NeuroImage 153 (2017) 319–335

Contents lists available at ScienceDirect

NeuroImage

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

# When math operations have visuospatial meanings versus purely symbolic definitions: Which solving stages and brain regions are affected?☆

Aryn A. Pyke[a,*], Jon M. Fincham[a], John R. Anderson[a]

[a] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, United States

ARTICLE INFO

*Keywords:*
Math problem solving
Mental stages
Symbolic
Visuospatial representation
Graphs

ABSTRACT

How does processing differ during purely symbolic problem solving versus when mathematical operations can be mentally associated with meaningful (here, visuospatial) referents? Learners were trained on novel math operations (↓, ↑), that were defined strictly symbolically or in terms of a visuospatial interpretation (operands mapped to dimensions of shaded areas, answer = total area). During testing (scanner session), no visuospatial representations were displayed. However, we expected visuospatially-trained learners to form mental visuospatial representations for problems, and exhibit distinct activations. Since some solution intervals were long (~10 s) and visuospatial representations might only be instantiated in some stages during solving, group differences were difficult to detect when treating the solving interval as a whole. However, an HSMM-MVPA process (Anderson and Fincham, 2014a) to parse fMRI data identified four distinct problem-solving stages in each group, dubbed: 1) encode; 2) plan; 3) compute; and 4) respond. We assessed stage-specific differences across groups. During encoding, several regions implicated in general semantic processing and/or mental imagery were more active in visuospatially-trained learners, including: bilateral supramarginal, precuneus, cuneus, parahippocampus, and left middle temporal regions. Four of these regions again emerged in the computation stage: precuneus, right supramarginal/angular, left supramarginal/inferior parietal, and left parahippocampal gyrus. Thus, mental visuospatial representations may not just inform initial problem interpretation (followed by symbolic computation), but may scaffold on-going computation. In the second stage, higher activations were found among symbolically-trained solvers in frontal regions (R. medial and inferior and L. superior) and the right angular and middle temporal gyrus. Activations in contrasting regions may shed light on solvers' degree of use of symbolic versus mental visuospatial strategies, even in absence of behavioral differences.

## Introduction

Different strategies for solving a math problem can involve different types of mental representations and different neural substrates, and may have different implications for transfer and future achievement (e.g., Geary, 2011; Price et al., 2013; Pyke et al., 2015). Strategies and instructional materials involving visuospatial representations are of particular interest. Some famous mathematicians report relying heavily on mental imagery to guide their mathematical thinking (Tall, 2006; see Hadamard (1945) for a discussion of Einstein), and students' spontaneous construction and use of effective visuospatial representations can predict their math problem-solving performance (Blatto-Vallee et al., 2007; Hembree, 1992; van Garderen, 2006). Such strategies presumably contribute to correlations between spatial ability and math performance (e.g., Clements and Battista, 1992; Gathercole and Pickering, 2000; Kyttälä and Lehto, 2008; Reuhkala, 2001; for a review see Mix and Cheng (2011). That said, the use of visuospatial representations is not always helpful (Berends and van Lieshout, 2009; Booth and Koedinger, 2012; Hegarty and Kozhevnikov, 1999; Larkin and Simon, 1987; Presmeg, 1997, 2006). Such research contrasting behavioral performance across symbolic versus visuospatial strategies of various types and in various contexts is on-going (e.g., for reviews see Arcavi (2003), Hembree (1992) and Presmeg (2006)). However, far less research has investigated the neural substrates supporting visuospatial mental strategies during math problem solving. More generally, we regard visuospatial referents as a way to operationalize a more fundamental contrast of interest: between solution processes when the operations have semantic meaning versus solution processes characterized by rote calculation.

As a very simple example, one might interpret a problem like 4*5 in

☆ Support for this research was provided by National Science Foundation grants (DRL-1007945) and (DRL-1420008) to John R. Anderson.

* Corresponding author.
*E-mail address:* aryn.pyke@gmail.com (A.A. Pyke).

http://dx.doi.org/10.1016/j.neuroimage.2017.03.046
Received 25 November 2015; Accepted 20 March 2017
Available online 29 March 2017
1053-8119/ 

abstract symbolic terms, or in visuospatial terms, such that the answer represents the area of a rectangle of width 4 and height 5. Regardless of whether the problem is interpreted symbolically or visuospatially, one might compute the numerical answer using the same arithmetic steps (e.g., 4*5 = 4+4+4+4+4), however in the latter case the answer and operands are associated with visuospatial meanings.

The present research was motivated by two main questions. First, do neural substrates differ when learners solve math problems with purely symbolic procedures versus when learners can mentally associate the problems with meaningful – here, visuospatial - referents? Second, is the role of visuospatial and/or semantic processing restricted to specific stages during problem solving (e.g., initial interpretation vs. computation)? To explore these questions, we introduced learners to novel math operations (↓, ↑), that were defined either in terms of purely symbolic computation algorithms (symbolically-trained group) or included visuospatial referents (visuospatially-trained group). Both groups then mentally solved problems in a scanner. We then applied a relatively novel analysis processes to segment the fMRI solving-interval data into distinct mental stages (Anderson and Fincham, 2014a) – to allow us to assess whether there were stage-specific (vs. overall) neural differences across groups.

*Role of visuospatial math representations: Depicting relative magnitudes*

A general feature of effective visuospatial math representations (e.g., number lines, strips and graphs) is that they spatially represent the relative magnitudes of relevant quantities (e.g., as locations, lengths or areas) (e.g., Beckmann, 2004; Murata, 2008; Lewis, 1989). Prior research suggests that math learning and transfer benefit from learners having knowledge about the magnitudes of problem elements (Siegler and Ramani, 2009; Whyte and Bull, 2008), and the magnitude relations among these elements, which characterize the operation (Booth and Siegler, 2008; Slavit, 1998). For example, Booth and Siegler (2008) found that children were better able to memorize or estimate answers for specific addition facts when, during training, they had been exposed not only to the symbolic fact (e.g., 5 + 4 = 9; 18 + 16 = 34), but also to shaded bars representing the magnitudes of each operand and the sum. For arithmetic word problems, spatial representations of relative quantities with strips/bars are commonly taught and used in countries like Singapore (Beckmann, 2004; Lee et al., 2007; Lee et al., 2010) and Japan (Murata, 2008), where students exhibit high math achievement as indexed by the Programme for International Student Assessment (Organisation for Economic Co-operation and Development: OECD, 2014). Visuospatial referents have also been used in the instruction of more advanced math topics like quadratics (Hoong et al., 2010) and integral calculus, where an answer can be represented as the area of a two-dimensional region. In our experiment we investigate the impact of training learners with two–dimensional spatial referents (vs. purely symbolic procedures) on their activation patterns when they later mentally solve problems.

*Neural substrates for visuospatial mental representations in math problem solving*

We hypothesized that visuospatially-trained solvers might exhibit more activation in some regions to support processing the mental imagery and relational information inherent in (mental) visuospatial referents of problems. Some prior research on math cognition may shed light on which regions might be implicated in supporting such visuospatial mental referents.

One visuospatial math representation hypothesized to support numerical cognition is the *mental number line* for representing the magnitudes of symbolic numbers. The brain region most commonly associated with the mental number line is the horizontal intra-parietal sulcus (HIPS; for meta-analyses see Cohen Kadosh et al. (2008) and Dehaene et al. (2003), whose activation in is modulated by the numerical distance between numbers in a comparison task (e.g., small: 2 vs. 3; large: 2 vs. 9; Pinel, Dehaene, Riviere & LeBihan, 2001). Other studies have implicated the angular gyrus (AG) in mental number line processes (e.g., Cattaneo et al., 2009; Göbel et al., 2001). The posterior superior parietal lobule (PSPL) extending into the precuneus is also implicated in such numerical tasks (Pinel et al., 2001) and in domain-general visuospatial procesing, so Dehaene et al. (2003) suggest it may modulate attention along the mental number line. If the semantic roles of the HIPS, AG, PSPL and precuneus generalize beyond the canonical mental number line to other visuospatial math representations that convey magnitude relations, we might expect increased activation in such regions when solvers can mentally associate problems with more general visuospatial representations versus purely symbolic procedures.

Neuroscience studies exploring the use of more general and varied visuospatial math representations sometimes contrast conditions in which the problem stimuli themselves are symbolic versus visuospatial – for example: processing a sequences of quantities represented as digits versus sets of dots (or mixed, Piazza et al., 2007); adding digits versus dots (Venkatraman et al., 2005); comparing graphs versus equations (Thomas et al., 2010); and solving for relations depicted as bar lengths versus symbolic expressions (Lee et al., 2010).

Note, however, that we are ultimately interested in a slightly different type of contrast: when a solvers' *mental interpretation* may be either purely symbolic (e.g., 4+4+4+4+4) or visuospatial (area of rectangle) for the *same problem stimulus* (4*5). However, since mental imagery is known to share many neural substrates with perception (Ganis et al., 2004), neural differences in processing visuospatial versus symbolic math ***stimuli*** may foreshadow neural differences in processing visuospatial versus symbolic ***mental interpretations*** (of a common problem stimulus).

Interestingly, some studies emphasize the similarity of activation patterns in some regions across symbolic and visuospatial math stimuli. For example, Piazza et al. (2007) found that when learners saw a sequence of similar quantities (e.g., 18, 17, 19, 17, 19, 18,...) followed by a new quantity (20 or 50), the response to the new quantity in the HIPS and frontal regions depended on the difference between the new quantity and the familiar quantities (e.g., 20 is near; 50 is far). Importantly this effect was notation-independent – that is, it occurred regardless of whether or not the new quantity was displayed in the same format (digits or dots) as the quantities in the original sequence. They did however report that the fusiform gyrii and left lingual gyrus were sensitive to a format change.

Other stimulus-contrast studies report differences in other regions. For example, Thomas et al. (2010) reported greater activity when participants processed graphs (vs. corresponding linear and quadratic equations) not only in a bi-lateral occipital region but also in the right posterior superior parietal lobe (PSPL), precuneus, right postcentral gyrus, and right middle temporal gyrus. In such experimental designs where the stimuli differ across conditions, despite clever controls, it is not always clear which activation differences are just due to stimulus format differences versus distinct mental semantic and visuospatial solution processes.

Other studies have controlled stimulus format (e.g., symbolic expressions or word problems) but still found distinct activation patterns when learners used visuospatial *mental* strategies (e.g., Lee et al., 2007; Zago et al., 2001; Zago et al., 2008; Zarnhofer et al., 2013). For word problems, Zarnhofer et al. (2013) found that a measure reflecting the self-reported degree of use of mental visualization strategies was correlated in both hemispheres with activation in occipital regions, the lingual gyrus, calcarine gyrus, cuneus, and thalamus; and with right hemisphere (only) activation in the fusiform and superior, middle and inferior temporal gyrii.

Additional evidence comes from learners who can use mental abacus imagery to solve problems. When children trained on abacus use did exact mental addition in a scanner, several regions were more active than among

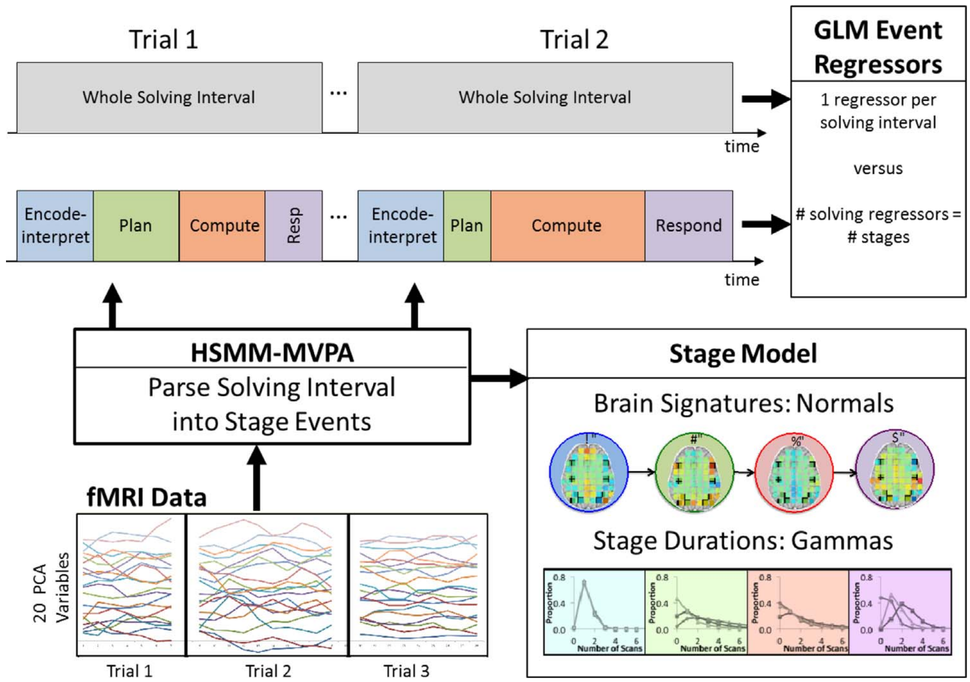


**Fig. 1.** An illustration of the HSMM–MVPA process yielding a 4-stage solution. As an alternative to treating each solving interval as a homogenous boxcar, the intermediate stages can provide finer grained event regressors in a GLM analysis of brain activity. Input to the HSMM–MVPA are scans organized within trials, where each scan consists of 20 PCA scores. The stage model characterizes each stage via 20 PCA means that define its brain signature (individual scans are normally distributed around these means) and parameters that define the gamma distribution for a stage's duration.

children who had not been abacus-trained (Du et al., 2013): medial prefrontal cortex (MPFC), right caudate, right thalamus, right superior temporal cortex, and right angular gyrus. Another study by Chen et al. (2006) also implicates the PSPL and precuneus (among others) in mental abacus imagery, though the contrast was between adding a sequence of 10 4-digit versus 1-digit numbers within abacus-experts, rather than a between group contrast between abacus experts and non-experts. Further, the difficulty contrast (4-digit > 1-digit serial addition) also implicated PSPL/precuneus regions among non-experts.

