Psychological Research (2025) 89:34
https://doi.org/10.1007/s00426-024-02063-3

ORIGINAL

Implications of neural integration of math and spatial experiences for math ability and math anxiety

Raeanne N. Martell[1] · Richard J. Daker[1] · H. Moriah Sokolowski[2,3] · Daniel Ansari[4] · Ian M. Lyons[1]

Received: 21 February 2024 / Accepted: 13 November 2024

Abstract

Mathematical and spatial abilities are positively related at both the behavioral and neural levels. Much of the evidence illuminating this relationship comes from classic laboratory-based experimental methods focused on cognitive performance despite most individuals also experiencing math and space in other contexts, such as in conversations or lectures. To broaden our understanding of math-space integration in these more commonplace situations, we used an auditory memory-encoding task with stimuli whose content evoked a range of educational and everyday settings related to math or spatial thinking. We used a multivariate approach to directly assess the extent of neural similarity between activity patterns elicited by these math and spatial stimuli. Results from whole-brain searchlight analysis revealed a highly specific positive relation between math and spatial activity patterns in bilateral anterior hippocampi. Examining individual variation in math-space similarity, we found that greater math-space similarity in bilateral anterior hippocampi was associated with *poorer* math skills and *higher* anxiety about math. Integration of neural responses to mathematical and spatial content may not always portend positive outcomes. We suggest that episodic simulation of quotidian contexts may link everyday experiences with math and spatial thinking—and the strength of this link is predictive of math in a manner that diverges from math-space associations derived from more lab-based tasks. On a methodological level, this work points to the value of considering a wider range of experimental paradigms, and of the value of combining multivariate fMRI analysis with behavioral data to better contextualize interpretations of brain data.

Introduction

Researchers have long been interested in the relationship between mathematical and spatial thinking (Galton, 1880). At the behavioral level, this relationship is supported by studies indicating that higher levels of extant spatial abilities are correlated with higher math performance (Atit et al., 2022; Gunderson et al., 2012; Mix & Cheng, 2012; Young et al., 2018), and training of spatial abilities is associated with improvements in math abilities (Cheng & Mix, 2014). There is also behavioral evidence that numerical information biases spatial orienting (Fischer et al., 2003) and performance on line-bisection tasks (Calabria & Rossetti, 2005; Fischer, 2001). Similarly, spatial information has been shown to bias numerical parity judgements (Dehaene et al., 1993; Fischer & Shaki, 2014).

At the neural level, similar brain regions, particularly the intraparietal cortex, tend to be activated during math and spatial lab tasks (e.g., Hawes & Ansari, 2020; Hubbard et al., 2005; Newcombe et al., 2019). However, it is important to note there is a dearth of evidence showing neural overlap between responses to math and spatial tasks within the same study. In the few studies that have (see Kaufmann et al., 2008; Simon et al., 2004; Zago et al., 2008), each has examined neural activity during active cognitive engagement with math and space in the context of standard laboratory-based paradigms (e.g., solving arithmetic problems, completing spatial rotation tasks, etc.).

However, such laboratory tasks measure only a small subset of potentially relevant cognitive and affective processes that comprise the greater set of our experiences with math

✉ Ian M. Lyons
iml30@georgetown.edu

1 Department of Psychology, Georgetown University, Washington, DC 20057, USA

2 Department of Psychology, Toronto Metropolitan University, Toronto, ON M5B 1W7, Canada

3 Rotman Research Institute, Baycrest Hospital, North York, ON M6A 2E1, Canada

4 Department of Psychology and Faculty of Education, Western University, London, ON N6A 3K7, Canada

Published online: 10 December 2024

and spatial thinking. Against this backdrop, researchers are recognizing the importance of individual engagement with math and spatial thinking in educational and everyday contexts, especially for the formation and maintenance of attitudes about math and spatial thinking. However, it is unclear if the current evidence of a math-space link is specific to the experimental settings used previously, or if they generalize to stimuli that evoke educational and everyday contexts as well.

