between grades ($F_{1,32} = 0.01$, $p = .94$). All coordinates are reported in MNI space.

fMRI processing

Event-related statistical analysis was performed according to the general linear model. Activation was modeled as epochs with onsets time-locked to the presentation of the first stimulus in each trial and with a duration of 2 seconds (i.e. the trial duration). For arithmetic scans, hits (i.e. correct responses in problems with a true answer) were sorted by operation (multiplication, subtraction) and problem size (smaller, larger), and a regressor of no interest coded other trials (correct rejections, false alarms and misses). For localizer scans, all correct trials were sorted by trial type (language, numerical). Null trials were further modeled in a separate regressor for each localizer scan and each arithmetic task. All epochs were convolved with a canonical hemodynamic response function. The time series data were high-pass filtered (1/128 Hz), and serial correlations were corrected using an autoregressive AR (1) model.

ROI definition

Language-related and numerical processing ROIs were defined independently for each participant. However, to ensure that each subject-specific ROI fell within the same structure, the selection of subject-specific ROIs was constrained by the group activation map (Fedorenko, Hsieh, Nieto-Castanon, Whitfield-Gabrieli & Kanwisher, 2010; Prado *et al*., 2011).

First, for each subject, language trials (versus null trials) were contrasted with numerical trials (versus null trials) to generate the language localizer contrast, while the reverse comparison generated the numerical localizer contrast.

Second, these individual contrasts were submitted to one-sample *t*-tests across participants. The resulting statistical maps were thresholded for significance using a voxelwise threshold of $p < .01$ (uncorrected) and a clusterwise threshold of $p < .05$ (FWE corrected for multiple comparisons). Note that each localizer contrast was masked (inclusively) by the voxels in which the first term of the comparison was positive (inclusive mask thresholded at $p < .05$ uncorrected) to ensure that (1) all of the voxels in the language localizer contrast were more active in language than null trials and (2) all of the voxels in the numerical localizer contrast were more active in numerical than null trials. Language-related clusters were found in the left MTG (−46, −42, 2, Z = 3.14) and in the left Inferior Frontal Gyrus (IFG) (−42, 24, 10, Z = 3.23) (see Figure 2a). Numerical processing clusters were found in the right IPS (34, −50, 54, Z = 4.51) and the right PSPL (30, −76, 34, Z = 4.51) (see Figure 2b).

Third, ROI masks were defined as 6 mm spheres around the local maximum of each cluster and each ROI mask was intersected with individual localizer contrasts. For each localizer contrast, only the 30 most active voxels within the ROI mask (i.e. approximately 50% of the voxels) were selected for further analyses in each subject.

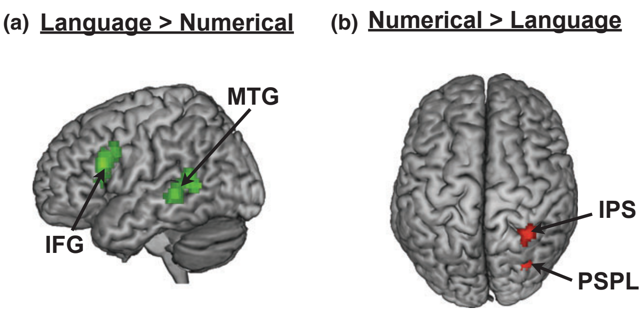

Figure 2 *Brain regions activated in the localizer trials. (a) Brain regions more activated in language than numerical trials. IFG: inferior frontal gyrus, MTG: middle temporal gyrus. (b) Brain regions more activated in numerical than language trials. IPS: Intraparietal Sulcus, PSPL: posterior superior parietal lobule. All activations are overlaid on a 3D rendering of the MNI-normalized anatomical brain.*

ROI analyses

For each participant, we calculated the average activity for each operation and problem size within an ROI by averaging the signal across the voxels within that ROI. Activity for each condition was calculated with respect to the fixation baseline (Luna, Velanova & Geier, 2010). This baseline makes few cognitive demands and any grade-related changes of activity are thus more likely to reflect changes in the mechanisms underlying arithmetic problem-solving than in those supporting visual fixation (Church, Petersen & Schlaggar, 2010).

We tested our main hypotheses regarding grade-related differences of activity in parietal and temporal regions using a GLM analysis with the continuous predictor Grade and the within-subject factors Operation (Multiplication, Subtraction), Problem size (Smaller, Larger), and ROI (MTG, PSPL, IPS). Significant or close to significant effects were further explored by more restricted GLM analyses carried out in each ROI or for each type of operation, as well as by correlation analyses.

Finally, overall patterns of brain activity across all participants were visualized and analyzed separately in lower and higher graders. This was achieved by median-splitting participants as a function of grade ($n = 17$ in each group). Brain activity of the resulting groups was analyzed via GLMs with the within-subject factors ROI, Operation and Problem size. Significant or close to significant effects were further explored by Bonferroni *t*-tests.

Whole-brain analyses

Data were also complementary analyzed with whole-brain analyses (see Supplementary Methods).

Results

Behavioral analyses

Behavioral performance is shown in Figure 3. Accuracy in the scanner was 91% on average for all arithmetic problems (range = 77%–100%), with an average of 90% (range = 71%–100%) and 92% (range = 71%–100%) for multiplication and subtraction, respectively. Response time (RT) across arithmetic problems was 1132 ms (range = 463 ms–1621 ms), with an average of 1100 ms (range = 494 ms–1699 ms) and 1165 ms (range = 431 ms–1691 ms) for multiplication and subtraction, respectively. No significant difference in either accuracy ($F_{1,32} = 2.23$, $p = .15$)

or RT ($F_{1,32}$ = 1.97, p = .17) was observed between multiplication and subtraction problems across subjects. There was, however, a problem-size effect across participants: Larger problems were associated with higher error rates and longer RTs than smaller problems (error rates: $F_{1,32}$ = 33.98, p < .00001, RTs: $F_{1,32}$ = 57.44, p < .00001) (see Figure 3a and 3c). There was also an interaction of Operation × Problem-size, indicating that the problem size effect was larger in multiplication than in subtraction (error rates: $F_{1,32}$ = 14.01, p = .0007, RTs: $F_{1,32}$ = 2, p = .17). Finally, there was a relatively modest improvement in overall performance with grade (error rates: $F_{1,32}$ = 2.94, p = .096, RTs: $F_{1,32}$ = 4.01, p = .05) (see Figure 3b and 3d). Improvements were mostly apparent for smaller problems. Performance improved with grade for smaller multiplication (error rates: r = −0.34, p = .048; RTs: r = 0.29, p = .09) and smaller subtraction (error rates: r = −0.35, p = .04; RTs: r = −0.37, p = .03). No significant improvement in performance was observed for larger multiplication (error rates: r = 0.24, p = .17; RTs: r = −0.14, p = .42) and only RTs decreased with grade for larger subtraction (error rates: r = −0.06, p = .74; RTs: r = −0.36, p = .04). Finally, there was no interaction of Grade × Operation (error rates: $F_{1,32}$ = 0.52, p = .48, RTs: $F_{1,32}$ = 1.55, p = .22), and of Grade × Problem-size (error rates: $F_{1,32}$ = 0.18, p = .67, RTs: $F_{1,32}$ = 0.20, p = .66). There was also no interaction of Grade × Operation × Problem size (error rates: $F_{1,32}$ = 1.64, p = .21, RTs: $F_{1,32}$ = 1.85, p = .18). Overall, arithmetic education affected multiplication and subtraction performance to the same extent.