Adders unfamiliar with abacus use may nonetheless use mental imagery to visualize the multi-digit addends in a vertically aligned format to facilitate mental computation. A visualization strategy of this type (involving symbolic representations) was reported in studies by Zago et al. (2001) by learners solving horizontally presented multi-digit multiplication and addition problems (Zago et al., 2008). When learners used mental imagery to support symbolic multiplication (37×14 vs. 3×4) they had more activity in occipital and cerebellar regions, in bi-lateral IPS and fusiform, and left supramarginal, superior frontal and precentral areas. They attributed increased fusiform activity to demands of visually processing the numerical stimuli (see also Dehaene and Cohen, 1995; Schmithorst and Brown, 2004). However, the fusiform has been otherwise implicated in mental imagery during math problem solving (Zarnhofer et al., 2103) and in non-math contexts such as visualizing images for concrete nouns (D'Esposito et al., 1997), imagining faces (O'Craven and Kanwisher, 2000), and other items (Ishai et al., 2000).

In summary, based on the prior research discussed above, several regions might be expected to show increased activity to support visuospatial mental referents of math problems (e.g., among others: HIPs, AG, PSPL, precuneus, supramarginal, middle and superior temporal, fusiform, and lingual gyrus). The further question of interest was whether any such neural differences would be localized to specific stages during problem solving.

### *Parsing the solving process into a sequence of distinct solving stages (HSMM-MVPA)*

While solving a complex problem, the solver presumably goes through a sequence of 'unobservable' mental stages. As a high-level example, the solving process could be subdivided into stages for problem representation versus problem solution (e.g., Lewis, 1989). It seemed highly plausible that processing differences (e.g., temporal and neurological) across strategies might only occur in certain stages, so we wanted to explore this possibility.

One approach that has been used to gain insight on intermediate processes is to segment the solving task into explicit subtasks (e.g., problem representation: Lee et al., 2007; Lewis, 1989; vs. problem solution: Lee et al., 2010). However, in the current research, we applied a relatively novel method for parsing fMRI data that affords us insight on the solving sub-stages while learners progress naturally through them. This method combines Hidden Semi-Markov Modeling with Multi-Voxel Pattern Analysis (HSMM-MVPA; Anderson and Fincham, 2014a, 2014b).

In HSMM-MVPA, a stage identified by this process reflects a period of time with a relatively constant brain activation pattern, the stage's brain signature. Due to the temporal resolution of fMRI data, these stages will typically be on the order of a second to several seconds in length. This method has been used to discover brain signatures and stage durations in contexts where learners were all initially trained on a common solution strategy (e.g., symbolic, Anderson and Fincham, 2014a, 2014b), and when data were collapsed across learning conditions (instruction vs. discovery learning, Anderson et al., 2014). In the current research, one objective was to extend this stage-parsing process to discover possible differences in the problem-solving stage structure among learners trained symbolically versus visuospatially, by creating separate models for each group and then comparing them (i.e., in terms of the number, nature and durations of stages).

In another experiment that involved only symbolic training with problems somewhat similar to ours, Anderson and Fincham (2014a, 2014b) reported that problem-solving appeared to consist of 4 stages, which they interpreted as: i) encode, ii) plan, iii) compute; and iv) respond. Beyond the obvious requirement for input (encode) and output processes (response), their middle two stages are compatible with Mayer's (1985) suggestion that the problem solution stage is subdivided into solution planning (selecting a solution procedure) and solution execution (carrying out the computations).

The ability to find the stage structure via HSMM-MVPA affords important opportunities for additional insight and analysis. For example, beyond exploring brain activation patterns in a traditional manner that treats the solving interval as a whole (i.e., modeling it as a homogenous boxcar regressor in a GLM), our objective was also to model activation patterns within each of the intermediate problem solving stages. Thus, as shown in Fig. 1, the HSMM-MVPA furnishes stage boundaries that allow each stage to be modeled as an event in a GLM design matrix. We specifically hoped to gain insight about when a visuospatial interpretation would play a role during solving. One possibility was that distinctive (i.e., visuospatial) processing might be localized only to the initial stage(s) of solving to reveal the needed computation, but that arithmetic computation would then progress without further reference to the visuospatial interpretation. If so, brain signatures in the computation stage would be comparable across groups (symbolic and visuospatial). Alternately, the influence and use of a visuospatial mental representation might be ongoing during computation. For example, solvers might use visuospatial mental representations to support piecemeal computation - doing one arithmetic step and then referring back to their mental representation to update their progress or determine the next step.

### *The present study*

Differences in solution strategies and problem interpretations are not always evident in behavioral measures, so we sought to explore brain activation differences during mental problem solving among learners trained to form visuospatial referents versus learners trained to use purely symbolic procedures. We introduced students to two new mathematical operators designated by ↓ and ↑. During the training session (day 1, not in scanner), both groups were asked to solve the same sets of problems (e.g., 3↑4=; 5↓2=), but the symbolic versus visuospatial training differed in three key respects. First, the operators (↓ and ↑) were initially defined either via a purely symbolic formula (e.g., Fig. 2b) or via a visuospatial interpretation (e.g., Fig. 2d). These visuospatial representations conveyed magnitude relations: operands governed the dimensions of two-dimensional graphical regions and the answer was represented by total area. Second, on two training blocks, problem solving was scaffolded such that before computing the numerical answer, the symbolically-trained group had to complete the appropriate formula (e.g., Fig. 3b), whereas the visuospatially-trained group had to generate the graphical representation (Fig. 3d). Finally, on core (vs. scaffolded) training blocks, learners initially just saw the problem and had to mentally compute the answer, but, after each response (as feedback) visuospatially-trained learners saw the appropriate visuospatial representation for the problem and symbolically-trained learners saw the appropriate formula.

After mastering the processes for solving problems during the training session (outside the scanner), learners were tested the following day in the scanner. In the scanner the two groups were presented with identical problem stimuli – that is, just the problems themselves (e.g., 3↑4=), which had to be mentally solved. This problem-only presentation and mental problem-solving task matched that in the core training blocks, however, in the scanner, no visuospatial representations or formulae were displayed even as feedback. Having matched stimuli across groups in the scanner was advantageous because it allowed us to attribute any differential processing to different mental interpretations (vs. stimulus formats) (cf. Lee et al., 2010; Thomas et al., 2010; Venkatraman et al., 2005).

This work extends a previous analysis (Pyke et al., 2015), which did not include HSMM-MVPA modeling and was based on only a subset of the current data (N=40 vs. 80). That work contrasted only the visuospatial training exemplified in Fig. 2c versus the symbolic training exemplified in Fig. 2b. However, those two characterizations differ not only in the presence (or not) of a visuospatial interpretation, but also in the sequence of arithmetic computation steps required to reach an answer. We resolved this potential confound in the current work by introducing a symbolic analog (Fig. 2a) with the same computation steps as required for Fig. 2c; and a visuospatial analog (Fig. 2d) with the same computation steps as those in Fig. 2b. Thus, overall, the arithmetic steps needed to compute answers for these problems were matched across the symbolic and corresponding visuospatial conditions; all that differed was whether or not operands/answers were also associated with visuospatial dimensions/areas. Having items that include two symbolic operationalizations for our arrow operations (Fig. 2a & b), each with its own corresponding visuospatial interpretation (Fig. 2c & d, respectively) also helps make our materials more general than using a single item/operationalization.

We expected that learners privy to visuospatial referents for problems would engage in extra processing to mentally associate problems with mental visuospatial referents. Specifically, such extra processing might be expected in regions found to be especially active in conditions involving visuospatial stimulus formats and/or visuospatial mental strategies (e.g., HIPs, AG, PSPL, precuneus, supramarginal, middle and superior temporal gyrus, fusiform, and lingual gyrus). In contrast, we did not have clear expectations about whether any regions might show increased activity in the group trained to use purely symbolic procedures. Note that both groups of solvers shared a requirement to perform similar arithmetic computations (symbol manipulation) to obtain answers, thus it seemed possible that the visuospatially-trained group might simply incur extra demands to support visuospatial mental representations.

We further wanted to assess whether or not such extra processing might be limited to certain stages during problem solving (e.g., early encoding and interpretation). To understand activation differences as they might occur in distinct stages during problem solving, we generated and compared separate (symbolic vs. visuospatial) stage models based on fMRI data using HSMM-MVPA (Anderson and Fincham 2014a, 2014b).

## Method

### *Participants*

One hundred and one right-handed participants (44 female; mean age=24.2 years, SD=5.4) recruited from the university community were given course credit or payment for their participation.

### *Procedure*

The study entailed two 1.5 h sessions on consecutive days. Training was provided in Session 1, outside the scanner. Then they were tested in Session 2, which took place in a Siemens 3 T Verio Scanner.

#### *Training session*

As detailed in Pyke et al. (2015), learners first did an 8 min arithmetic pre-test (French et al., 1963), then practiced using an occluded numeric keypad to prepare them to type their answers without looking in the scanner. Half of the learners were then assigned to be symbolically-trained (half in A1SO, half in A2SO in Fig. 2) and half to be visuospatially-trained (half in A1VS, half in A2VS). Each learner was then accordingly trained on the two novel math operators: up arrow, ↑, and down arrow, ↓. These operations were initially defined via example problems: for example a symbolically-trained learner might be given the examples in Fig. 2a (or 2b) to study, whereas a visuospatially-trained learner might be given the examples in Fig. 2c (or 2d) to study.

Training problems were of Standard form $b{\uparrow}h$ = X and $b{\downarrow}h$ = X, where participants solved for X on the right (e.g., versus other possible positions for the missing value, X, in Transfer problems like, $3{\uparrow}X$ =5). During training, $b$ and $h$ were integers from 1–9 and 2–6, respectively,

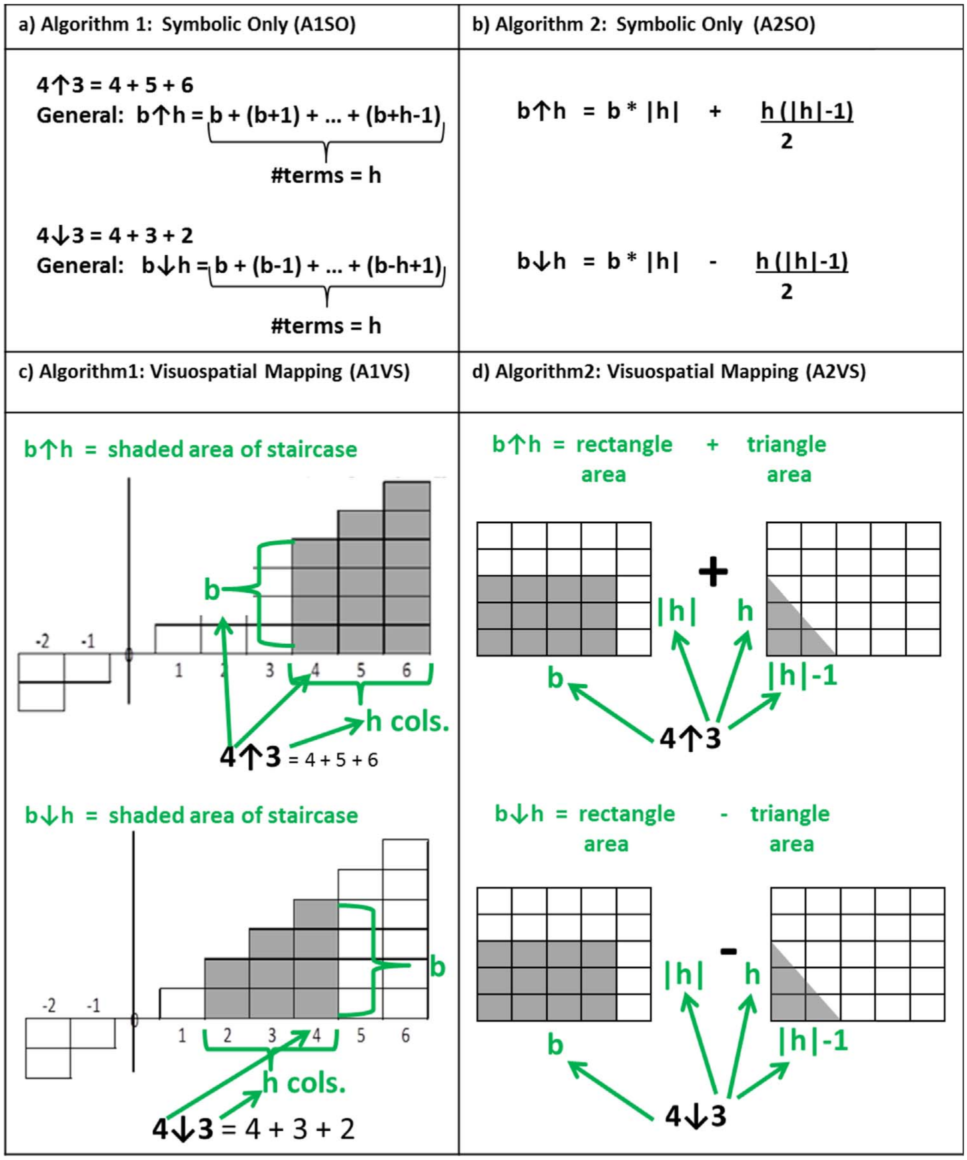


**Fig. 2.** Purely symbolic versus visuospatial characterizations for teaching arrow problems. Materials included two alternate symbolic characterizations (a) & (b), and beneath each is a corresponding visuospatial interpretation (panels c & d, respectively). The green arrows (c, d) are included for illustrative purposes but did not appear in the training examples. On actual training trials, the staircase framework in (c) extended from −5 to 15 on the horizontal axis and the grids in (d) were 9 by 9 (see Fig. 3) to accommodate problems with b in 1–9 and h in 2–6.

yielding 90 unique training problems (45 per operation). These were partitioned into 8 blocks, the first and last with 12 problems each and the rest with 11 problems each.