In terms of analytic framework, examples in the current literature on math-space relations that combine both behavioral and neural evidence remain relatively rare. This leaves an incomplete picture of what neural overlap between math and space actually implies for a given individual. Additionally, prior work pointing to common neural substrates during math and spatial lab tasks has relied on identifying overlap between univariate activation maps (Hawes et al., 2019; Kaufmann et al., 2008; Simon et al., 2004; Zago et al., 2008). This approach is limited as, at best, it allows only for weak reverse inference, which is of limited theoretical value in terms of inferring shared neural mechanisms (Chatham & Badre, 2019; Coltheart, 2013; Poldrack, 2006; Woolrich, 2012). Further, given overlap is generally computed at the group level, it is not straightforward how one might provide better context to allow for stronger inference, for instance by associating overlap with behavioral outcomes at the individual level.

An alternative approach is representational similarity analysis (RSA), which directly tests for similarity in distributed activation patterns (Arbuckle et al., 2019; Dimsdale-Zucker & Ranganath, 2018; Kriegeskorte & Diedrichsen, 2019), thereby allowing for stronger inferences about shared neural mechanisms (Calzavarini & Cevolani, 2022; Hutzler, 2014; Poldrack & Farah, 2015). Beyond respecting the fact that neural processing is likely distributed in nature (Kriegeskorte & Diedrichsen, 2019), this approach allows for stringent controls also at the level of distributed activity patterns and on a per subject basis. Crucially, because similarity estimates are computed at the individual level, one can straightforwardly relate neural similarity estimates with behavioral outcomes. This in turn allows one to better contextualize and therefore draw more principled reverse inferences with regard to shared neural mechanisms and their behavioral implications. For this reason, here we focus our theoretical hypothesis testing and interpretation on RSA-based assessment of shared neural representation of everyday math and spatial situations.

Current study

As noted above, the neural bases of math-space integration have been previously explored via several active, cognitive laboratory tasks. Important as such work is, such tasks are representative of only a fragment of the contexts in which math-space integration might be seen, and thus capture only a small subset of processes potentially relevant to the sum of one's overall experiences with math and spatial thinking. To more fully understand the neural bases of math-space integration, it is therefore important to expand the range of contexts in which this link is investigated.

With this in mind, our first aim was to understand the neural substrates supporting math-space integration using stimuli depicting educational and everyday mathematical and spatial contexts. We operationalized this by testing for neural similarity when participants listened to auditory stimuli depicting domain-specific educational and everyday mathematical and spatial contexts (presented to participants as a memory task). The second aim of this study was to identify how math-space integration relates to cognitive and affective behavioral math outcomes.

The logistics of fMRI design limit the extent to which brain activity can be measured during actual everyday events. Hence, we approximated these situations by measuring neural activity during the encoding phase of a domain-specific verbal memory task. Participants listened to sentences that described a range of educational and everyday situations. In each situation, the topic of the sentence was either mathematical or spatial. An example math sentence was, "You study for a final mathematics test." An example spatial sentence was, "You memorize maps for a geography test." Recognition memory for the sentences was subsequently tested after participants exited the scanner. Stimuli were presented in different modalities in the encoding and recognition phases to further encourage semantic comprehension of the sentences.

Given the novel nature of our stimuli, as a preliminary step, we conducted univariate analyses to check whether current results are reasonably congruent with prior literature. To test our primary theoretical aim, however, we used whole-brain RSA to identify brain areas showing significant unique similarity for sentences depicting math and spatial content. To address our second aim, we then tested whether math-space similarity at the individual level predicted cognitive and affective math outcomes. In particular, we tested whether greater math-space similarity predicted more desirable math outcomes (higher math performance, lower math anxiety) or less desirable math outcomes (lower math performance, higher math anxiety).

Methods

Participants

Participants included 59 first-year university students at Western University (Canada). Of these, 2 participants were excluded due to technical problems with the auditory equipment, 3 participants were excluded due to excessive motion, and 1 participant was excluded because they did not follow instructions in responding to tone stimuli (see procedure details below). The final sample included in all analyses was thus $N = 53$ (36 female, mean age = 18.51 years). All stimuli and procedures were approved by the Western University Ethics Review Board, and all participants provided written consent.

Procedure

The results reported in the current study are part of a larger dataset. However, all hypotheses tested and results reported herein are unique to the current paper. Participants completed two sessions during the fall semester and the winter semester, respectively. In the first session, participants completed the study's in-lab behavioral component comprised of a battery of psychological surveys and cognitive tasks. Relevant behavioral measures described below are from this session (with the exception of the post-scan memory task). The order of the behavioral measures was randomized. In the second session, participants completed a set of functional tasks while in a magnetic resonance imaging scanner (i.e., fMRI). More specifically, participants completed several functional runs in a random order and a high-resolution anatomical scan. Here, we are most interested in the 3 functional runs containing only the auditory sentence stimuli. Participants completed the post-scan memory task immediately after exiting the scanner.