Performance in the localizer trials was also high, with an average of 87% for language trials (range = 61%–98%) and 88% for numerical processing trials (range = 63%– 100%), and with an average of 1251 ms for language trials (range = 601 ms–1825 ms) and 1052 ms for numerical trials (range = 507 ms–1670 ms). No significant difference in accuracy was observed between tasks ($F_{1,32}$ = 0.66, p = .42), but language trials were significantly longer than numerical trials ($F_{1,32}$ = 43.25, p < .00001). Finally, no improvement in performance was observed over grade (error rates: $F_{1,32}$ = 0.05, p = .82; RTs: $F_{1,32}$ = 1.47, p = .23). The interaction of Grade × Trial type was also not significant (error rates: $F_{1,32}$ = 1.39, p = .25; RTs: $F_{1,32}$ = 3.56, p = .07).

ROI analyses: temporal and parietal ROIs

We ran a GLM analysis with the factors ROI (MTG, PSPL, IPS), Operation (Multiplication, Subtraction),

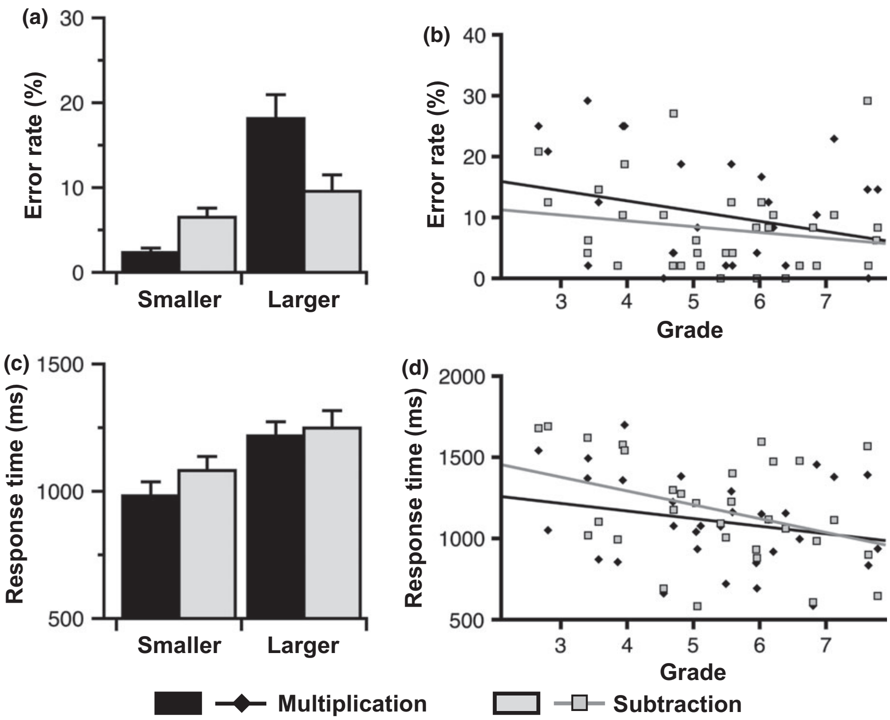

Figure 3 *Arithmetic performance. (a) Error rates as a function of operation and problem size across all subjects. (b) Error-rates as a function of operation and grade. Smaller and larger problems were combined because grade-related differences did not vary with problem size (see text). (c) RTs as a function of operation and problem size across all subjects. (d) RTs as a function of operation and grade. Smaller and larger problems were combined because grade-related differences did not vary with problem size (see text).*

Problem size (Smaller, Larger), and Grade on brain activity. First, we found a tendency for an interaction of Problem size × Grade ($F_{1,32}$ = 4.13, p = .05), revealing that grade-related increases of activity were more pronounced for smaller than larger problems across ROIs. Second, we found a main effect of ROI ($F_{2,64}$ = 26.14, p < .0001), indicating that overall levels of activity (versus baseline) differed between ROIs. Critically, however, this effect varied with operation and grade, as revealed by a significant interaction of ROI × Operation × Grade ($F_{2,64}$ = 6.65, p = .002). Post-hoc GLM analyses were then conducted separately for each ROI. These analyses revealed (1) a tendency for stronger grade-related increases of activity for multiplication than subtraction in the MTG ($F_{1,32}$ = 2.89, p = .099) (see Figure 4b), (2) significantly stronger grade-related increases of activity for subtraction than multiplication in the PSPL ($F_{1,32}$ = 6.56, p = .01) (see Figure 5b), and (3) no significant grade-related differences of activity in the IPS ($F_{1,32}$ = 1.21, p = .28) (see Supplementary Figure 1b). Problem size interacted with Operation and Grade in none of the ROIs (MTG: $F_{1,32}$ = 0.001, p = .98; PSPL: $F_{1,32}$ = 2.13, p = .15; IPS: $F_{1,32}$ = 0.052, p = .82). Furthermore, grade-related increases of activity for multiplication were stronger in the MTG than in the PSPL ($F_{1,32}$ = 4.82, p = .04), but the difference between MTG and IPS was not reliable ($F_{1,32}$ = 2.64, p = .11). Conversely, grade-related increases of activity for subtraction were stronger in the PSPL than in the MTG ($F_{1,32}$ = 5.82, p = .02), but no difference was observed between IPS and MTG ($F_{1,32}$ = 1.21, p = .28). Finally, correlation analyses revealed a positive linear relationship between activity associated with smaller multiplication and grade in the MTG (r = 0.38, p = .03). This correlation was not significant in either the PSPL (r = –0.24, p = .16), or in the IPS (r = –0.05, p = .76). Conversely, we found a positive correlation between activity associated with smaller subtraction and grade in the PSPL (r = 0.45, p = .007). This correlation was not significant in either the IPS (r = 0.23, p = .20), or in the MTG (r = –0.26, p = .14). No significant correlations were observed between activity associated with larger problems and grade in any ROI (all ps > .13). Direct comparisons between regression lines indicated that grade-related increases of activity tended to be stronger for smaller than larger multiplication in the MTG ($F_{1,32}$ = 2.88, p = .099) (see Figure 6a), whereas no difference between smaller and larger subtraction was observed in the PSPL ($F_{1,32}$ = 0.28, p = .87).