Standard problems (e.g., 4↑3=X) were presented one at a time on the screen in black font on a white background. Learners pressed ENTER when they had solved the problem, then they typed the answer using an occluded numeric keypad, and pressed ENTER again to submit it. A problem would time out if learners took longer than 30 s to compute an answer and/or longer than 5 s to type it. Otherwise, learners' answers appeared in blue font below the problem as they typed them. Upon pressing ENTER after their response, learners got feedback which consisted of three elements: i) their answer turned green if correct or red if incorrect; ii) the X in the problem was replaced with the correct answer, framed in a box; and iii) as per the learner's training condition (Fig. 2), either a symbolic or visuospatial representation for the particular problem appeared under the learner's answer. Thus, symbolically-trained learners, were shown as feedback the correct expression for solving the problem with the appropriate numerical values filled in (e.g., 4↑3=4+5+6 for A1SO or 4↑3= 4*|3| + 3(|3|−1)/2 for A2SO). Visuospatially-trained learners saw the shaded region(s) corresponding to the problem: either a section of a staircase (A1VS, Fig. 2c) or an appropriately scaled rectangle and triangle (A2VS, Fig. 2d), with the appropriate numbers substituted into the dimensions (e.g., the height of the rectangle would be labelled 3). Feedback was displayed for 5 s when the learner's answer was correct and 7 s when incorrect.

The first and last training blocks (1 & 8) were 'scaffolded' – that is, before mentally computing an answer for each problem, learners first had to explicitly generate its symbolic or visuospatial representation (as per their condition). The framework for this scaffolding task is illustrated in Fig. 3. Symbolically-trained learners had to produce/complete a numeric expression by typing in appropriate values for the current problem (Fig. 3a, b). The values they typed into the text box(es) appeared in blue. Visuospatially-trained learners used the mouse to delineate the relevant area(s) that should be shaded for the current problem on an initially un-shaded graph/grid framework (Fig. 3c, d). Learners trained via A1VS (see Figs. 2c, 3c) clicked on the columns in the staircase framework that were relevant to the current problem, (e.g., columns with heights 4, 5 and 6 for 4↑3=X). Learners trained via

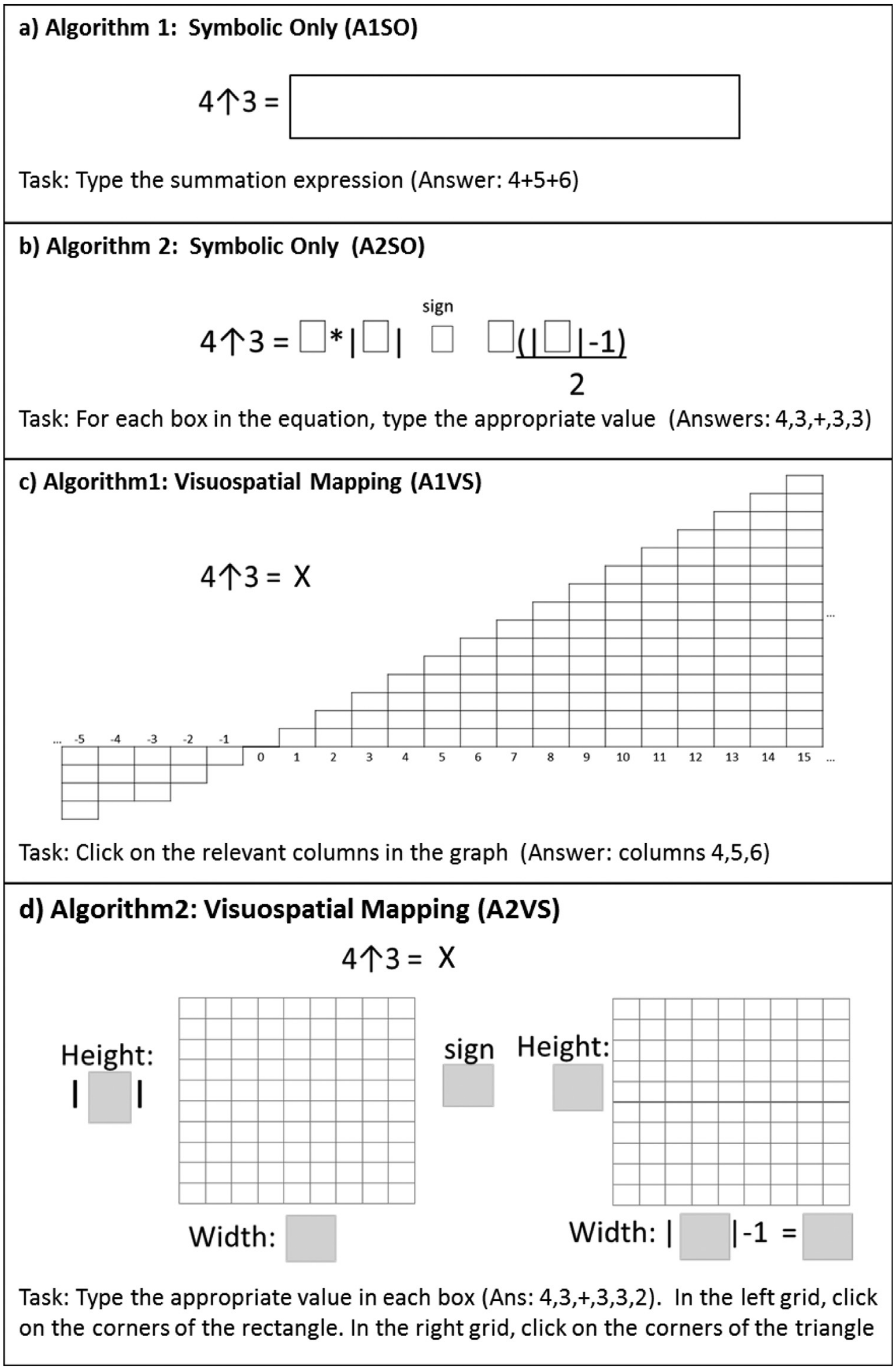


**Fig. 3.** Scaffolded training task for each of the Fig. 2 examples. Prior to mentally computing answers, learners generated purely symbolic (a, b) or visuospatial (c, d) representations (as per their condition) for each problem in the first and last training blocks.

A2VS (see Figs. 2d, 3d) clicked on points in the left grid to designate the appropriate locations of the corners for the rectangle for the current problem, and similarly clicked on points in the right grid to designate the appropriate corners of the triangle. The shapes were always anchored to the lower left corner of their respective grids so learners only had to click on the remaining corners to describe each shape. The program would shade in blue the area(s) that the visuospatially-trained learner had selected to represent the current problem.

In these scaffolded blocks, after learners had generated a symbolic or visuospatial representation for the problem (as per their condition), they would mentally compute and enter an answer as normal, and would receive the usual feedback, which included either the correct symbolic expression (in black) or visuospatial representation (shaded in grey) for their condition. Additionally, however, to help learners notice any possible mismatch between the representation they had generated and the correct one, for symbolically-trained learners, their generated expression also remained visible (in blue) above the correct one (in black). For visuospatially-trained learners, their selected area(s) remained visible in blue outline superimposed on the framework with correct region shaded in grey.

In summary, after studying definitions for the operations via symbolic or visuospatial examples (Fig. 2), training trials for symbolically- versus visuospatially-trained learners also differed in two key ways. First, in the initial and final block of training, prior to mentally computing an answer, symbolically-trained learners first had to enter/complete a symbolic expression with appropriate numbers for the current problem, whereas visuospatially-trained learners first generated the appropriate visuospatial representation for the problem within a grid framework provided on the screen. Second, feedback on every training trial included the correct symbolic expression for the problem for symbolically-trained learners, but the correct the visuospatial representation for visuospatially-trained learners.

*Testing session*

To ensure learners remembered their training, for two example problems (1↑6 and 1↓6) symbolically-trained learners were asked to write out appropriate symbolic expressions and visuospatially-trained learners had to designate the correct area(s) in a paper-and-pencil version of the task in Fig. 3c/d. Feedback was provided. The rest of the session took place in a Siemens 3 T Verio Scanner. During the structural scan, solvers did an initial practice block (8 Standard problems with *b* within 2–9 and *h* within 2–5) to help them acclimate to solving problems and using the keypad interface in the novel scanner environment. Functional scanning occurred during the remaining 8 blocks, which provide the data for the study. Each began with a Standard practice problem (not analyzed), followed by a randomly ordered mix of 4 Standard Problems (2 up-arrow problems: b↑h=X; and 2 down-arrow problems: b↓h=X) and 6 Transfer Problems (an unknown b: e.g., X↑h=a, an unknown h: e.g., b↑X=a, and 4 Relational Transfer, see Table A). Standard problems in the scanner included some with a negative operand (50% of problems, e.g., −4↑3; 4↑−3; −4↓3; 4↓−3). The solution process for such problems follows readily from the positive operand examples (Fig. 2). For example, Fig. 2c, they can intuit that a negative second operand requires them to reverse the direction (right or left) to extend the area from the starting column.

These standard problems (of the forms b↑h=X and b↓h=X), which have very well-defined solution steps (see Fig. 2), provided the data for creating models of the solving structure (stages) for visuospatially- versus symbolically-trained solvers. Within these standard format problems, operand type (negative vs. positive) was collapsed in our main analyses as it was not a factor of interest. Unknown b and unknown h problems were excluded because they do not provide visuospatial solvers with all the necessary dimensions with which to construct a well-defined visuospatial representation for the problem. This issue was shared by some relational transfer problems, which had the most variable formats and solution procedures and were characterized by quite poor performance, so they were also excluded from the models. Such transfer problems were a focus in a prior paper which was based on half of the current subjects, and which did not involve an HSMM-MVPA stage analysis (Pyke et al., 2015).

Scanner problems were presented in a similar way to problems in the core training blocks – that is, only the problem was displayed and students had to mentally solve it. In the scanner, however, feedback was always 7 s, and all students got the same type of feedback – the correct answers (no formulae or visuospatial representations were shown with the feedback in the scanner session). Thus, in contrast to training, physical presentation was identical across groups, which allows us to attribute any behavioral or neural differences to different mental (vs. stimulus) representations across groups. To distract participants from thinking about the prior problem and allow brain activity to return to a relatively constant level between problems, a repetition detection task (12 s) was inserted after the feedback on each trial: A fixation cross (3 s) was followed by a series of letters (1.25 s each) and learners were instructed to press ENTER when the same letter appeared twice in a row.

*fMRI acquisition and analysis*

*Acquisition & preprocessing.* Images were acquired using gradient echo-echo planar image (EPI) acquisition on a Siemens 3 T Verio Scanner with a 32 channel RF head coil, with 2 s repetition time (TR), 30 ms echo time (TE), 79° flip angle, and 20 cm field of view (FOV). On each TR, 34 axial slices (3.2 mm) were acquired using a 64×64 matrix. Voxels were 3.2 mm high by 3.125×3.125 mm$^2$. The anterior commissure-posterior commissure (AC-PC) line was on the 11th slice from the bottom.

Acquired images were pre-processed and analyzed using AFNI (Cox, 1996; Cox and Hyde, 1997). Functional images were motion-corrected using 6-parameter 3D registration, slice-time centered at 1 s, and normalized such that voxel time series within each block in each subject had a mean value of 100. Functional data were then co-registered to a common reference structural MRI by means of a 12-parameter 3D registration and smoothed with a 6 mm full-width-at-half-maximum 3D Gaussian filter to accommodate individual differences in anatomy.

*Whole-brain analysis.* In a preliminary traditional analysis, the whole solving interval on each relevant trial (correct Standard format problems) was treated as a single homogenous 'stage', as shown at the top of Fig. 1. Estimates of engagement (beta weights) were obtained by using a general linear model (GLM). The model involved a regressor (boxcar) for the whole solving interval, and a separate boxcar for the feedback interval. There was also a baseline model of an order-4 polynomial to account for general signal drift. The design matrix regressors were constructed by convolving the boxcar functions with the standard SPM hemodynamic function (Friston et al., 2011). Group level analyses were performed on these first-level beta estimates. Whole-brain exploratory analyses of average beta weight per voxel were conducted.

We also investigated whether there were any regions that differed in activation across groups within corresponding intermediate stages during solving (e.g., stage 1: encode). As detailed in the next subsection the stage models were generated using imaging data that had been reduced in dimensionality (e.g., principal components). However, to compare the activations across groups (visuospatial vs. symbolic) we returned to voxel-level resolution. As shown in Fig. 1, the stage structure from the HSMM-MVPA models for each group can effectively be used to subdivide the solving interval into internal 'events' for a general linear model (GLM). The regressors used corresponded to the *n* stage occupancy probabilities. As when looking at the solving interval as a single stage, regressors were convolved with the standard SPM hemodynamic function and the model also included a regressor for the feedback interval and a polynomial baseline model.