Instruments

Behavioral and fMRI stimuli and tasks were presented via E-Prime 2.0; surveys were presented via Qualtrics.

Math measures (behavioral)

Math Ability Participants completed a series of challenging mental arithmetic problems (mean RT = 9.9 s, mean accuracy = 81%). For each problem, participants were required to supply their own answers. Problems were a mix of addition (e.g., 67 + 95 + 52), subtraction (e.g., 283–97), multiplication (e.g., 36 × 7), and division (e.g., 522 ÷ 9). Each operation type was presented separately in a 3 min block, and scores were the total number of problems correctly answered across all 4 blocks within the specified time-limit (3 min per operation) (Sokolowski et al., 2019). A higher score thus indicates higher arithmetic ability ($M = 47.72$, $s = 21.47$, observed range: 11–121). Cronbach's α for this measure was 0.89.

Math anxiety Math anxiety was measured using the short math anxiety rating scale (SMARS) (Alexander & Martray, 1989). This scale asks participants to indicate how anxious they would feel in 25 different scenarios, each of which involved math (e.g. "Being given a set of addition problems to solve on paper"). Participants responded using a 5-point Likert scale (0 = not at all; 4 = very much). Possible scores ranged from 0 to 100. A higher score indicates higher math anxiety ($M = 28.28$, $s = 19.95$, observed range: 4–77). Cronbach's α for this measure was 0.96.

Spatial measures (behavioral)

Spatial ability Spatial ability was measured using accuracy on a standard mental rotation task (MRT) (Shepard & Metzler, 1971). In this task, participants were shown pairs of objects—half of which were congruent (the same object twice), and the other half of which were incongruent (two different objects). Participants were asked to mentally rotate these objects and use a button box to indicate whether the pair of objects were congruent or incongruent. Here, a higher score indicates higher spatial ability ($M = 80.8\%$, $s = 12.0\%$, observed range: 55.1–98.0%). Reliability for MRT measures with stimuli akin to those used here is generally acceptable: $\alpha = 0.87$ (Caissie et al., 2009).

Spatial anxiety Spatial anxiety was measured using the sum of the three subscales from Lyons et al. (2018). This scale asks participants to indicate how anxious they would feel in 24 different scenarios, each of which involved one of three spatial elements: mental manipulation, navigation or imagery (e.g., "Asked to imagine and mentally rotate a 3-dimensional figure"). Participants responded using a 5-point Likert scale (0 = not at all, 4 = very much). A higher score indicates higher overall spatial anxiety ($M = 33.72$, $s = 15.16$, observed range: 1–66). Cronbach's α for this measure was 0.86.

Control measures (behavioral)

Working memory capacity Working memory capacity was measured using the automated reading span task (R-Span) (Conway et al., 2005; Redick et al., 2012). In this task, participants were asked to read sentences and determine their semantic validity (e.g., "Andy was stopped by the policeman because he crossed yellow heaven," is syntactically valid, but not semantically). Participants were given detailed instructions and multiple examples to ensure they understood this aspect of the task (see Conway et al., and Redick et al., for complete task details). After each sentence, participants were also cued to remember a letter. At the end of each block of sentences (3–7 sentences), participants were prompted to recall the cued letters in the same order they were presented. Working memory scores are based on the number of correctly recalled letters. Hence, a higher score indicates a higher working memory capacity ($M = 46.77$, $s = 21.47$, observed range: 12–75). Reliability for this task is generally acceptable: $\alpha = 0.75$ (Đokić et al., 2018).

General anxiety General anxiety was measured using the trait subscale of the State-Trait Anxiety Inventory (STAI) (Spielberger et al., 1970), where a higher score indicates higher general anxiety ($M = 39.25$, $s = 10.76$, observed range: 22–72). Note we used just the trait subscale because the math and spatial anxiety measures are phrased at the trait level, thus making the trait subscale of the STAI the more appropriate control measure. Cronbach's α for this measure was 0.93.