The analysis above suggests that education is associated with important differences in arithmetic-related activity of the MTG and PSPL. To visualize and evaluate the impact of these differences on overall patterns of brain activity in these ROIs, we median-split participants into two groups (i.e. lower and higher graders) on the basis of their level of education (as measured by grade). We then ran GLM analyses with the factors ROI (MTG, PSPL), Operation (Multiplication, Subtraction) and Problem size (Smaller, Larger) on brain activity of lower and higher graders. In lower graders (average grade = 4.26, range = 2.66–5.49), there was only a significant interaction of Problem size × Operation ($F_{1,16}$ = 5.11, p = .04), indicating that the Problem-size effect was larger for subtraction than for multiplication across ROIs (see Figure 4c, left and Figure 5c, left). The interaction of ROI × Operation was not significant ($F_{1,16}$ = 1.90, p = .19). In higher graders (average grade = 6.71, range = 5.57–7.99), however, the interaction of ROI × Operation approached significance ($F_{1,16}$ = 3.52, p = .08). This effect was driven by greater activity for multiplication than subtraction in the MTG (see Figure 4c, right) and greater activity for subtraction than multiplication in the PSPL (see Figure 5c, right). Post-hoc Bonferroni t-tests indicated that the effect of operation was not reliable in the MTG. However, in the PSPL, there was significantly more activity for (1) smaller subtraction than smaller multiplication (p = .01) and (2) larger subtraction than larger multiplication (p = .002).

In the IPS, GLM analyses with the factors Operation (Multiplication, Subtraction) and Problem size (Smaller, Larger) on brain activity of lower and higher graders did not reveal any significant main effects or interaction (see Supplementary Figure 1c).

ROI analyses: frontal ROI

As in our previous study on adults (Prado *et al.*, 2011), localizer trials identified a language-related ROI in the left IFG (see Supplementary Figure 2a). Brain activity in this region was analyzed via a GLM analysis with the factors Operation (Multiplication, Subtraction), Problem size (Smaller, Larger), and Grade. We only found a trend for an interaction of Problem size × Grade ($F_{1,32}$ = 3.52, p = .08), indicating greater grade-related decrease of activity for smaller than larger problems. Follow-up GLMs conducted separately for each operation showed that this effect was observed for multiplication ($F_{1,32}$ = 3.54, p = .07) (see Figure 6b), but not for subtraction ($F_{1,32}$ = 0.68, p = .42). Interestingly, this pattern seems to be the opposite to that observed in the MTG, where we found greater grade-related increases of activity for smaller than larger multiplication. To directly compare the patterns in the IFG and MTG, we ran a GLM analysis with the factors ROI (MTG, IFG), Problem size (Smaller, larger) and Grade on brain activity associated with multiplication. We found a significant interaction of ROI × Problem size × Grade ($F_{1,32}$ = 10.23, p = .003),

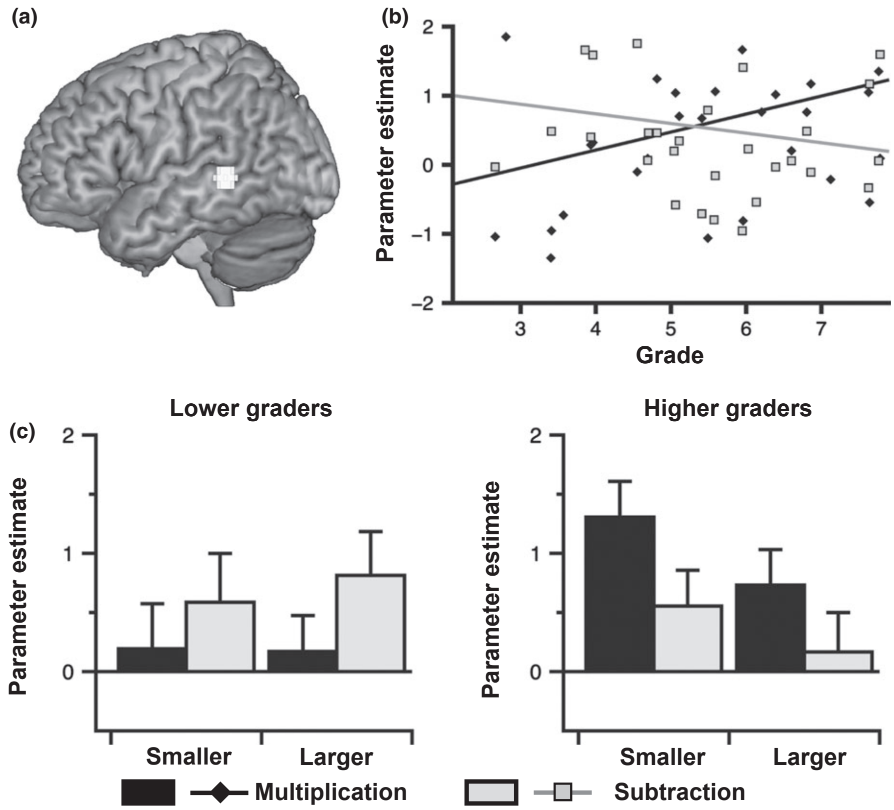

Figure 4 *Left middle temporal gyrus (MTG). (a) Location of the left MTG ROI overlaid on a 3D rendering of the MNI-normalized anatomical brain. (b) Activity in the left MTG as a function of operation and grade. Smaller and larger problems were combined because grade-related differences did not vary with problem size (see text). (c) Median-split analysis. Activity in the left MTG as a function of operation and problem size in lower (left) and higher (right) graders.*

indicating that the patterns of grade-related differences observed for smaller versus larger multiplication in the IFG and MTG were statistically different.

GLM analyses with the factors Operation (Multiplication, Subtraction) and Problem size (Smaller, Larger) on brain activity of lower and higher graders (after median split) did not reveal any significant effects or interactions (see Supplementary Figure 2c).

Whole-brain analyses

Results from whole-brain analyses are detailed in the Supplementary Results.

Discussion

In the present cross-sectional study, we used fMRI to measure the neural correlates of arithmetic processing in children from 2nd to 7th grade. Our results suggest that the neural systems underlying grade-related differences in arithmetic processing are operation-dependent.