*Stage modeling (HSMM-MVPA).* We used the imaging data to parse the trials into stages using a HSMM-MVPA analysis sequence similar to that in Anderson and Fincham (2014a, 2014b). A basic overview of this process is depicted in Fig. 1.

To reduce the dimensionality of the fMRI data for MVPA, voxels in each slice were aggregated into larger 2×2 'megavoxel' regions (12, 481 total). Excluding those with an excess of extreme values yielded 8755 regions. Excluded regions with extreme values were mostly at the edges of the brain or in the top or bottom slice, possibly reflecting variations in anatomy that were not entirely dealt with in co-registration. The BOLD response in each retained megavoxel was calculated as the percent change from a linear baseline defined from the first scan of the trial (start of fixation prior to problem) to the first scan of the next trial. To estimate the shape of the signal driving the BOLD response (and

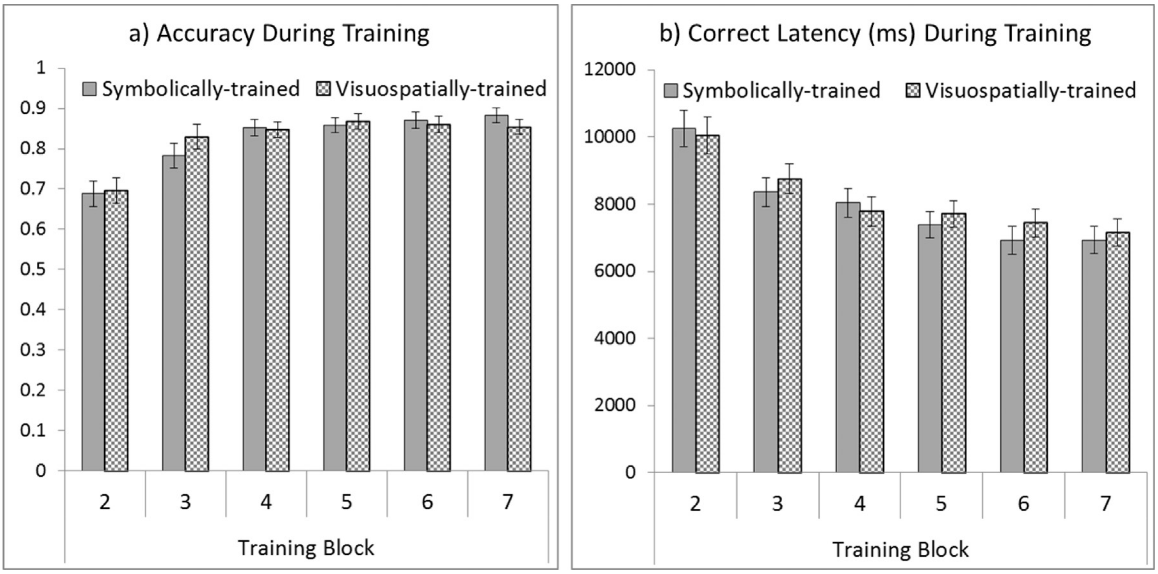


**Fig. 4.** Learning curves during the training session for both symbolically- and visuospatially-trained solvers for a) accuracy and b) correct latencies (in ms). Error bars are standard error.

reverse the lag), it was deconvolved with an SPM hemodynamic response function (difference of gammas: gamma(6,1)-gamma(16,1)/6; Friston et al., 2011) with a Wiener filter (Glover, 1999) with noise parameter of .1. Thus we had an estimate of the (unlagged) activity during each 2 s scan in each of the 8755 megavoxel regions. To detect/remove anomalous values, these activity estimates were also thresholded ( < =5% above/below baseline).

Since activities in these regions were correlated and not independent, to further reduce dimensionality we applied a spatial Principal Component Analysis (PCA) of megavoxel activity. Each megavoxel was treated as a variable that varied over scans, trials and subjects. The PCA process allows the 3-D brain activity pattern in each scan to be represented as a weighted sum of a set of basic patterns (components). The first 20 components (accounting for 54% of the total variance) were retained, and each was standardized to have a mean of 0 and standard deviation of 1.

The data of interest for modeling were scans during the correct solving of Standard format problems (b↑h=X or b↓h=X, solve for X). The PCs for these scans were normalized (mean=0, std=1) within each subject. Separate HSMM-MVPA models were generated using scans from the visuospatially- versus symbolically-trained solvers. As shown at the bottom of Fig. 1, the relevant trials for a group, each represented as 20 PC values per scan, served as input to the HSMM-MVPA stage-modeling procedure (for a detailed description see Anderson and Fincham (2014a, 2014b)). The outputs of each model (visuospatial vs. symbolic) include characterizations of the stages such as their **brain signatures** (the mean values of the 20 PC dimensions in that stage) and the distributions of durations (gamma distribution). Basically, MVPA can characterize a brain activation pattern within a temporal segment of data, and the HSMM complements the MVPA by assessing different ways to temporally segment the data stream into a particular number of stages. It estimates a set of parameters that characterize the segmentation in a way that maximizes the probability of the brain activity given a certain number of stages. The estimation process to identify the appropriate parameters is iterative. When the estimation process settles on a set of parameters, it provides a scan-by-scan parse of each trial in terms of **stage occupancies**, which are the probabilities of being in a particular stage on a particular scan.

## Results

Of the 101 participants, twenty-one did not furnish usable fMRI data: 2 failed to master the occluded keypad during training, 11 did not master solving problems during training (required threshold over 70% correct in the last 4 core training blocks), 3 did not show up for the scan session, 4 could not be scanned due to size constraints or claustrophobia, and one had a brain abnormality identified by the fMRI technician. Thus, 80 participants provided fMRI data for analysis (44% female, mean age = 23.6, SD = 4.9). Of these, 40 were trained on arrow problems symbolically (half with A1SO and half with A2SO in Fig. 2); and 40 were trained using visuospatial referents (half with A1VS, half with A2VS). Training groups did not differ significantly in age, gender distribution, or performance on an arithmetic pre-test (*ps* $< .05$). A key goal was to detect differences in the neural substrates implicated among solvers who were versus were not privy to visuospatial referents for the problem operations. Analyses were based on activations on correct Standard format problems. Latencies are from the presentation of the problem stimulus until the solver pressed ENTER after responding. The Greenhouse-Geisser adjustment was used where applicable.

### *Behavioral performance*

Our primary focus was on performance and activation patterns during the Testing Session in the scanner. However we also examined performance (accuracy and correct latency) during the core (unscaffolded) training blocks using 2 (Training: symbolic vs. visuospatial) X 6 (Training Blocks: 2–7) ANOVAs. During training, there were no main effects of training group (visuospatial vs. symbolic) on either accuracy, $F(1,78) = .02$, $\eta_p^2 = .00$, p=.896, or correct latency, $F(1,78) = .12$, $\eta_p^2 = .00$, p=.734. There were also no interactions between training group and training block for either accuracy, $F(4.16, 324.42) = .84$, $\eta_p^2 = .01$, p=.507, or correct latency, $F(3.67, 286.02) = .61$, $\eta_p^2 = .01$, p=.645. Thus, as shown in Fig. 4, learning curves were comparable across the visuospatial and symbolic training groups in the training session. In the final core training block learners achieved a mean accuracy of 86.8% (mean latency = 7045 ms).

In terms of performance in the testing (scanner) session on Standard Formation problems, a 2 (Training: symbolic vs. visuospatial) X 2 (Operand Type: positive vs. negative) ANOVA revealed that accuracy was comparable across symbolically and visuospatially-trained solvers (78% and 75%), $F(1,78) = .98$, $\eta p^2 = .01$, p=.326. The difference in correct latencies (9.8 s and 10.9 s) also did not reach significance, $F(1,78) = 3.07$, $\eta p^2 = .04$, p=.084. On problems with negative operands solvers were less accurate (63% vs. 90%); $F(1,78) = 124.77$, $\eta_p^2 = .62$, $p < .001$, and had longer latencies (12.6 s vs. 8.1 s),

$F(1,78) = 351.41$, $\eta p^2 = .82$, $p < .001$. Negative operand problems had not been practiced prior to the scanner session and often involved computations with negative numbers. Interactions of operand type and training type were not significant.

Similar performance across visuospatially- and symbolically-trained solvers may be unsurprising because the stimuli and computation steps to solve these problems were matched overall across groups. However, in the next sections we discuss brain imaging data, which can be useful for providing information not behaviorally apparent about the possible use of different cognitive representations and strategies (as in Lee et al. (2010) and Sohn et al. (2004)). Like the above behavioral analyses, the following analyses of neural activation patterns are based on correct Standard Format Problems. However, for the neural models we collapsed across operand type (positive and negative) to increase power and focus on the factor of primary interest – visuospatial versus symbolic training. That said, the HSMM-MVPA stage models of these groups yielded information about the stages (duration and activity patterns) on a trial by trial basis so we were still able to explore how operand type (positive vs. negative) affected stage durations (Section 3.3.2) and check for group activation differences in each stage among only positive or negative problems (footnote 2).

*Whole-brain analysis of the full solving interval*

This section discusses analyses of activation levels (betas) when modeling the solving interval as a whole (homogenous boxcar), as depicted at the top of Fig. 1, rather than partitioned into stages in accord with the HSMM-MVPA models. Our voxel-wise significance threshold of p=.001 yielded a whole-brain alpha estimated to be less than 0.05 by simulation for clusters with more than 20 contiguous voxels (Cox, 1996; Cox and Hyde, 1997).

*Regions mutually active in visuospatial and symbolic groups*

Although our primary interest was in regions with distinctive activations across groups, a preliminary conjunction analysis revealed extensive areas of significant activation shared across the visuospatially- and symbolically-trained solvers. As shown in Fig. 5 (top panel), these mutual activation patterns include key regions found in other studies involving numeric and arithmetic tasks (for reviews see Arsalidou and Taylor (2011) and Dehaene et al. (2003)) such as: HIPS, PSPL, insula, caudate, fusiform and (pre)frontal regions like BA10 and the lateral inferior pre-frontal cortex (LIPFC; Anderson and Fincham, 2014b).

*Regions showing significant contrasts (visuospatial vs. symbolic)*

The between-groups contrast modeling the solving interval as a whole yielded no significant clusters with greater visuospatial than symbolic activation, nor any clusters with greater symbolic than visuospatial activation. Since solution intervals were long (~10 s) and mental representations might only be instantiated during some stages during solving (e.g., to initially interpret the problem), it is perhaps not surprising that differences might become undetectable if treating the whole solving interval as a homogeneous event in a GLM. In the next section we discuss the results of the stage modeling process to parse the solving interval into a sequence of mental sub-stages. This allowed us assess whether there were any activation differences across groups local to each stage.

*Stage models: visuospatial versus symbolic*

Using an HSMM-MVPA methodology (Anderson and Fincham, 2014a), separate models were created to discover the temporal structure (sequence of stages with distinctive brain signatures) of the problem-solving process for visuospatially- versus symbolically-trained learners. This allowed us to discover the number of distinctive stages in each group, the durations of these stages and their brain activation patterns, and to compare these characteristics across groups.

*Number of stages*

To determine the appropriate number of stages for a group (say, visuospatial), we created five hypothetical models, with the number of stages ranging from 1 to 5. To select which model (how many stages) was most appropriate, we used split-half validation. An n-stage model was created based on half the learners in a group, and the fit (likelihood of the data) was assessed for the other half of the learners who were excluded from the model. For the visuospatial group, the model subjects (N=20) and test subjects (N=20) each consisted of 10 learners trained via Fig. 2b and 10 via Fig. 2c. Similarly, for the symbolic group the model and test sets each contained an equal number of learners trained via Fig. 2a and Fig. 2b.

This process was repeated for an n+1 stage model. If a significant majority (sign test: > = 15 of 20) of the test subjects (i.e., the half excluded from the model) were better fit by the n+1 than the n stage model, then the n+1 stage model was preferred.

The split-half analysis process was applied to each group (visuospatial vs. symbolic) separately, and in each case indicated a 4-stage model was appropriate.[1] For symbolically-trained subjects, the number of test subjects with better fits for *n+1* versus *n*-stage models, for *n*=1–5 were, respectively: 19 subjects (2-stages better than 1), 18 (3-stages better than 2), 15 (4 stages better than 3), and 13 (*ns*, 5 stages *not* better than 4). For the visuospatial group the numbers were: 17 (2 stages better than 1), 16 (3 stages better than 2), 16 (4 stages better than 3), and 9 (*ns*: 5 stages *not* better than 4). The 4-stage models are compatible with a model found for another experiment with only symbolic learners solving a subset of similar problems (Anderson and Fincham, 2014a; Anderson and Fincham, 2014b). In keeping with that work, the stages were interpreted as: i) encode; ii) plan; iii) compute; and iv) respond. To support the interpretation that the four stages in two current groups (visuospatial and symbolic) roughly corresponded in character, we verified that correlations of brain signatures (principle components) were higher between corresponding stages (*r* values ranged from .50 to .97) than between non-aligned stages (*r* values ranged from −.59 to .35).