Auditory memory task (fMRI)

To approximate educational and everyday contexts while in the fMRI scanning environment, we had participants listen to sentences describing educational and everyday situations. Each sentence described a situation primarily involving one of three different types of content: math, spatial reasoning, or reading. Sentences were initially adapted from surveys and questionnaires that assess attitudes about different topics across a range of settings. Sentences were subsequently modified to be more squarely situated within the relevant content domain (math, space, reading). Finally, sentences were refined and selected to equate the different cognitive domains in terms of basic auditory and linguistic properties (temporal duration, number of words, number of syllables, etc.). The average duration of the sentences was 3264 ms (range 1919–5365 ms), and the three domains did not differ from one another in terms of average duration (all $ps > 0.05$). See Appendix 1 for a complete list of all sentence stimuli presented in the scanner.

All sentences were presented aurally via MR-compatible headphones. Participants listened passively to the sentences. Sporadic tones occurred during the functional runs, which served as non-semantic auditory control stimuli. Furthermore, to encourage participants to maintain attention to all auditory inputs, participants were instructed to press a button whenever they heard a tone. Three participants were removed for failing to do so. In each functional run, participants heard 30 sentences (10 from each cognitive domain: math, space, reading). Across the 3 functional runs, participants thus heard 30 sentences from each cognitive domain. Sentence order was randomized within each run. Each functional run began and ended with a 15 s fixation period to estimate baseline, and each run contained three tone/button-press events to ensure participant attention. Events (sentences and tones) were presented in a rapid-event-related fashion, using a power-law inter-trial-interval (ITI) ranging from 2731 to 8487 ms (mean = 4711 ms).

We also took several steps to increase the likelihood that participants accessed the semantic content of the sentences. First, prior work has demonstrated that attending to the semantic content of linguistic inputs increases memory for those inputs (Craik & Lockhart, 1972; Ferré, 2003). To increase the likelihood that participants were attending to the semantic content of the sentences in the current study, participants were informed that their memory for these sentences would be tested after exiting the scanner. Memory for the sentences was indeed tested, and participants performed well above chance (see post-scan memory task below for further details). Finally, we conducted a task validation analysis to check whether brain areas traditionally associated with semantic processing were activated when listening to the sentences. Second, stimuli were presented in different modalities in the encoding (auditory) and recognition (visual) phases to further encourage semantic comprehension of sentence content (e.g., Karlsson et al., 2013).

Post-Scan recognition memory task

Shortly after exiting the scanner, participants completed a recognition memory task. Memory task data from one of the 53 participants was lost due to a technical error. For mean-based analyses, the analytic N for this task was 52. For correlation and regression analyses, the missing participant's data were imputed via linear interpolation (results did not vary substantially if the participant was omitted altogether). Participants saw 90 sentences in written form (presented one at a time). Half of these stimuli were repeated from the scanning environment; half were new sentences. Repeated and new sentences were equally divided among the three sentence types (math, space, reading). The sentence order was randomized. Participants' task was to indicate if a given sentence was one they recognized from their time in the scanner ('old') or not ('new'). Overall accuracy on this task was well above chance: $M = 79.3\%$, $s = 8.7\%$, $p = 1.1\text{E}{-29}$ (where chance = 50%). Memory accuracy did not differ significantly ($ps > 0.05$) across the three sentence-types (math: $M = 79.1\%$, $s = 10.3\%$; space: $M = 80.7\%$, $s = 11.3\%$; reading: $M = 78.0\%$, $s = 10.5\%$). In sum, participants appeared to be accessing the semantic content of the sentences, and their memory for said content did not differ between the sentence types. Cronbach's α for this measure was 0.74.

fMRI data acquisition and preprocessing

Data acquisition parameters

MRI data were collected via a 3 T Siemens Magnetom Prisma scanner, using a 32-channel head coil. A high-resolution whole-brain anatomical scan (T1-weighted) was acquired with the following parameters: repetition time (TR) = 2300 ms, flip-angle = 9°, 192 sagittal slices (1 mm thickness, no skip), in-plane resolution of 1 × 1 mm (240 × 256 matrix), total field of view (FOV) of 192 × 240 × 256 mm, final anatomical resolution = 1 mm^3 isometric voxels. Functional scans were acquired using an echo-planar imaging (EPI) sequence (T2*-weighted) with the following parameters: TR = 2500 ms, echo time (TE) = 30 ms, flip-angle = 78°, 44 axial slices (3 mm thickness, no skip), ascending-interleaved acquisition, in-plane resolution = 3 × 3 mm (70 × 70 matrix), total FOV = 210 × 210 × 132 mm, final functional resolution = 3mm^3 isometric voxels. Each functional run comprised an average of 119 TRs for an average duration of 4.96 min per functional run (note that runs varied slightly in duration because auditory stimuli varied in duration and their presentation was randomized across runs).