First, we found that multiplication was associated with grade-related increases of activity in a language-related region of the left MTG, rather than in any of the numerical-processing regions of the PSPL and IPS. There is growing evidence that regions of the left mid-temporal cortex (including the left MTG) are involved in the storage of math facts in verbal memory. In adults, the left MTG and adjacent regions of the left superior temporal gyrus are typically more active when multiplication problems are compared to subtraction (Andres, Michaux & Pesenti, 2012; Andres, Pelgrims, Michaux, Olivier & Pesenti, 2011; Prado *et al.*, 2011) and addition problems (Zhou, Chen, Zang, Dong, Chen, Qiao & Gong, 2007). Activity in the left temporo-parietal cortex (including the left MTG) also increases with training when adults learn complex multiplication problems (Ischebeck, Zamarian, Egger, Schocke & Delazer, 2007). To our knowledge, the present study is the first to show grade-related differences of activity associated with arithmetic in this region in children. Previous studies, however, have associated processing impairments in the left MTG with both math and language deficits in

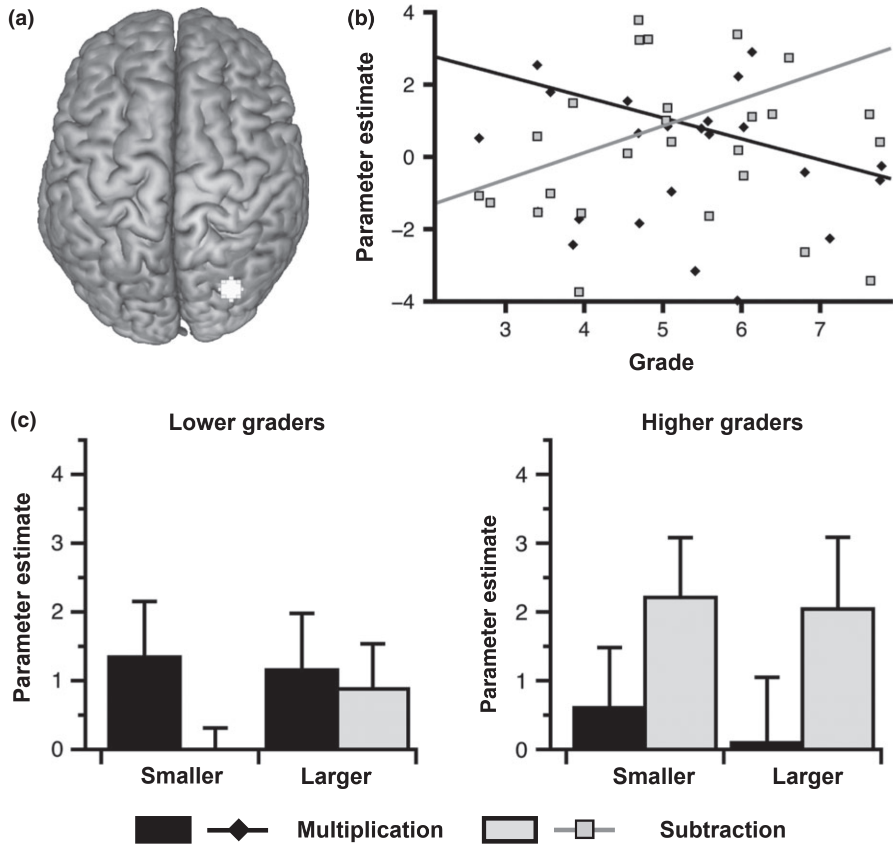

Figure 5 *Right posterior superior parietal lobule (PSPL). (a) Location of the right PSPL ROI overlaid on a 3D rendering of the MNI-normalized anatomical brain. (b) Activity in the right PSPL as a function of operation and grade. Smaller and larger problems were combined because grade-related differences did not vary with problem size (see text). (c) Median-split analysis. Activity in the right PSPL as a function of operation and problem size in lower (left) and higher (right) graders.*

children (Ashkenazi, Rosenberg-Lee, Tenison & Menon, 2012). The idea that the left MTG might be involved in the storage of math facts is consistent with its known role in the representation of lexico-semantic information (Blumenfeld, Booth & Burman, 2006; Booth, Burman, Meyer, Gitelman, Parrish & Mesulam, 2002; Fiebach, Friederici, Muller & von Cramon, 2002). Reliance on this region increases with age when children make semantic judgments on words (Chou, Booth, Bitan, Burman, Bigio, Cone, Lu & Cao, 2006), and the amount of left MTG activity has been related to the level of semantic association between two words (Chou *et al.*, 2006; Chou, Chen, Wu & Booth, 2009). Therefore, grade-related increases of activity in the left MTG might reflect increases in the strength of semantic associations between multiplication problems and their solutions with years of education. Such increases are more likely to occur for smaller than larger multiplication problems because smaller problems are generally the most practiced in elementary school (Ashcraft & Christy, 1995). This is consistent with the fact that we found greater age-related increases of activity for smaller than larger multiplication in the left MTG.

Interestingly, we found an opposite pattern of activity for multiplication in another language-related region, the left IFG. Brain activity tended to decrease with grade in this region, albeit to a greater extent for smaller than larger multiplication. This effect might reflect the decreasing demands in executive control and working memory as arithmetic becomes more and more practiced. Specifically, the left IFG has been associated with the effortful control and retrieval of verbal semantic knowledge (Bookheimer, 2002) and the selection between competing representations (Badre & Wagner, 2007). Therefore, a grade-related decrease of activity in the left IFG might reflect a decrease in the reliance on executive control. This decrease might be due to a strengthening of the associations between multiplication problems and their answers in the left MTG, an effect mostly apparent with the most practiced smaller

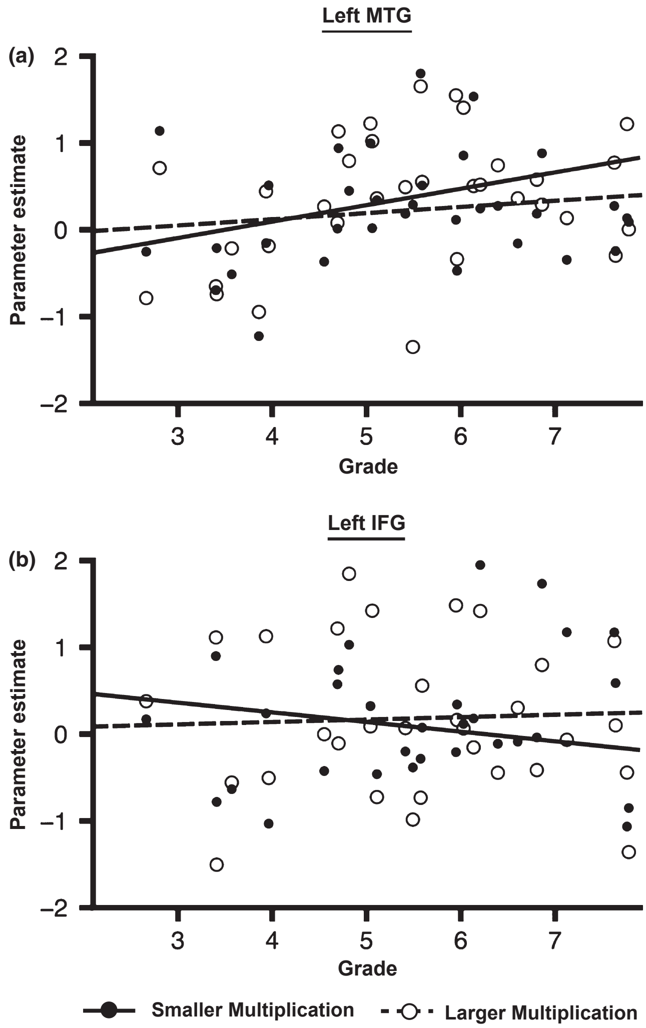

Figure 6 *Effect of multiplication problem size in language-related ROIs. (a) Activity in the left MTG as a function of multiplication problem size and grade. (b) Activity in the left IFG as a function of multiplication problem size and grade.*

problems. Overall, our results are consistent with the idea that mastering multiplication is characterized by an increase in verbal retrieval efficiency, thereby supporting the fact-retrieval hypothesis of multiplication learning (Siegler, 1988).