*Duration of stages*

Within each group (model), stage durations varied from trial to trial. Mean stage durations for correct Standard problems (upon which the models were based) are displayed in Fig. 6. The models were generated using data from all Standard Format problems (collapsed across positive and negative operands), but since they yield stage duration estimates on a trial by trial basis, this allowed us to also break down durations by operand type. Durations were analyzed with a 2 (Training: visuospatial vs. symbolic) X 4 (Stage: encode, plan, compute, respond) X 2 (Operand Type: positive vs. negative) ANOVA. The Greenhouse-Geisser adjustment was used where applicable. There was an interaction of stage with operand type, $F(1.75,136.56) = 47.06$, $\eta p2 = .38$, $p < .001$. Problems with negative operands exhibited only slightly longer encoding and response stages (< 0.5 s), likely because there was an extra negative sign character in the problem to encode, and sometimes an extra negative sign character in the response

[1] A leave-out-out-cross-validation (LOOCV) approach was also explored, but, when the same data is used in different plies for estimation and prediction LOOCV has complications due to non-independence of the plies (Bengio and Grandvalet, 2004). This approach suggested there might be as many as 6 solving stages, but examination revealed that the 6 stage models included 3 stages that closely matched the first 3 stages in the corresponding four-stage model (mean PCA correlation = 0.95), and two final stages that were quite similar to each other and to the last stage in the 4-stage models (mean PCA r = 0.95). The main difference in the 6 stage models was the inclusion of an extra stage that straddled (was similar to) two of the stages in the 4-stage models (e.g., Spatial stage 3 of 6 correlated 0.44 with stage 2 of 4 and 0.47 with stage 3 of 4). Some stage transitions occur within a scan so activity patterns in such scans are hybrids of the straddled stages. Treating hybrid scans as separate stages in their own right may sometimes improve fit, but, our focus was on identifying a set of distinct stages in like the 4-stage models supported by the split-half approach (max. inter-state PCA correlation in 4-stage models = 0.12 vs. 0.82 in 6-stage models).

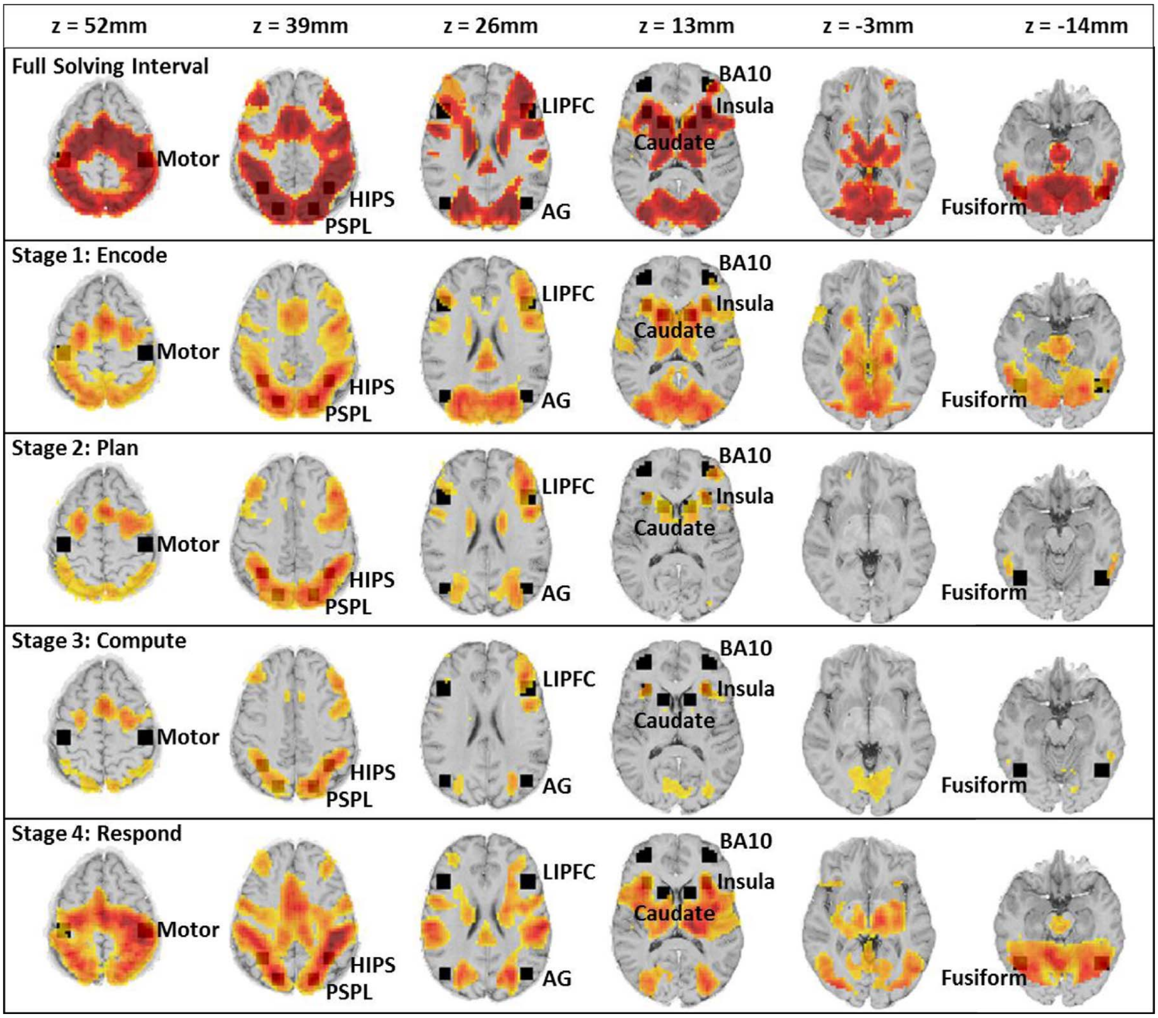


**Fig. 5.** Regions mutually active among Visuospatially- and Symbolically-trained solvers in the solving interval as a whole (top panel), and within each of the four solving stages. For reference, dark boxes (underlay) roughly delineate regions commonly associated with math tasks: AG (Angular Gyrus, Talairach center: ± 41,−65, 37; voxels:4x4×4), BA10 (Brodmann Area 10, ± 34,47,8; voxels:5x5×4); Caudate ( ± 13,10,7; voxels:4x4×4); Fusiform ( ± 42,−60,−8; voxels: 4x4×3); HIPS ( ± 34,−49,45; voxels:4x4×4; adapted from Cohen Kadosh et al. (2008)), Insula ( ± 31,20,5; voxels:2x3×4), LIPFC ( ± 43,23,24; voxels:5x5×4); Motor ( ± 49,−19,50; voxels:5x5×4); and PSPL ( ± 19,−68,55; voxels:4x4×4). For each brain slice the z-coordinate provided is for x=y=0 in Talairach coordinates (radiological convention: image left = participant's right).

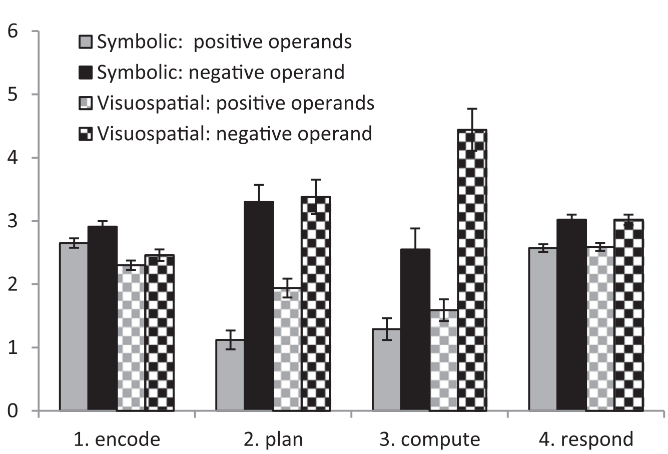


**Fig. 6.** Mean stage durations (s) for the 4 solving stages in the models of Visuospatially-trained versus Symbolically-trained solvers, for correct Standard Format Problems (b↑h=X or b↓h=X) with positive operands (e.g., 4↑3=X) or with a negative operand (e.g., −4↑3=X.; 4↑−3=X). Error bars are standard error.

to type (e.g., 4↓2 = 9 vs. −4↓2 = −9). The durations of the middle stages (planning and computation) exhibited more dramatic effects of negative operands (Fig. 6), as would be expected, since solvers have to adapt to and/or perform arithmetic with negative numbers.

There was also two-way interaction of training group and stage, $F(2.10,163.92) = 15.82$, $\eta_p^2 = .17$, $p < .001$. The mean duration of the response stage in the testing session was almost exactly matched across the visuospatially- and symbolically-trained groups, $t(78)=-0.15$, $p=.880$ (ns). This makes sense because the groups were exposed to equivalent sets of problems, and thus incurred comparable demands for typing answers. However, on average visuospatially-trained solvers had faster encoding times to interpret the problems, $t(78)=4.36$, $p < .001$, but longer planning stages, $t(78)=-2.42$, $p=.018$, and computation stages, $t(78) = -3.28$, $p =.002$, than solvers trained to form purely symbolic interpretations.

There was also a 3-way interaction of training type, stage and operand type (positive or negative), $F(1.75, 136.56) = 13.48$, $\eta_p^2 = .15$, $p < .001$. This mainly reflects the particularly long computation stage for negative operand problems in the visuospatial group.

*Activation patterns in the stages*

For these whole-brain analyses, each stage furnished a GLM regressor, as shown in the second time-line in Fig. 1. The HSMM-MVPA model provides information about the stage durations for each individual trial, so this information was used to generate stage regressors appropriate to each trial. Thus, stage regressors that are temporally tailored to each trial to manage differences in stage durations across trials and groups.

*Mutual visuospatial and symbolic activations within each solving stage*. Fig. 5 shows the large regions of mutual activation (conjunction analysis) for each stage, and gives a sense of the patterns of activation differentiating the stages. For example, during encoding and responding, solvers observe problems and/or their own responses as they type them, so these stages are characterized by more activity in visual processing regions like the occipital lobe. Similarly, the response stage has pronounced activity in the precentral motor region of the left hemisphere (Fig. 4, slice z = 52) associated with the requirement to type answers with the right hand. In the computation stage, activation seems the least wide-spread and is especially localized to areas like PSPL, HIPS, right frontal areas (Fig. 4, slice z = 39 mm), which are commonly associated with arithmetic (Arsalidou and Taylor, 2011), and also the cuneus (slice z = −3). Although perhaps not obvious in Fig. 5, the planning stage, which involves metacognitive processing to identify, retrieve and/or adapt an procedure to use to reach a solution was characterized by greater activation in frontal regions like BA10 and BA8 relative to their mean activation in the other stages (see also Anderson and Fincham, 2014b).

*Contrasts across visuospatial and symbolic groups within each solving stage*. Table 1 and Fig. 7 summarize the clusters exhibiting significant contrasts between visuospatially- and symbolically-trained solvers in each stage. As for duration differences, activation differences were detected in the first three stages. During encoding and computation, regions tended to be more active among visuospatially- (vs. symbolically) trained learners but the reverse was true for contrasts in stage 2 (planning). Regions more active in the visuospatial group during the encode stage included: bilateral supramarginal, precuneus, cuneus, lingual, and parahippocampas gyri, as well as left middle temporal regions and the right insula. As evident in the yellow regions in Fig. 7, there were notable consistencies (overlaps) among regions more active among visuospatially-trained solvers during encoding (stage 1) and again during computation (stage 3) in the supramarginal gyri, the precuneus, and the left parahippocampal gyrus. This suggests that distinct (presumably visuospatial) processing is not limited to the early interpretation of the problem but continues to play a role during computation.

The above activation contrasts were collapsed across negative and positive operand problems because group contrasts within only positive or negative problems were hampered by lower power.[2] To address the question of possible differences in group contrasts among positive and negative operand problems we did a supplementary analysis on our Table 1 regions with separate regressors for positive and negative operand problems in each stage to determine if any of these regions were characterized by an interaction of Operand Type (Positive vs. Negative) by Group (Visuospatial vs. Symbolic). Only one region, the left parahippocampal (stage 3, region 23 in Table 1) was characterized by a significant interaction, $F(1,78) = 3.96$, $\eta_p^2=.048$, $p = .050$, such that the group difference (visuospatial > symbolic) was greater among positive than negative operand problems, though significant within each subset, $t(55.3)$[3] $= 2.73$, $p=.008$ and $t(78) = 2.77$, $p=.007$, respectively.

## Discussion

The present research explored how training solvers to form purely symbolic versus meaningful visuospatial referents for math problems can impact their subsequent mental solving processes. For our problems, behavioral performance measures tended to be comparable across our visuospatially- and symbolically-trained groups. However, brain imaging data can be useful for providing information not behaviorally apparent about the use of different cognitive representations and strategies (as in Lee et al. (2010) and Sohn et al. (2004)).

*Problem solving stage-structure revealed by HSMM-MVPA*

One key objective was to assess whether the temporal stages of the mental solution processes were similar across solvers who had been trained to interpret problems visuospatially versus purely symbolically. We applied an HSMM-MVPA method for parsing fMRI data that afforded us insight on the solving sub-stages as learners progressed naturally through them. Separate models for the visuospatially- and symbolically-trained solvers indicated that both could be characterized by a basic 4-stage solution procedure: encode, plan, compute, and respond. However the stage models also revealed that there were temporal (Fig. 6) and neural differences (Fig. 7) across groups in all stages but the response stage. Note that the HSMM-MVPA process operated on the neural imaging data for the trials and was blind to any key press events, so it inferred the boundaries of each stage, including the response stage, based solely on a shift in brain activation patterns. Thus, the fact that the 'blind' HSMM-MVPA processes applied separately to each group produced this correspondence in the temporal and neural characterizations for their response stages provides some confidence in the validity of the analysis and models. Notably without the insight afforded by the stage-parsing of the HSMM-MVPA, no overall temporal or activation differences were found to be significant for the solving interval as a whole. Differences that are stage-specific may become diluted and undetectable in a traditional analysis.

Conceivably, visuospatially-trained solvers might have been able to bypass visuospatial processing and associate problems directly with arithmetic procedures, which, as demonstrated by the symbolic group, are sufficient to solve the problems. Such an interpretation might seem supported by the non-significant contrasts across groups in the solving interval as a whole. However, stage-wise differences were detected, so knowledge of visuospatial referents impacted mental solution processes.