Image preprocessing

fMRI data were preprocessed using BrainVoyager 20.6 (Brain Innovation, Maastricht, Netherlands). Functional time-series were slice-time corrected using cubic-spline interpolation, corrected for head motion (6 parameters, trilinear/sinc interpolation), and subjected to a high-pass temporal filter (GLM-Fourier basis set) that included linear detrending. Functional images were then aligned to anatomical images using 12-parameter affine fine-tuning alignment, and both were then transformed into standardized Talairach space. Functional data were not spatially smoothed. Participants who exhibited excessive head-motion [total motion > 1 functional voxel (3 mm) and/or sudden movements > half a functional voxel (1.5 mm)] were removed from the dataset (N = 3).

Univariate analyses

Univariate model

A random effects (RFX) univariate general linear model (GLM) was computed using BrainVoyager 20.6. For each participant in each voxel, the model estimated betas for 4 event-types (math, space, reading and tone events) plus baseline. In each functional run, there were 33 events: 10 each for math, space and reading, and 3 tone events. Across the three runs, there were thus a total of 30 each of math, space and reading events, and 9 tone events. Math, space and reading events were modeled for the duration of the stimulus; tone events were modeled from the onset of the tone to the participant's response. Events were then convolved assuming a standard 2-gamma hemodynamic response function (HRF) model. A fifth beta, baseline, was estimated as the average deflection from the global mean for each participant in each voxel (i.e., each participant's random intercept in that voxel). Betas were computed as %-signal change. The resulting GLM dataset (all betas for all subjects in all voxels) was then exported to Matlab for similarity-based analyses.

Auditory task verification

Despite the inclusion of an attention check (tones) and a memory encoding component, participants passively listened to the sentences. Hence, we checked that participants indeed semantically comprehended the sentences by testing for greater activation for sentences relative to tones in brain areas traditionally associated with sentence comprehension and semantic processing of language, such as left anterior inferior frontal gyrus (IFGa) and middle-to-anterior portions of left superior temporal gyrus (STGm) (Chang et al., 2015; Enge et al., 2021; Price, 2010; Rodd et al., 2015). We did so via the conjunction of the following RFX contrasts: (Math Sentences > Tones) ∩ (Spatial Sentences > Tones) ∩ (Reading Sentences > Tones). The overall conjunction model was thresholded at $p < 0.005$, and subsequently cluster-level corrected for multiple comparisons using a Monte-Carlo simulation procedure (Forman et al., 1995) at $\alpha < 0.01$. This threshold was yoked to those used for the RSA searchlight analysis.

Univariate overlap between math and space

To situate the everyday sentences within prior literature that has used primarily lab-based tasks, we conducted a preliminary test for overlapping brain activity for math and space sentences using a traditional univariate approach. Here we also controlled for domain-specificity of sentence content. We did so via the conjunction of the following RFX contrasts: (Math Sentences > Reading Sentences) ∩ (Space Sentences > Reading Sentences). The overall conjunction model was thresholded at $p < 0.005$, and cluster-level corrected at $\alpha < 0.01$. This threshold was yoked to those used for the RSA searchlight analysis.

Representational similarity analysis (RSA)

To address our primary aim of testing for unique math-space similarity at the neural level, we conducted whole-brain (searchlight) RSA. RSA inputs were unsmoothed functional voxels (3mm^3). For the searchlight procedure, spherical ROIs extended a maximum of 3 functional voxels in each direction from the center-voxel (10.5 mm radius) for a maximum ROI size of 123 functional voxels.

Data triage and ROI validation

Valid ROIs for RSA were identified in the following manner. First, at the individual participant-level, each voxel was checked to ensure it contained valid betas for all 5 parameters; if it did not, it was omitted for that participant. Next, if data were omitted from a given voxel for more than 10% of participants (6 or more), then it was omitted from consideration for all participants. Again at the participant-level, a given ROI was considered valid only if, after accounting for voxels omitted via the aforementioned restrictions, it contained at least 50% of the maximum number of voxels (62 or greater). For RSA, one computes the correlation between a pair of betas across a given set of voxels at the level of the individual subject. The degrees of freedom (*df*) for a given RSA value (*r* value) are thus dependent upon the number of voxels in the ROI. The above triage measures thus ensured that (1) only voxels with valid data for the large majority of participants were considered, and (2) that the *df* for a given ROIs was not less than 50% of those of any other ROI. As the precision of a correlation depends on its *df*, this, in turn, ensured the precision of RSA estimates did not substantially vary across different parts of the brain (especially boundary conditions).