Unlike the results obtained with multiplication, we did not find any increases of activity in language-related regions for subtraction. Instead, subtraction was associated with grade-related increases of activity in a numerical processing region of the right PSPL. Activation in the PSPL is often reported in studies investigating the manipulation of numerical quantities in both children (Kaufmann, Wood, Rubinsten & Henik, 2011) and adults (Arsalidou & Taylor, 2011). Critically, our finding is consistent with two previous studies showing that activity associated with addition increases over development in this region. First, Rosenberg-Lee *et al*. (2011) found enhanced activity in the right PSPL as a result of only one year of education. Second, Cho *et al*. (2012) found that the only region in which brain activity increased with age in 2nd and 3rd graders was the right PSPL. In contrast to the IPS, the PSPL is not thought to be specific to quantity-based processing but is also involved in tasks involving visuo-spatial attention in children (Krinzinger, Koten, Horoufchin, Kohn, Arndt, Sahr, Konrad & Willmes, 2011) and adults (Simon, Mangin, Cohen, Le Bihan & Dehaene, 2002). It has thus been proposed that the PSPL may support some of the spatial and attentional processes associated with the procedural manipulation of numbers (Dehaene *et al*., 2003). Because subtraction performance tended to improve with grade, grade-related increases of PSPL activity might reflect the acquisition of efficient calculation procedures that might facilitate subtraction problem-solving (Baroody, 1994; Fayol & Thevenot, 2012). We speculate that such fast and efficient strategies rely on the 'recycling' of PSPL mechanisms involved in rapid shifts of attention (Dehaene & Cohen, 2007; Hubbard, Piazza, Pinel & Dehaene, 2005). Support for this idea comes from a recent fMRI study on adults. Knops, Thirion, Hubbard, Michel and Dehaene (2009) demonstrated overlapping patterns of activity in the PSPL for (1) subtraction and leftward saccades and (2) addition and rightward saccades. Therefore, the authors argued that operations such as subtraction and addition involve calculation procedures relying on leftward and rightward shifts (removing and adding quantities, respectively) along a mental number line. In the present study, the grade-related increases of activity observed in the right PSPL suggest that these attentional mechanisms might be increasingly recycled during arithmetic education.

Although grade-related increases of PSPL activity only reached significance for smaller subtraction problems, the lack of reliable increase for larger problems is difficult to interpret because it might simply result from a lack of power (i.e. only correct trials were analyzed and there were less of them in larger than smaller problems). In fact, grade-related increases did not statistically differ between smaller and larger problems in the PSPL. This indicates that such increases were not dependent upon the difficulty of the problem but only varied as a function of the type of operation (i.e. subtraction vs. multiplication). This finding is consistent with the claim that automatic procedures based on shifts along the mental number line may be triggered automatically every time a subtraction (but not a multiplication) problem is presented in adults, irrespective of the size of the operands (Fayol & Thevenot, 2012). However, because the representation of large numbers tends to be less precise than that of small numbers on the mental number line (Cohen Kadosh, Tzelgov & Henik, 2008), the accuracy of procedures based on such shifts is likely to decrease when problem size increases. Therefore, if shifts might be sufficient for solving smaller problems, solving larger problems might require the use of additional procedures, such as decomposition or transformation. One of the possible regions mediating these more effortful procedures might be the left IPS. This region has been related to procedural complexity in adults (Kong, Wang, Kwong, Vangel, Chua & Gollub, 2005), and was more activated for larger than smaller subtraction across children in the present study (see Supplementary Results).

In our view, the acquisition of efficient procedures relying on attentional shifts provides the best explanation for the grade-related increases of PSPL activity observed here. However, this does not necessarily mean that the PSPL in its entirety only supports these sorts of automatized procedures during arithmetic. For example, some regions of the PSPL are involved in spatial working memory (Wager & Smith, 2003) and might contribute to strategies requiring an effortful maintenance and manipulation of numbers, such as decomposition. These demands are likely to increase with the size of a problem. Consistent with this idea, De Smedt *et al*. (2011) found more activity for large than small subtraction in children at coordinates nearly identical (33, −70, 46 versus 33, −69, 48) to a region of the PSPL involved in manipulating items (e.g. numbers) in working memory (Wendelken, Bunge & Carter, 2008). Although we did not find a problem size effect in this region of the PSPL (see Supplementary Results), it is important to consider that our children had higher arithmetic fluency overall than the participants tested by De Smedt *et al*. (2011) (103.7 vs. 87.6 on the Math Fluency subtest of the WJ-III). Our participants might have thus found larger problems less difficult and have relied to a lesser extent on effortful procedures requiring spatial working memory. Overall, the possible heterogeneity of the PSPL highlights the importance of using localizers in future studies to identify the nature of the different PSPL subregions involved in arithmetic procedures.

We did not observe any (grade-related or overall) differences between operation or problem size in the right IPS. In this region, activity remained high and stable across both lower and higher graders (for both subtraction and multiplication) (see Supplementary Figure 1c). It is always difficult to interpret a null finding. However, we note that our results are consistent with the idea that the right IPS is involved in the automatic representation of numerical magnitude early on during development (Cantlon, Brannon, Carter & Pelphrey, 2006) and is activated whenever numbers are presented (Eger, Michel, Thirion, Amadon, Dehaene & Kleinschmidt, 2009). This might be the case for both smaller and larger problems, and even when participants retrieve multiplication facts from memory (Jost, Khader, Burke, Bien & Rosler, 2009). Our findings suggest that the right PSPL, more than the right IPS, might be the best candidate for supporting the development of procedural strategies in children.