One hypothesis mentioned in the introduction was that visuospatially-trained solvers might only differ from symbolically-trained solvers in the initial stage of interpreting the problem, but that subsequent stages would be comparable. Our results, however, support an alternative hypothesis that processing differences persist past encoding into the planning and computation stages.

*Overview of stage interpretations*

In this section, we present an interpretation of how solution processes may be partitioned across the stages in the two groups. This interpretation is informed by the training manipulation (visuospatial vs. symbolic), stage duration patterns, mutual activation patterns, and evidence from stage models in other problem solving studies

[2] The analysis of negative problems detected no group differences in stages 1 or 2 and only one region with a significant difference (visuopatial > symbolic) in stage 3 which was an 80 voxel subset of Table 1 region 20 (left inferior parietal). The group contrast among positive problems detected no differences in stages 2 or 3, and seven regions with differences (visuospatial > symbolic) in stage 1: Six were subsets of Table 1 regions (1, 2, 3, 4, 7 and 10), and the seventh in the medial frontal gyrus (BA10, 27 voxels). However, if a region (or voxel) in a given stage is detected only in the analysis of positive problems not negative problems (or vice versa) this could simply reflect an issue of power and might not license the conclusion that the effect is only relevant to one subset of problems.

[3] Equal variances not assumed since Levene's test failed.

**Table 1**
Regions from whole-brain analyses which exhibited a significant contrast of whether learners were trained with or without a visuospatial interpretation, by solving stage.

| Regions Identified in Whole-Brain Analyses[a,b] | Taliarach Peak x,y,z[c] | Brodmann Area(s) | Voxels | $t(78)$[d] max(mean) |
|---|---|---|---|---|
| **Stage 1 (Encoding): Visuospatial > Purely Symbolic VS-Symbolic** | | | | |
| 1. R. Post.Cingulate/Lingual/Parahipp./Cuneus | 14,−53,−1 | 18,19,30 | 152 | 5.09(3.81) |
| 2. R. Precuneus/Cingulate (extends to Left) | 4, −54,34 | 7,31 | 142 | 4.26(3.69) |
| 3. L. Cuneus (extends to R. Cuneus) | −2, −78,12 | 17,18 | 91 | 4.62(3.77) |
| 4. L. Precuneus | −17, −72,34 | 7 | 79 | 4.44(3.74) |
| 5. L. Middle/Superior Temporal Gyrus | −63,−20,−3 | 21/22 | 74 | 5.15(3.89) |
| 6. L. Culmen/Parahippocampal Gyrus | −26,−36,−12 | 36 | 59 | 4.93(3.87) |
| 7. L. Supramarginal Gyrus (into Inf. Parietal) | −48,−54,31 | 40 | 58 | 4.26(3.70) |
| 8. R. Supramarginal Gyrus (into Angular) | 57,−57,32 | 39/40 | 43 | 4.31(3.75) |
| 9. L. Middle Temporal | −57,−2,−14 | 21 | 35 | 4.52(3.75) |
| 10. L. Middle Temporal | −45,−69,14 | 39 | 29 | 4.61(3.82) |
| 11. R. Insula | 35,−26,14 | 13/41 | 28 | 4.18(3.67) |
| 12. L. Lingual/Fusiform Gyrus | −23,−63,−6 | 18/19 | 22 | 4.31(3.79) |
| 13. L. Lingual Gyrus | −14,−81,−6 | 18 | 20 | 4.37(3.78) |
| 14. L. Cuneus | −11,−88,29 | 19 | 20 | 4.16(3.70) |
| **Stage 2 (Planning): Purely Symbolic > Visuospatial** | | | | |
| 15. R. Medial Frontal Gyrus | 4,37,40 | 8/9 | 63 | −4.46(−3.70) |
| 16. R. Middle Temporal Gyrus | 63,−43,4 | 22 | 55 | −4.41(−3.73) |
| 17. L. Superior Frontal Gyrus | −26,56,−1 | 10 | 24 | −4.48(−3.71) |
| 18. R. Inferior Frontal Gyrus | 41,27,3 | 45/47 | 20 | −4.19(−3.75) |
| 19. R. Angular /Middle & Superior Temporal | 48,−59,22 | 39 | 20 | −4.20(−3.71) |
| **Stage 3 (Computation): Visuospatial > Purely Symbolic** | | | | |
| 20. L. Inferior Parietal Lobule (overlaps 7) | −48,−58,51 | 7,40 | 215 | 5.20(3.93) |
| 21. L. Precuneus/Cingulate Gyrus (overlaps 2) | −8,−48,37 | 7,31 | 63 | 4.78(3.82) |
| 22. R. Angular Gyrus ( & SupraMarg.,overlaps 8) | 57,−60,35 | 39/40 | 61 | 4.34(3.70) |
| 23. L. Parahippocampal Gyrus (overlaps 6) | −20,−27,−13 | 35 | 27 | 4.75(3.90) |
| 24. L. Middle Temporal Gyrus | −29,−55,25 | 39 | 20 | 4.77(3.86) |

[a] Clusters were identified using a voxel-wise p =.001 ($t$-threshold 3.413) and cluster size threshold > 20 voxels which yields a brain-wide alpha estimated to be less than 0.05 by simulation.
[b] Region labels and Brodmann numbers were obtained for the Taliarach-Tournoux Atlas using the "Where am I" function in the AFNI fMRI analysis software.
[c] Peak co-ordinates correspond to the maximum $t$-value in the cluster.
[d] $t$-values are the maximums and means of the voxel-wise t-values within each cluster from the whole-brain analysis.

(e.g., Anderson and Fincham, 2014a, 2014b; Anderson et al., 2014). In Sections 4.3 and 4.4 we will focus the regions that are differentially active between groups in particular stages.

***Stage 1: Encoding***. As reflected in its name, our first problem-solving stage includes encoding the symbols in the problem stimulus (e.g., "4↑3=X"). In the scanner, these problem stimuli were identical across groups. As in the current study (Fig. 5), the encoding stage is generally characterized by relatively high activity in visual areas (e.g., occipital and fusiform; Anderson and Fincham, 2014a; Anderson et al., 2014). Unsurprisingly, this stage was slightly longer in both groups for problems which included a negative operand – that is, an extra character to encode (−4↑3=X vs. 4↑3=X). However, since this stage could take over 2.5 s overall, it presumably also involved some interpretation beyond basic encoding – i.e., a preliminary characterization or classification of the problem.

More specifically, we suggest that some heightened regional activations among visuospatially-trained solvers in this first stage support forming a preliminary mental visuospatial representation for the problem (see also Section 4.3). For example, a visuospatially-trained solver might image a staircase (Fig. 1c) or a rectangle and a triangle (Fig. 1d). At this first stage, this process may entail the retrieval and instantiation of an abstract template with unspecified or default dimensions. As summarized

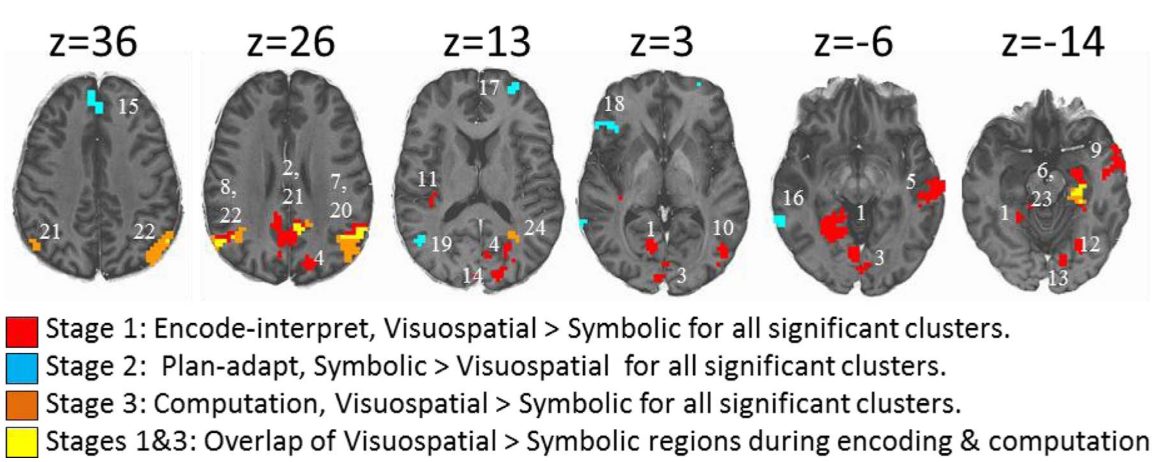


**Fig. 7.** Significant clusters from group contrasts in stages 1, 2 and 3 (warm colors: Visuospatial > Symbolic; cool colors: Symbolic > Visuospatial). No significant clusters were found in stage 4 (response). Clusters are labeled with their numbers in Table 1. For each brain slice (radiological convention: image left = participant's right), the z-coordinate provided is for x = y = 0 in Talairach coordinates.

in Appendix A, the stimulus set not only included Standard format problems (e.g., 4↑3=X), but also problems with the missing value, X, in other locations (4↑X=15 and X↑3=15) and problems containing two arrow operators (e.g., 4↓4=X↓5). For the representation scheme in Fig. 1c, all problems with one arrow operator) can be accommodated by an abstract single-staircase template, but for problems containing two arrow operators (4↓4=X↓5) an appropriate visuospatial template might involve two separate staircases that are equal in area. Thus, more generally, we can think of forming a preliminary visuospatial interpretation as an act of classifying the problem (e.g., 1-staircase vs. 2-staircase template).

The general act of making a preliminary classification of a stimulus problem also applies to symbolically-trained solvers. Symbolically-trained solvers, being unaware of visuospatial interpretations, must interpret problems solely on their symbolic stimulus structure. Procedurally they could compare the format of the current problem against stored templates representing other encountered problem formats. Thus, these different problems (4↑3=X and X↑3=15 and 4↑X=15) would match distinct format templates in symbolic classification (n↑m=X vs. X↑n=m vs. n↑X=m, where n and m are unspecified integer variables). This grain size of classification is logical for the symbolic group because these different formats then correspond to the requirement to select different algorithms in the subsequent planning stage. In absence of a higher-level 'meaning', the focus of the symbolic group would be on what to do (vs. what the representation looks like). This finer grain size of symbolic classification might account for the slightly longer stage 1 duration in the symbolic group. However, the symbolic classification process would not as strongly engage the visuo-spatial regions that showed heightened activation in the visually trained group.

***Stage 2: Planning***. We suggest that planning involves using the encoded operand values and problem classification from stage 1 to identify or adapt a method to reach a solution (Anderson and Fincham, 2014b). Planning processes generally involve working memory and frontal and prefrontal areas (Arsalidou and Taylor, 2011; Dehaene and Cohen, 1997). Frontal areas (e.g., LIPFC and BA10) were active in this stage (Fig. 5, also Anderson and Fincham, 2014a; 2014b), as was the caudate. A review by Grahn et al. (2008) associates the caudate with the excitation of correct action schemas and selection of appropriate sub-goals.

This planning stage sometimes took longer for visuospatial solvers. A possible reason for this effect is that the finer-grained format-based problem classification presumed for symbolic solvers in stage 1 may make planning very expedient by cuing the retrieval of an algorithm customized for the specific problem format (e.g., n↑m=X vs. X↑n=m vs. n↑X=m). Then symbolic solvers could just initialize the variables in this customized equation/algorithm with the operand values they encoded in stage 1 and proceed to the computation stage.

Visuospatially-trained solvers also have to initialize the values of the spatial dimensions for their visuospatial templates with the operand values they encoded in stage 1. Some spatial dimension values are transparent from the given operands. For example, for 4↑3=X, the first operand, 4, directly specifies the location and height of the first column (Fig. 1c) or the base of the rectangle (Fig. 1d). Such transparent dimensions might even be initialized in stage 1. However, the operand values to not directly specify other dimensions (e.g., the base of the triangle in Fig. 1d). Having initialized relevant dimensions for the representation, visuospatial solvers may proceed to determine how to compute the missing 'dimension' X – which, for Standard Problems, amounts to computing area. They can retrieve the equation/algorithm to compute area and initialize it with their dimension values. As such, planning may sometimes be slightly longer in the visuospatial group because it may be mediated by their visuospatial representation – that is, they may first initialize their visuospatial representation dimensions then determine and initialize the needed computation algorithm with these dimensions. In the symbolic group, the fine-grained symbolic classification in stage 1 may directly index the appropriate algorithm and they just have to initialize the values in this algorithm.[4] This possible mediation by a visuospatial representation may reduce some demands in executive function and working memory that may be required by the symbolic group to mentally represent the abstract purely symbolic equations/algorithms during planning. This may account for higher activations in some frontal regions in the symbolic group during planning.

The presence of negative operands increased planning duration substantially in both groups. Since both groups practiced problems with only positive operands prior to the scanner session, we suspect that even when solvers initially encounter a problem with a negative operand (−4↑3=X or 4↑−3=X), their default classification process may function as if all operands were positive. Then during planning, solvers in both groups tailor must their representation/approach to accommodate negative operands if present (for a more detailed procedural discussion, please see Anderson and Fincham (2014b)).

***Stage 3: Computation***. The computation stage involves performing the arithmetic required to calculate the answer. The duration of this stage is related to the number of arithmetic steps required (e.g., Anderson et al., 2016; Anderson and Fincham, 2014a). In the current context, arithmetic demands were balanced overall across visuospatial and symbolic solvers. However, computation duration was longer in the visuospatial group and there was evidence of distinct processing (visuospatial > symbolic activations) in several regions common to those in the interpretation stage 1. These differences may reflect the demands of maintaining or referring back to the mental visuospatial representation during computation – for example, to track their progress or re-derive the next required computation step.