Similarity computation

The output of the whole-brain searchlight was a map with each voxel containing a matrix of similarity values (*r* values). A given *r* value from the matrix of a given voxel represented the correlation between the specified pair of betas across the voxels within the ROI sphere with the voxel in question as its center-point. In our case, we were primarily interested in the *unique* similarity between activity patterns elicited by listening to sentences about everyday math and everyday spatial situations. We thus computed *partial* similarity matrices, in which the correlation between any given pair of variables controlled for the influence of all other variables in the matrix. The input variables were the betas for math, space, and reading sentences, as well as the tone events. The partial-*r* value relating math and space betas thus accounted for the activity patterns elicited by reading and tone events. In this way, we controlled for both general auditory monitoring (tone events) as well as semantic processing and memory encoding of content unrelated to math or space. These similarity estimates—computed at the individual participant level—thus capture shared patterns of neural activation that are unique to sentences evoking math and space. However, we saw it as insufficient simply to demonstrate the existence of unique math-space similarity relations in the brain. For instance, a given brain region might show significant unique math-space similarity, but much stronger unique math-reading similarity. From the perspective of understanding math-space relations in the brain, it is difficult to see how one might interpret such a result.

To overcome this limitation, we turned to a more stringent test requiring one to demonstrate that the unique math-space similarity in (the sphere of voxels surrounding) a given voxel is also greater than the math-reading and space-reading similarity in (the sphere of voxels surrounding) that voxel. This approach required us to directly contrast partial-*r* values for each participant in each voxel. To do so (for a given participant and voxel), we extracted the math-space, math-reading and space-reading partial-*r* values from the similarity matrix. We then computed the equivalent of a planned contrast, comparing the math-space partial-*r* against the average of the math-reading and space-reading partial-*r*s. First, to normalize the distribution of r-values, partial-rs were z-transformed using the standard Fisher-z transformation: $z = \text{atanh}(r)$. Next the comparison was computed using the Steiger's z-test (Steiger, 1980) for comparing correlations within the same correlation matrix (i.e., within-sample correlations): $z_\Delta = \left(z_{ab} - z_{ac}\right)\frac{\sqrt{N-3}}{\sqrt{2 \bullet h\left(1-r_{bc}\right)}}$, where z_{ab} is the average of the z-transformed math-space partial-*r*, z_{ac} is the average of the z-transformed math-reading and space-reading partial-*r*s, and N is the number of voxels in that ROI for that participant. Further, $h = \frac{1-(f \bullet m^2)}{1-m^2}$, where $f = \frac{1-V_{bc}}{2\left(1-m^2\right)}$, and $m = \frac{r_{ab}^2 + r_{ac}^2}{2}$. Note that V_{bc} is the standardized variance common to the two correlations of interest for the contrast (r_{ab} and r_{ac}), but because we are dealing with partial-*r*s, we know this shared variance is reduced to 0; hence, for our purposes, we can set all instances of V_{bc} to 0.

The result of the above procedure was that, for each participant in each voxel, we obtained a z_Δ that quantifies the difference between (unique) math-space similarity and the (average unique) similarity between these variables and reading. Our final step was to test whether this difference was consistent across subjects, which entailed a one-sample t-test of z_Δ values against 0 (i.e., one z_Δ for each of the 53 participants). This resulted in a whole-brain map of t-scores ($df = 52$), which was thresholded at $p < 0.005$ and subsequently cluster-level corrected for multiple comparisons using a Monte-Carlo simulation procedure (Forman et al., 1995) at $\alpha < 0.01$ (minimum ROI size: 25 functional voxels). Because a searchlight approach implicitly smooths similarity maps due to shared voxels between ROIs, we used an estimated smoothing kernel of 2 functional voxels (or 6mm) to account for this implicit smoothing. Note also that we adopted a more liberal voxelwise threshold ($p < 0.005$) and more conservative cluster-level threshold ($\alpha < 0.01$) because, as detailed above, our procedure for identifying specific math-space similarity at the voxel level was already relatively conservative. Hence, we adopted a voxelwise threshold that would be less likely to yield Type II errors given our overall conservative approach (Lieberman & Cunningham, 2009), but balanced this with a more conservative cluster-level threshold to also avoid Type I errors.