Overall, our findings appear to contradict previous claims about the development of single-digit subtraction. It has been proposed that the repeated use of procedures in children leads to a representation of arithmetic facts in long-term memory (Geary, 1996; Siegler, 1996). For example, Siegler's strategy-choice model assumes that the repeated use of counting to solve 8 − 6 leads to an association between this problem and the answer 2 (Siegler, 1987). According to this model, activity in numerical processing regions should thus decrease over development, while activity in language-related regions should increase. Not only did we observe neither of these effects, but activity in a numerical processing region of the PSPL increased with grade. Most evidence for such fact-retrieval models comes from studies showing that, as compared to adults (Campbell & Xue, 2001), children (1) report using more procedural strategies (Barrouillet & Lépine, 2005; Barrouillet *et al*., 2008; Carpenter & Moser, 1984), and (2) exhibit longer RTs (Ashcraft & Fierman, 1982; Groen & Parkman, 1972). However, some calculation procedures (e.g. shifts along a mental number line) might be so practiced during arithmetic education that they might become as fast as retrieval and are unlikely to reach consciousness (Barrouillet & Thevenot, 2013; Fayol & Thevenot, 2012). Therefore, neither self-reports nor RTs might be able to disentangle retrieval from fast and efficient procedural strategies: participants might then mistakenly report a problem that has been calculated as being retrieved (Fayol & Thevenot, 2012). Rather than necessarily reflecting a shift from effortful procedures to efficient retrieval, the age-related decrease in error rates and rates of self-reported procedures could thus also be explained by an increase in the efficiency of procedures (at least for subtraction).

More recently, cognitive neuroscience studies have examined the neural correlates of self-reports. Findings from these studies can be reinterpreted in the light of our results. For example, Grabner and De Smedt (2011) found that problems self-reported to be retrieved in adults were associated with higher event-related (de-) synchronization in the theta band than those self-reported to be calculated. The authors interpreted this effect as reflecting retrieval from long-term memory (Grabner & De Smedt, 2011). However, theta oscillations have also been linked to shifts of attention (Green, Doesburg, Ward & McDonald, 2011; Green & McDonald, 2008) and might thus also reflect procedures based on such shifts.

Other fMRI studies have found that problems reported as being retrieved from memory are associated with different brain regions than problems reported being calculated. First, Grabner, Ansari, Koschutnig, Reishofer, Ebner and Neuper (2009) found that problems reported to be retrieved are associated with more activity than those reported to be calculated in the left AG in adults. This cluster, however, was closer to the posterior region of the left AG that is increasingly deactivated as tasks become more demanding (e.g. in larger vs. smaller problems) than to the anterior region which is thought to be involved in fact retrieval (Grabner, Ansari, Koschutnig, Reishofer & Ebner, 2013). The involvement of the left AG in this study might thus result from differences in performance (e.g. fast versus slow procedures) rather than fact retrieval *per se* (Wu, Chang, Majid, Caspers, Eickhoff & Menon, 2009). In the present study, whole-brain analyses did not reveal any grade-related differences in the left AG for multiplication (see Supplementary Results). However, differences of activity associated with multiplication were subtle and could not be detected at the whole-brain level in the left MTG either. Therefore, our experiment might have lacked statistical power at the whole-brain level to observe changes of activity in the left AG during arithmetic learning. Future studies using localizers allowing for an identification of the region of the left AG involved in retrieval might investigate this possibility. Second, children who frequently report using a retrieval strategy exhibit (1) enhanced activity (Cho *et al*., 2012) and (2) processing differences (Cho *et al*., 2011) in the hippocampus (as well as in several other brain regions) when compared to children who more frequently report using procedural strategies (e.g. decomposition, counting). Together with other studies suggesting enhanced arithmetic-related hippocampal activity in children than adults (De Smedt *et al*., 2011; Rivera *et al*., 2005), these findings have been interpreted as reflecting the mapping between arithmetic problems and their answers. However, enhanced hippocampal activity may also reflect the mapping between operations (e.g. subtraction, addition) and corresponding procedures (e.g. leftward shift, rightward shift). Indeed, the hippocampus is involved in stimulus–response mapping during both implicit and explicit learning in children (Casey, Thomas, Davidson, Kunz & Franzen, 2002; Thomas, Hunt, Vizueta, Sommer, Durston, Yang & Worden, 2004) and interacts with the striatum to enable procedural learning in adults (Albouy, Sterpenich, Balteau, Vandewalle, Desseilles, Dang-Vu, Darsaud, Ruby, Luppi, Dequeldre, Peigneux, Luxen & Maquet, 2008). Overall, none of these studies provide unequivocal evidence for the idea that small problems cannot be solved by means of fast and efficient procedures.

In conclusion, the present cross-sectional study indicates that acquiring expertise in two fundamental arithmetic operations might not rely on the same neurodevelopmental correlates. Whereas mastering single-digit multiplication may be associated with increasing reliance on verbal retrieval, learning single-digit subtraction may be associated with the acquisition of procedures. This might be because multiplication is explicitly taught by verbal rote in school, whereas subtraction is not. Alternatively, procedures might be well suited for operations such as subtraction and addition, but might be very effortful for multiplication. Rote memorization may thus be the most efficient way to solve multiplication problems (Zamarian, Ischebeck & Delazer, 2009). This idea is supported by a study showing that training complex multiplication and subtraction problems in adults leads to increased activity in retrieval-related regions of the left temporo-parietal cortex only for multiplication, even when the same training method is used for both operations (Ischebeck *et al*., 2007). In children, future studies are needed to evaluate the extent to which differences in learning methods account for the dissociation observed here. Nonetheless, our results indicate that fluency in arithmetic may be achieved by both increasing reliance on verbal retrieval and by greater use of efficient quantity-based procedures, depending on the operation.

Acknowledgements

This research was supported by a grant from the National Institute of Child Health and Human Development (HD059177) to James R. Booth. We thank Daniel Ansari and two anonymous reviewers for their helpful comments on an earlier version of this manuscript. We are also grateful for insightful discussions with Catherine Thevenot.

References

Albouy, G., Sterpenich, V., Balteau, E., Vandewalle, G., Desseilles, M., Dang-Vu, T., Darsaud, A., Ruby, P., Luppi, P.H., Dequeldre, C., Peigneux, P., Luxen, A., & Maquet, P. (2008). Both the hippocampus and striatum are involved in