As was the case in the planning stage, negative operands increased computation durations in both groups. Negative operands often necessitated doing arithmetic with negative numbers which is less practiced than arithmetic with positive numbers. The computation durations were especially extended on negative operand problems for visuospatially-trained solvers. One reason for this may be the counter-intuitive nature of computing areas for regions (columns, rectangles, triangles) with negative dimensions like height.

***Stage 4: Response.*** This final stage involves planning and executing a motor sequence to type the obtained answer. In this and other studies (e.g., Anderson and Fincham, 2014a; Anderson et al., 2014), this stage is characterized by relatively high activity in the motor cortex controlling the response hand (Fig. 5). Duration tends to increase with the complexity of the response (e.g., number of digits, Anderson and Fincham, 2014a), which is consistent with our longer response durations for negative operand problems, which often involve an extra negative sign character in the answer (e.g., 4↓2 = 9 vs. −4↓2 = −9). Since response demands were equalized across visuospatially- and symbolically-trained solvers, it was unsurprising (but methodologically validating) that our models revealed no group differences in this stage.

*Regions with higher activation in visuospatially-trained solvers*

During the encoding and computation stages we detected regions more active among visuospatially- versus symbolically-trained solvers.[5] During encoding these regions included: bilateral supramarginal, precuneus, cuneus, posterior cingulate, parahippocampus, and left lingual/fusiform gyri and left middle and superior temporal regions and right insula. Four of these regions again emerged in the computation stage: precuneus, right supramarginal/angular, left supramarginal/inferior parietal, and left parahippocampal gyrus.

In our stage interpretations (Section 4.2), we suggested that visuospatially-trained solvers may image a general visuospatial repre-

[4] Negative operands however may introduce additional planning demands for the symbolic group. This is detailed in Anderson and Fincham (2014b) for the case in Fig. 1a.

[5] We detected no regions significantly more active in the symbolic group in these stages.

sentation for the problem in stage 1, and may refer back to that representation during stage 3 computation to track their progress or deduce the next required computation step. Given that visuospatially-trained solvers were trained to interpret problems visuospatially, the suggestion that they mentally did so in the scanner seems plausible.

All of our clusters which were more active among visuospatially-trained solvers are in anatomical regions that have been previously implicated in mental imagery (Ganis et al., 2004). Some of these regions have been identified in prior studies contrasting visuospatial versus symbolic *stimulus formats* also emerged in our contrast of visuospatial versus symbolic *mental interpretations* of common problem stimuli (e.g., 3↑4=X): Specifically, the precuneus (Thomas et al., 2010); cuneus (Lee et al., 2010); insula (Venkatraman et al., 2005); and lingual/fusiform gyri (Lee et al., 2010; Piazza et al., 2007; Thomas et al., 2010). Also, some of these stimulus-contrast regions (precuneus, cuneus & lingual/fusiform) have also been implicated in math studies associated with the use of *mental* visuopatial strategies and/or imagery: precuneus (Chen et al., 2006; Dehaene et al., 2003; Pinel et al., 2001); cuneus (Lee et al., 2010; Zarnhofer et al., 2013); lingual and/or fusiform gyri (Lee et al., 2007; Zago et al., 2001; Zarnhofer et al., 2013). Such studies also implicated other regions similar to our encode and compute regions including: left supramarginal gyrus (Zago et al., 2001), left middle temporal lobe (Lee et al., 2007); and right angular gyrus (Du et al., 2013).

The Superior Temporal Gyrus (region 5 in Table 1) has (also) been associated with linguistic functions (e.g., phonological processing, Binder et al., 1994). Nonetheless, Ganis et al. (2004) found the superior temporal region was active during both visual perception and mental imagery. Also Zamboni et al. (2013) found that activation and structural density in this area were associated with success in a visuospatial association task (Placing Test). That said, an alternate phonological interpretation for the group contrast in this region might be that it reflects more use of subvocalization during problem interpretation in the visuospatial group, because this area has been associated with the perception of inner speech (e.g., Shergill et al., 2002). However, this phonological (vs. visuospatial) interpretation may be less likely because we did not simultaneously find increased activation in frontal areas associated with the *production* of inner speech (Shergill et al., 2002).

A subset of other regions in Stage 1 have been associated with the default mode network (Buckner et al., 2008) – for example, the posterior cingulate (regions 1 & 2 in Table 1), lateral temporal cortex (10 in Table 1), inferior parietal areas (7 in Table 1), and parahippocampus. As reviewed by Buckner et al. (2008), the default mode network includes regions that tend to be mutually correlated and more active when participants are at rest (i.e., in absence of external stimuli and/or task demands) - thus these regions presumably support stimulus-independent thought such as daydreaming and reliving episodic memories. As such, a possible interpretation of higher activity in these regions in the visuospatial (vs. symbolic) group is that visuospatially-trained solvers may have been entertaining more task-unrelated thoughts while interpreting the problems. On the other hand, a general function of default network regions may be to support mental imagery (Daselaar et al., 2010; Hassabis et al., 2007). Thus, heightened activation in these regions in the visuospatially-trained group may support mental imagery of the problems (vs. an increase in task-unrelated thoughts).

Many of our regions implicated in visuospatial processing or mental imagery are also implicated in the processing of stimulus ‘meaning’ more generally. Accordingly, in meta-analyses of semantic studies (Binder et al., 2009; Wang et al., 2010) many regions implicated in the semantic interpretation of language (e.g., words vs. non-words; concrete vs. abstract nouns) are strikingly similar to our stage 1 regions that were more active for participants who had been trained with visuospatial referents: supramarginal and angular gyri; middle temporal gyrus; posterior cingulate and precuneus; parahippocampus; and fusiform. Note that several of these ‘semantic’ regions also correspond to default network regions, which may reflect their common functional involvement of imagery – note semantic studies often contrast imageable vs. non-imageable nouns. In particular, such non-math studies suggest a semantic and imagery role for the parahippocampal gyrus (Binder et al., 2009; Ganis et al., 2005; Wang et al., 2010), which (like the precuneus & supramarginal gyrus) had enhanced activation among visuospatially-trained solvers during both encoding and computation, but was not identified in the prior math studies we discussed.

The supramarginal and angular gyrii may also fulfill a semantic role here. Work by Ansari (2008), Grabner et al. (2013) and Grabner et al. (2009) has implicated the left angular and left supramarginal gyri in the process of associating a math problem stimulus with a mental referent (i.e., the *symbol-referent mapping hypothesis*). In their context, the mental referents in question were specific numerical answers for problems that had been over-trained to permit solution by direct retrieval. It is possible that this mapping functionality of the left angular and/or supramarginal gyri might generalize to associating a problem stimulus to a mental visuospatial referent (e.g., associating 4↑3 with a mental image of a rectangle and triangle with appropriate relative dimensions as in Fig. 2d). Our finding of increased left supramarginal activation during stage 1 in the visuospatial group seems potentially compatible with this extended interpretation of their hypothesis.

*Regions with higher activation in symbolically-trained solvers*

Visuospatially-trained solvers had to arithmetically compute answers in a similar manner to their symbolically-trained counterparts. Due to this shared requirement, although we expected some extra visuospatial/semantic processing in the visuospatial group, we were unsure whether there would be regions more active among symbolic solvers at any stage of solving. However, in the current study, we found heightened activations in the symbolically-trained group during the planning stage.

Above (Section 4.2) we had suggested a possible interpretation for higher frontal activations among symbolic solvers during planning. Specifically, for visuospatial solvers, their planning process might be mediated by their visuospatial representation in a way that might incur a time cost but reduce executive and working memory demands incurred by symbolic solvers to mentally represent abstract purely symbolic equations/algorithms during planning.

More specifically, several of the regions more active in the symbolic group have been previously identified as relevant to parsing and comparing symbolic math expressions (Maruyama et al., 2012): middle temporal gyrus, superior frontal gyrus, inferior frontal gyrus, and angular gyrus.

Effects in the right angular gyrus are somewhat challenging to interpret in this and other stages because although it was sometimes more active among symbolically-trained solvers (i.e., during planning), it was sometimes more active among visuospatially-trained solvers (i.e., during encoding and computing). Other studies have also reported right angular gyrus activation during both symbolic and visuospatial conditions (Lee at al., 2007; Thomas et al., 2010). For example the right angular gyrus was active whether participants compared symbolic equation stimuli or graph stimuli (conjunction analysis, Thomas et al., 2010), and also when participants formed either symbolic or visuospatial mental representations for word problems (Lee et al., 2007). One possibility is that, since the angular gyurs is known to support multiple functions (Seghier, 2013), this region could be serving different functions for the two groups. It might support symbolic representation and processing during planning for the symbolic group (Maruyama, 2012), but visuospatial processing (e.g., Arsalidou and Taylor, 2011; Cattaneo et al., 2009; Göbel et al., 2001) for the visuospatial group. For example, transcranial magnetic stimulation of the angular gyrus was found to disrupt performance of a visuospatial search task and a number comparison task using the putative mental number line (Göbel et al., 2001). Alternately, the right angular gyrus maybe serving a general/common visuospatial function in both groups – for mental imagery of graphical representations in the visuospatially-trained group and for mental imagery and manipulation of symbolic representations in the symbolically-trained group.

## Conclusions

During complex problem solving, strategy differences are not always evident behaviorally or when analysing the fMRI data of the solution interval as a whole. An HSMM-MVPA process allowed us to parse the solution interval into stages (encode, plan, compute, respond) and find stage-specific differences in stage durations and activations across learners trained to associate math operations with visuospatial referents versus purely symbolic algorithms. This manipulation operationalized a more general contrast of interest between solution processes imbued with some semantic meaning versus rote calculation. Accordingly, during the early interpretation of our math problems, solvers able to generate visuospatial referents exhibited greater activation in regions also associated with imagery and linguistic semantics such as the supramarginal and angular gyri; fusiform and parahippocampal gyri; middle temporal gyrus; posterior cingulate and precuneus (Binder et al., 2009). Visuospatially-trained solvers also exhibited enhanced activation in several such regions during computation. Thus, we suggested that rather than using a proceduralized/automatic computation algorithm like symbolically-trained solvers, visuospatially-trained solvers may deduce the specific computations necessary for the current problem on-the-fly from their mental visuospatial representation. Since the two strategies can support similar behavioral performance, neither is necessarily better than then other in general. In the current context, our stage interpretations suggest the two strategies may partition workload differently across regions and stages.

In the current research, we manipulated whether or not learners were privy to visuospatial referents for the problem operations. However, in future work, learners' prior training/conceptions of an operation may be unknown or they might exercise a choice between visuospatial or symbolic approaches. Activations in these regions might become useful as a potential index of a solver's use of visuospatial interpretation strategies, because, as demonstrated in the current study, such strategy differences are not always apparent in behavioral measures.

## Appendix A

See Table A1.

**Table A1**
Types of problems presented during the testing session in the scanner on Day 2.

| Problem Type | Example Problems | Principle and/or Solution Process |
|---|---|---|
| **Standard Format** | | See Fig. 2 for solution examples |
| Positive Operands (16 items/subj) | 3↓2=X<br>3↑2=X | |
| Negative Operand | | |
| Negative b (8 items) | −3↑2=X<br>−2↓X=6 | Formulas/Examples apply as for Positive Operands |
| Negative h (8 items) | 3↑−2=X<br>2↓−5=X | Formulas apply as normal; Negative h reverses direction for addition |
| **Unknown Operand** | | |
| Unknown b (8 items) | X↑2=7<br>4↓X=7 | Learners may guess at X then check answer using solution procedure for standard format problems |
| Unknown h (8 items) | 4↑X=9<br>2↓X=3 | Learners may guess at X then check answer using solution procedure for standard format problems |
| **Relational Transfer** | | |
| Relating Up & Down Problems (8 items) | 31↑4=31↓X<br>19↓4=X↑4 | For positive integers B & H,<br>UpDown-1: B↓H = B↑(-H) and B↑H = B↓(-H)<br>UpDown-2: B↓H = (B-H+1)↑H |
| Consecutive Operand Problems | 35↓3=(34↓2)+X | For positive integers, B & H,<br>Consecutive-1: B↓H= (B-1)↓(H-1) + B |
| (8 items) | 26↓15=(X↓14) +26 | Task: Solve for the final constant value to add<br>Consecutive 2: B↓H= (B-1)↓(H-1) + B |

(continued on next page)

**Table A1** *(continued)*

| Problem Type | Example Problems | Principle and/or Solution Process |
|---|---|---|
| Mirror Problems (8 items) | 5↓5=X↓6<br>9↓X=9↓(X+1) | Task: Solve for an n or b value.<br>Mirror-1: B↓B= B↓(B+1),<br>The origin column (area=0) contributes nothing |
| | 50↓100 = X | Mirror-2: B↓2B = B and B↓(2B+1) = 0 |
| | 30↓61 =X | If area crosses the origin, negative columns cancel corresponding positive columns, for integer B > 0. |
| Rule Problems | 5↓X = 7↓X | 0-Rule: For any integer B, B↑0=0 and B↓0=0 |
| (8 items) | 5↑2 = 11↑X | 1-Rule: For any integer B, B↑1=B and B↓1=B |

## References


Anderson, J.R., Fincham, J.M., 2014a. Discovering the sequential structure of thought. Cogn. Sci. 38, 322–352.

Anderson, J.R., Fincham, J.M., 2014b. Extending problem-solving procedures through reflection. Cogn. Psychol. 74, 1–34.

Anderson, J.R., Lee, H.S., Fincham, J.M., 2014. Discovering the structure of mathematical problem solving. NeuroImage 97, 163–177.