consolidation of motor sequence memory. *Neuron*, **58**(2), 261–272.
Andres, M., Michaux, N., & Pesenti, M. (2012). Common substrate for mental arithmetic and finger representation in the parietal cortex. *NeuroImage*, **62**(3), 1520–1528.
Andres, M., Pelgrims, B., Michaux, N., Olivier, E., & Pesenti, M. (2011). Role of distinct parietal areas in arithmetic: an fMRI-guided TMS study. *NeuroImage*, **54** (4), 3048–3056.
Arsalidou, M., & Taylor, M.J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, **54**(3), 2382–2393.
Ashburner, J., & Friston, K.J. (2005). Unified segmentation. *NeuroImage*, **26**(3), 839–851.
Ashcraft, M.H., & Christy, K.S. (1995). The frequency of arithmetic facts in elementary texts: addition and multiplication in Grades 1–6. *Journal for Research in Mathematics Education*, **26**, 396–421.
Ashcraft, M.H., & Fierman, B.A. (1982). Mental addition in 3rd, 4th and 6th graders. *Journal of Experimental Child Psychology*, **33**, 216–234.
Ashkenazi, S., Rosenberg-Lee, M., Tenison, C., & Menon, V. (2012). Weak task-related modulation and stimulus representations during arithmetic problem solving in children with developmental dyscalculia. *Developmental Cognitive Neuroscience*, **2**(Suppl 1), S152–S166.
Badre, D., & Wagner, A.D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, **45**(13), 2883–2901.
Baroody, A.J. (1983). The development of procedural knowledge: an alternative explanation for chronometric trends of mental arithmetic. *Developmental Review*, **3**, 225–230.
Baroody, A.J. (1984). A reexamination of mental arithmetic models and data: a reply to Ashcraft. *Developmental Review*, **4**, 148–156.
Baroody, A.J. (1994). An evaluation of evidence supporting fact-retrieval models. *Learning and Individual Differences*, **6** (1), 1–36.
Barrouillet, P., & Lépine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *Journal of Experimental Child Psychology*, **91**, 183–204.
Barrouillet, P., Mignon, M., & Thevenot, C. (2008). Strategies in subtraction problem solving in children. *Journal of Experimental Child Psychology*, **99**(4), 233–251.
Barrouillet, P., & Thevenot, C. (2013). On the problem-size effect in small additions: can we really discard any counting-based account? *Cognition*, **128**(1), 35–44.
Blumenfeld, H., Booth, J., & Burman, D. (2006). Differential prefrontal-temporal neural correlates of semantic processing in children. *Brain and Language*, **99**(3), 226–235.
Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, **25**, 151–188.
Booth, J., Burman, D., Meyer, J., Gitelman, D., Parrish, T., & Mesulam, M. (2002). Modality independence of word comprehension. *Human Brain Mapping*, **16**(4), 251–261.
Burgund, E.D., Kang, H.C., Kelly, J.E., Buckner, R.L., Snyder, A.Z., Petersen, S.E., & Schlaggar, B.L. (2002). The feasibility of a common stereotactic space for children and adults in fMRI studies of development. *NeuroImage*, **17**(1), 184–200.
Campbell, J.I., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, **130** (2), 299–315.
Cantlon, J.F., Brannon, E.M., Carter, E.J., & Pelphrey, K.A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biol*, **4**(5), e125.
Carpenter, T., & Moser, J.M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, **15**, 179–202.
Casey, B.J., Thomas, K.M., Davidson, M.C., Kunz, K., & Franzen, P.L. (2002). Dissociating striatal and hippocampal function developmentally with a stimulus–response compatibility task. *Journal of Neuroscience*, **22**(19), 8647–8652.
Cho, S., Metcalfe, A.W., Young, C.B., Ryali, S., Geary, D.C., & Menon, V. (2012). Hippocampal-prefrontal engagement and dynamic causal interactions in the maturation of children's fact retrieval. *Journal of Cognitive Neuroscience*, **24**(9), 1849–1866.
Cho, S., Ryali, S., Geary, D.C., & Menon, V. (2011). How does a child solve 7 + 8? Decoding brain activity patterns associated with counting and retrieval strategies. *Developmental Science*, **14**, 989–1001.
Chou, T.L., Booth, J.R., Bitan, T., Burman, D.D., Bigio, J.D., Cone, N.E., Lu, D., & Cao, F. (2006). Developmental and skill effects on the neural correlates of semantic processing to visually presented words. *Human Brain Mapping*, **27**(11), 915–924.
Chou, T.L., Chen, C.W., Wu, M.Y., & Booth, J.R. (2009). The role of inferior frontal gyrus and inferior parietal lobule in semantic processing of Chinese characters. *Experimental Brain Research*, **198**(4), 465–475.
Church, J.A., Petersen, S.E., & Schlaggar, B.L. (2010). The 'Task B problem' and other considerations in developmental functional neuroimaging. *Human Brain Mapping*, **31**(6), 852–862.
Cohen Kadosh, R., Tzelgov, J., & Henik, A. (2008). A synesthetic walk on the mental number line: the size effect. *Cognition*, **106**, 548–557.
Cooney, J.B., Swanson, H.L., & Ladd, S.F. (1988). Acquisition of mental multiplication skill: evidence for the transition between counting and retrieval strategies. *Cognition and Instruction*, **5**, 323–345.
Dehaene, S., & Cohen, D. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, **1**, 83–120.
Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, **56**(2), 384–398.
Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, **20**(3), 487–506.
De Smedt, B., Holloway, I.D., & Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation

during calculation in children with varying levels of arithmetical fluency. *NeuroImage*, **57**(3), 771–781.

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering cortical number coding from human brain activity patterns. *Current Biology*, **19** (19), 1608–1615.

Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. *Cognition*, **123**, 392–403.

Fedorenko, E., Hsieh, P.J., Nieto-Castanon, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, **104**(2), 1177–1194.

Fiebach, C.J., Friederici, A.D., Muller, K., & von Cramon, D.Y. (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience*, **14**(1), 11–23.

Geary, D.C. (1996). *Children's mathematical development: Research and practical applications*. Washington, DC: American Psychological Association.

Geary, D.C., & Burlingham-Dupree, M. (1989). External validation of the strategy choice model for addition. *Journal of Experimental Child Psychology*, **47**, 175–192.

Geary, D.C., Frensch, P.A., & Wiley, J.G. (1993). Simple and complex mental subtraction: strategy choice and speed of processing in younger and older adults. *Psychology and Aging*, **8**, 242–256.

Grabner, R.H., Ansari, D., Koschutnig, K., Reishofer, G., & Ebner, F. (2013). The function of the left angular gyrus in mental arithmetic: evidence from the associative confusion effect. *Human Brain Mapping*, **34**(5), 1013–1024.

Grabner, R.H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., & Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia*, **47**(2), 604–608.

Grabner, R.H., & De Smedt, B. (2011). Neurophysiological evidence for the validity of verbal strategy reports in mental arithmetic. *Biological Psychology*, **87**(1), 128–136.

Green, J.J., Doesburg, S.M., Ward, L.M., & McDonald, J.J. (2011). Electrical neuroimaging of voluntary audiospatial attention: evidence for a supramodal attention control network. *Journal of Neuroscience*, **31**(10), 3560–3564.

Green, J.J., & McDonald, J.J. (2008). Electrical neuroimaging reveals timing of attentional control activity in human brain. *PLoS Biol*, **6**(4), e81.

Groen, G.J., & Parkman, J.M. (1972). A chronometric analysis of simple addition. *Psychological Review*, **79**, 329–343.

Hubbard, E., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, **6**(6), 435–448.

Ischebeck, A., Zamarian, L., Egger, K., Schocke, M., & Delazer, M. (2007). Imaging early practice effects in arithmetic. *NeuroImage*, **36**(3), 993–1003.