Anderson, J.R., Pyke, A.A., Fincham, J.M., 2016. Hidden stages of cognition revealed in patterns of brain activation. Psychol. Sci. 27 (9), 1215–1226.

Ansari, D., 2008. Effects of development and enculturation on number representation in the brain. Nat. Rev. Neurosci. 9 (4), 278–291.

Arcavi, A., 2003. The role of visual representations in the learning of mathematics. Educ. Stud. Math. 52 (3), 215–241.

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.

Beckmann, S., 2004. Solving algebra and other story problems with simple diagrams: a method demonstrated in grade 4–6 texts used in Singapore. Math. Educ. 14 (1), 42–46.

Bengio, Y., Grandvalet, Y., 2004. No unbiased estimator of the variance of k-fold cross-validation. J. Mach. Learn. Res. 5 (Sep), 1089–1105.

Berends, I.E., van Lieshout, E.C., 2009. The effect of illustrations in arithmetic problem-solving: effects of increased cognitive load. Learn. Instr. 19 (4), 345–353.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 19 (12), 2767–2796.

Binder, J.R., Rao, S.M., Hammeke, T.A., Yetkin, F.Z., Jesmanowicz, A., Bandettini, P.A., Hyde, J.S., 1994. Functional magnetic resonance imaging of human auditory cortex. Ann. Neurol. 35 (6), 662–672.

Blatto-Vallee, G., Kelly, R.R., Gaustad, M.G., Porter, J., Fonzi, J., 2007. Visual–spatial representation in mathematical problem solving by deaf and hearing students. J. Deaf Stud. Deaf Educ. 12 (4), 432–448.

Booth, J.L., Koedinger, K.R., 2012. Are diagrams always helpful tools? Developmental and individual differences in the effect of presentation format on student problem solving. Br. J. Educ. Psychol. 82 (3), 492–511.

Booth, J.L., Siegler, R.S., 2008. Numerical magnitude representations influence arithmetic learning. Child Dev. 79 (4), 1016–1031.

Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network. Ann. N. Y. Acad. Sci. 1124 (1), 1–38.

Cattaneo, Z., Silvanto, J., Pascual-Leone, A., Battelli, L., 2009. The role of the angular gyrus in the modulation of visuospatial attention by the mental number line. NeuroImage 44 (2), 563–568.

Chen, F., Hu, Z., Zhao, X., Wang, R., Yang, Z., Wang, X., Tang, X., 2006. Neural correlates of serial abacus mental calculation in children: a functional MRI study. Neurosci. Lett. 403 (1), 46–51.

Clements, D.H., Battista, M.T., 1992. Geometry and spatial reasoning. In: Grouws, D.A. (Ed.), Handbook of Research on Mathematics Teaching and Learning: A Project of the National Council of Teachers of Mathematics. Macmillan, New York, NY, 420–464.

Cohen Kadosh, R., Lammertyn, J., Izard, V., 2008. Are numbers special? An overview of chronometric, neuroimaging, developmental, and comparative studies of magnitude representation. Progress Neurobiol. 84, 132–147.

Cox, R., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Cox, R., Hyde, J.S., 1997. Software tools for analysis and visualization of fMRI data. NMR Biomed. 10, 171–178.

Daselaar, S.M., Porat, Y., Huijbers, W., Pennartz, C.M., 2010. Modality-specific and modality-independent components of the human imagery system. Neuroimage 52 (2), 677–685.

Dehaene, S., Cohen, L., 1995. Towards an anatomical and functional model of number processing. Math. Cogn. 1, 83–120.

Dehaene, S., Cohen, L., 1997. Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. Cortex 33 (2), 219–250.

Dehaene, S., Piazza, M., Pinel, P., Cohen, L., 2003. Three parietal circuits for number processing. Cogn. Neuropsychol. 20 (3-6), 487–506.

D'Esposito, M., Detre, J.A., Aguirre, G.K., Stallcup, M., Alsop, D.C., Tippet, L.J., Farah, M.J., 1997. A functional MRI study of mental image generation. Neuropsychologia 35 (5), 725–730.

Du, F., Chen, F., Li, Y., Hu, Y., Tian, M., Zhang, H., 2013. Abacus training modulates the neural correlates of exact and approximate calculations in Chinese children: an fMRI study. Biomed. Res. Int.. http://dx.doi.org/10.1155/2013/694075.

French, J.W., Ekstrom, R.B., Price, I.A., 1963. Kit of Reference Tests for Cognitive Factors. Educational Testing Service, Princeton, NJ.

Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., Penny, W.D., 2011. Statistical Parametric Mapping: The Analysis of Functional Brain Images: The Analysis of Functional Brain Images. Academic Press, Cambridge, MA.

Ganis, G., Thompson, W.L., Kosslyn, S.M., 2004. Brain areas underlying visual mental imagery and visual perception: an fMRI study. Cogn. Brain Res. 20 (2), 226–241.

Gathercole, S.E., Pickering, S.J., 2000. Working memory deficits in children with low achievements in the national curriculum at 7 years of age. British J. Educ. Psychol. 70 (2), 177–194.

Geary, D.C., 2011. Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. Dev. Psychol. 47 (6), 1539.

Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429.

Göbel, S., Walsh, V., Rushworth, M.F., 2001. The mental number line and the human angular gyrus. NeuroImage 14 (6), 1278–1289.

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., 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. Hum. Brain Mapp. 34 (5), 1013–1024.

Grahn, J.A., Parkinson, J.A., Owen, A.M., 2008. The cognitive functions of the caudate nucleus. Progress Neurobiol. 86 (3), 141–155.

Hadamard, J., 1945. The Mathematician's Mind: The Psychology of Invention in the Mathematical Field. Princeton University Press, Princeton, NJ.

Hassabis, D., Kumaran, D., Maguire, E.A., 2007. Using imagination to understand the neural basis of episodic memory. J. Neurosci. 27 (52), 14365–14374.

Hegarty, M., Kozhevnikov, M., 1999. Types of visual–spatial representations and mathematical problem solving. J. Educ. Psychol. 91 (4), 684.

Hembree, R., 1992. Experiments and relational studies in problem solving: a meta-analysis. J. Res. Math. Educ. 23 (3), 242–273.

Hoong, L.Y., Fwe, Y.S., Yvonne, T.M.L., Subramaniam, T., Zaini, I.K.B.M., Chiew, Q.E., Karen, T.K.L., 2010. Concretising factorisation of quadratic expressions. Aust. Math. Teach. 66 (3), 19–24.

Ishai, A., Ungerleider, L.G., Haxby, J.V., 2000. Distributed neural systems for the generation of visual images. Neuron 28 (3), 979–990.

Kyttälä, M., Lehto, J.E., 2008. Some factors underlying mathematical performance: the role of visuospatial working memory and non-verbal intelligence. Eur. J. Psychol. Educ. 23 (1), 77–94.

Larkin, J.H., Simon, H.A., 1987. Why a diagram is (sometimes) worth ten thousand words. Cogn. Sci. 11 (1), 65–100.

Lee, K., Lim, Z.Y., Yeong, S.H., Ng, S.F., Venkatraman, V., Chee, M.W., 2007. Strategic differences in algebraic problem solving: neuroanatomical correlates. Brain Res. 1155, 163–171.

Lee, K., Yeong, S.H., Ng, S.F., Venkatraman, V., Graham, S., Chee, M.W., 2010. Computing solutions to algebraic problems using a symbolic versus a schematic strategy. ZDM 42 (6), 591–605.

Lewis, A.B., 1989. Training students to represent arithmetic word problems. J. Educ. Psychol. 81, 521–531.

Maruyama, M., Pallier, C., Jobert, A., Sigman, M., Dehaene, S., 2012. The cortical

representation of simple mathematical expressions. NeuroImage 61 (4), 1444–1460.
Mayer, R.E., 1985. Mathematical ability. In: Sternberg, R.J. (Ed.), Human Abilities: An Information Processing Approach. W.H. Freeman, New York: NY, 127–150.
Mix, K.S., Cheng, Y.L., 2011. The relation between space and math: developmental and educational implications. Adv. Child Dev. Behav. 42, 197–243.
Murata, A., 2008. Mathematics teaching and learning as a mediating process: the case of tape diagrams. Math. Think. Learn. 10 (4), 374–406.
O'Craven, K.M., Kanwisher, N., 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. J. Cogn. Neurosci. 12, 1013–1023.
Organisation for Economic Co-operation and Development (OECD), 2014. PISA 2012 Results: What Students Know and Can Do (Volume I, Revised Edition, February 2014): Student Performance in Mathematics, Reading and Science. Paris, France: OECD Publishing. Retrieved from ⟨http://www.oecd-ilibrary.org/education/pisa-2012-results-what-students-know-and-can-do-volume-i-revised-edition-february-2014_9789264208780-en⟩.
Piazza, M., Pinel, P., Le Bihan, D., Dehaene, S., 2007. A magnitude code common to numerosities and number symbols in human intraparietal cortex. Neuron 53 (2), 293–305.
Pinel, P., Dehaene, S., Riviere, D., LeBihan, D., 2001. Modulation of parietal activation by semantic distance in a number comparison task. NeuroImage 14 (5), 1013–1026.
Presmeg, N.C., 1997. Generalization using imagery in mathematics. In: English, L.D. (Ed.), Mathematical Reasoning: Analogies, Metaphors and Images. Routledge, New York, NY, 299–312.
Presmeg, N.C., 2006. Research on visualization in learning and teaching mathematics. In: Gutiérrez, A., Boero, P. (Eds.), Handbook of Research on the Psychology of Mathematics Education: Past, Present and Future. Sense, Rotterdam, The Netherlands, 205–235.
Price, G.R., Mazzocco, M.M., Ansari, D., 2013. Why mental arithmetic counts: brain activation during single digit arithmetic predicts high school math scores. J. Neurosci. 33 (1), 156–163.
Pyke, A., Betts, S., Fincham, J.M., Anderson, J.R., 2015. Visuospatial referents facilitate the learning and transfer of mathematical operations: extending the role of the angular gyrus. Cogn. Affect. Behav. Neurosci. 15 (1), 229–250.
Reuhkala, M., 2001. Mathematical skills in ninth-graders: relationship with visuo-spatial abilities and working memory. Educ. Psychol. 21 (4), 387–399.
Schmithorst, V.J., Brown, R.D., 2004. Empirical validation of the triple-code model of numerical processing for complex math operations using functional MRI and group Independent Component Analysis of the mental addition and subtraction of fractions. NeuroImage 22 (3), 1414–1420.
Seghier, M.L., 2013. The angular gyrus multiple functions and multiple subdivisions. Neuroscientist 19 (1), 43–61.
Siegler, R.S., Ramani, G.B., 2009. Playing linear number board games – but not circular ones – improves low-income preschoolers' numerical understanding. J. Educ. Psychol. 101 (3), 545–560.
Shergill, S.S., Brammer, M.J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R.M., McGuire, P.K., 2002. Modulation of activity in temporal cortex during generation of inner speech. Human. Brain Mapp. 16 (4), 219–227.
Slavit, D., 1998. The role of operation sense in transitions from arithmetic to algebraic thought. Educ. Stud. Math. 37 (3), 251–274.
Sohn, M.H., Goode, A., Koedinger, K.R., Stenger, V.A., Fissell, K., Carter, C.S., Anderson, J.R., 2004. Behavioral equivalence, but not neural equivalence—neural evidence of alternative strategies in mathematical thinking. Nat. Neurosci. 7 (11), 1193–1194.
Tall, D., 2006. A theory of mathematical growth through embodiment, symbolism and proof. Ann. Didact. Sci. Cogn. 11, 195–215.
Thomas, M.O., Wilson, A.J., Corballis, M.C., Lim, V.K., Yoon, C., 2010. Evidence from cognitive neuroscience for the role of graphical and algebraic representations in understanding function. ZDM Int. J. Math. Educ. 42 (6), 607–619.
van Garderen, D., 2006. Spatial visualization, visual imagery, and mathematical problem solving of students with varying abilities. J. Learn. Disabil. 39 (6), 496–506.
Venkatraman, V., Ansari, D., Chee, M.W., 2005. Neural correlates of symbolic and non-symbolic arithmetic. Neuropsychologia *43* (5), 744–753.
Wang, J., Condor, J.A., Blitzer, D.N., Shinkareva, S.V., 2010. Neural representations of abstract and concrete concepts: A meta-analysis of neuroimaging studies. Human Brain Mapping 31, 1459–1468.
Whyte, J.C., Bull, R., 2008. Number games, magnitude representation, and basic number skills in preschoolers. Dev. Psychol. 44 (2), 588–596.
Zago, L., Pesenti, M., Mellet, E., Crivello, F., Mazoyer, B., Tzourio-Mazoyer, N., 2001. Neural correlates of simple and complex mental calculation. NeuroImage 13 (2), 314–327.
Zago, L., Petit, L., Turbelin, M.R., Andersson, F., Vigneau, M., Tzourio-Mazoyer, N., 2008. How verbal and spatial manipulation networks contribute to calculation: an fMRI study. Neuropsychologia 46 (9), 2403–2414.
Zamboni, G., de Jager, C.A., Drazich, E., Douaud, G., Jenkinson, M., Smith, A.D., Tracey, I., Wilcock, G.K., 2013. Structural and functional bases of visuospatial associative memory in older adults. Neurobiol. Aging 34 (3), 961–972.
Zarnhofer, S., Braunstein, V., Ebner, F., Koschutnig, K., Neuper, C., Ninaus, M., Ischebeck, A., 2013. Individual differences in solving arithmetic word problems. Behav. Brain Funct. 9 (1), 28.