Jost, K., Khader, P., Burke, M., Bien, S., & Rosler, F. (2009). Dissociating the solution processes of small, large, and zero multiplications by means of fMRI. *NeuroImage*, **46**(1), 308–318.

Kang, H.C., Burgund, E.D., Lugar, H.M., Petersen, S.E., & Schlaggar, B.L. (2003). Comparison of functional activation foci in children and adults using a common stereotactic space. *NeuroImage*, **19**(1), 16–28.

Kaufmann, L., Wood, G., Rubinsten, O., & Henik, A. (2011). Meta-analyses of developmental fMRI studies investigating typical and atypical trajectories of number processing and calculation. *Developmental Neuropsychology*, **36**(6), 763–787.

Kawashima, R., Taira, M., Okita, K., Inoue, K., Tajima, N., Yoshida, H., Sasaki, T., Suqiura, M., Watanabe, J., & Fukuda, H. (2004). A functional MRI study of simple arithmetic – a comparison between children and adults. *Brain Research. Cognitive Brain Research*, **18**(3), 227–233.

Knops, A., Thirion, B., Hubbard, E.M., Michel, V., & Dehaene, S. (2009). Recruitment of an area involved in eye movements during mental arithmetic. *Science*, **324**(5934), 1583–1585.

Kong, J., Wang, C., Kwong, K., Vangel, M., Chua, E., & Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Brain Research. Cognitive Brain Research*, **22**(3), 397–405.

Krinzinger, H., Koten, J.W., Horoufchin, H., Kohn, N., Arndt, D., Sahr, K., Konrad, K., & Willmes, K. (2011). The role of finger representations and saccades for number processing: an FMRI study in children. *Frontiers in Psychology*, **2**, 373.

Kucian, K., von Aster, M., Loenneker, T., Dietrich, T., & Martin, E. (2008). Development of neural networks for exact and approximate calculation: a fMRI study. *Developmental Neuropsychology*, **33**(4), 447–473.

Luna, B., Velanova, K., & Geier, C.F. (2010). Methodological approaches in developmental neuroimaging studies. *Human Brain Mapping*, **31**(6), 863–871.

Poldrack, R.A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, **10**(2), 59–63.

Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A.S., Minas, J.E., & Booth, J.R. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Human Brain Mapping*, **32**(11), 1932–1947.

Rivera, S.M., Menon, V., White, C.D., Glaser, B., & Reiss, A.L. (2002). Functional brain activation during arithmetic processing in females with fragile X Syndrome is related to FMR1 protein expression. *Human Brain Mapping*, **16**(4), 206–218.

Rivera, S.M., Reiss, A.L., Eckert, M.A., & Menon, V. (2005). Developmental changes in mental arithmetic: evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex*, **15**(11), 1779–1790.

Robinson, K.M. (2001). The validity of verbal reports in children's subtraction. *Journal of Educational Psychology*, **93**, 211–222.

Rosenberg-Lee, M., Barth, M., & Menon, V. (2011). What difference does a year of schooling make? Maturation of brain response and connectivity between 2nd and 3rd grades during arithmetic problem solving. *NeuroImage*, **57**(3), 796–808.

Siegler, R.S. (1987). Strategy choice in subtraction. In J.A. Sloboda & D. Rogers (Eds.), *Cognitive processes in mathematics* (pp. 81–106). New York: Oxford University Press.

Siegler, R.S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, **117**, 258–275.

Siegler, R.S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R.S., & Shipley, C. (1995). Variation, selection, and cognitive change. In G. Halford & T. Simon (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Erlbaum.

Siegler, R.S., & Shrager, J. (1984). Strategy choices in addition and subtraction: how do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Lawrence Erlbaum.

Simon, O., Mangin, J., Cohen, L., Le Bihan, D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, **33** (3), 475–487.

Thomas, K.M., Hunt, R.H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & Worden, M.S. (2004). Evidence of developmental differences in implicit sequence learning: an fMRI study of children and adults. *Journal of Cognitive Neuroscience*, **16**(8), 1339–1351.

Wager, T.D., & Smith, E.E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cognive, Affective, and Behavioral Neuroscience*, **3**(4), 255–274.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. New York: The Psychological Corporation; Harcourt Brace & Company.

Wendelken, C., Bunge, S., & Carter, C. (2008). Maintaining structured information: an investigation into functions of parietal and lateral prefrontal cortices. *Neuropsychologia*, **46** (2), 665–678.

Wilke, M., Schmithorst, V.J., & Holland, S.K. (2002). Assessment of spatial normalization of whole-brain magnetic resonance images in children. *Human Brain Mapping*, **17** (1), 48–60.

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing Company.

Wu, S.S., Chang, T.T., Majid, A., Caspers, S., Eickhoff, S.B., & Menon, V. (2009). Functional heterogeneity of inferior parietal cortex during mathematical cognition assessed with cytoarchitectonic probability maps. *Cerebral Cortex*, **19**(12), 2930–2945.

Zamarian, L., Ischebeck, A., & Delazer, M. (2009). Neuroscience of learning arithmetic –evidence from brain imaging studies. *Neuroscience and Biobehavioral Reviews*, **33**(6), 909–925.

Zhou, X., Chen, C., Zang, Y., Dong, Q., Chen, C., Qiao, S., & Gong, Q. (2007). Dissociated brain organization for single-digit addition and multiplication. *NeuroImage*, **35**(2), 871–880.

Received: 1 February 2013
Accepted: 16 September 2013

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Right Intraparietal sulcus (IPS). (A) Location of the right IPS ROI overlaid on a 3D rendering of the MNI-normalized anatomical brain. (B) Activity in the right IPS as a function of operation and grade. Smaller and larger problems were combined because grade-related differences did not vary with problem size (see text). (C) Median-split analysis. Activity in the right IPS as a function of operation and problem size in lower (left) and higher (right) graders.

Figure S2. Left Inferior Frontal Gyrus (IFG). (A) Location of the left IFG ROI overlaid on a 3D rendering of the MNI-normalized anatomical brain. (B) Activity in the left IFG as a function of operation and grade. Smaller and larger problems were combined because grade-related differences did not vary with problem size (see text). (C) Median-split analysis. Activity in the left IFG as a function of operation and problem size in lower (left) and higher (right) graders.

Figure S3. Whole-brain analysis. (A) Regions showing a grade-related increase of activity for smaller subtraction. (B) Regions showing a greater grade-related increase of activity for smaller subtraction than smaller multiplication. (C) Regions showing more activity for larger than smaller subtraction across all subjects. (D) Regions showing more activity for larger than smaller multiplication across all subjects. All activations are overlaid on a 3D rendering of the MNI-normalized anatomical brain. PSPL: Posterior Superior Parietal Lobule, MOG: bilateral Middle Occipital Gyrus, Prec: Precuneus, IPS: Intraparietal Sulcus, ACC: Anterior Cingulate Cortex, DLPFC: Dorsolateral Prefrontal Cortex.