Data availability

Data for this project can be found on the Open Science Framework (OSF) at https://osf.io/zubq8/.

Results

Univariate results (preliminary analyses)

Preliminary analysis 1: task verification

The theoretical utility of the novel stimuli used here rely on the assumption that participants comprehended the semantic content of the everyday sentences as they were presented. Performance on the post-scan memory task suggested they did, but here we provide an additional test at the neural level. To that end, we found greater activation for all three sentence types relative to tones [(Math Sentences > Tones) ∩ (Spatial Sentences > Tones) ∩ (Reading Sentences > Tones)] in the middle portion of bilateral superior temporal gyrus (STGm), left IFGa, left IPS, orbital frontal cortex (OFC), dorso-medial primary motor cortex (M1dm) and left dorsal M1 (LM1d). Region details are given in Table 1; regions are visualized in Fig. 1. Left STGm and left IFGa in particular overlap with areas previously identified as part of the semantic network (Chang et al., 2015; Enge et al., 2021; Price, 2010; Rodd et al., 2015), especially in the context of sentence comprehension. Coupled with strong post-scan memory recognition for the sentences, this result lends weight to the view that participants were attending to and accessing the semantic content of the sentences. Note also that individual task contrasts (e.g., Math > Tones) did reveal not any regions qualitatively distinct from those listed in Table 1.

Table 1 ROI anatomical details

	Mean *x*	Mean *y*	Mean *z*	Volume (mm^3)
LSTGm	– 53.4	– 16.5	6.1	6336
RSTGm	55.0	– 11.3	5.1	2221
LIPS	– 39.4	– 63.6	33.9	1114
LIFGa	– 39.0	28.2	– 2.7	725
OFC	– 2.3	37.0	– 7.6	2049
LM1d	– 32.3	– 19.3	50.4	728
M1dm	– 0.6	– 22.1	59.8	3268

L/RSTGm left/right middle portion of the superior temporal gyrus, *LIPS* left intraparietal sulcus, *LIFGa* left anterior inferior frontal gyrus, *OFC* orbito-frontal cortex, *LM1d* left dorsal primary motor cortex, *M1dm* dorso-medial primary motor cortex.

Preliminary analysis 2: univariate overlap

As noted in the introduction, univariate overlap provides at best weak evidence for shared neural mechanisms (Chatham & Badre, 2019; Coltheart, 2013; Poldrack, 2006; Woolrich, 2012). However, because in this study we used a relatively novel stimulus set, in this section we present results from univariate analysis showing co-activation for math and spatial sentences. To be clear, the primary utility of this analysis is simply to situate the current stimuli within prior literature using lab-based math and spatial tasks, which has relied primarily on univariate overlap. Figure 2 shows the following contrast: (Math Sentences > Reading Sentences) ∩ (Space Sentences > Reading Sentences). This analysis revealed one significant region in the left inferior parietal lobe (LIPL, Fig. 2), that spanned the posterior portion of the left supramarginal gyrus (LSMGp) and the anterior portion of left angular gyrus (LAGd). This result is consistent with prior univariate results showing the parietal cortex to be a point of univariate overlap between lab-based math and spatial tasks requiring active responses (Hawes et al., 2019). This suggests our results with everyday sentence stimuli do not radically diverge from prior work using lab-based tasks, which is encouraging. However, both this analysis and the prior work with which it converges rely on univariate overlap, and as noted above, this approach is unsatisfactory when it comes to drawing strong inferences about mechanisms shared by math and spatial processing—be that of lab-based tasks or sentences depicting quotidian situations. In the following sections, to more properly address the core aims of the study, we turn to RSA to provide a means of drawing stronger inferences about shared math and spatial neural mechanisms (Calzavarini & Cevolani, 2022; Hutzler, 2014; Poldrack & Farah, 2015).

Primary analysis 1: whole-brain RSA results

The primary aim of the paper was to identify evidence of shared neural mechanisms in quotidian contexts involving math and spatial reasoning. An effective means of drawing strong inferences in this regard is RSA (Calzavarini & Cevolani, 2022; Hutzler, 2014; Poldrack & Farah, 2015), so here we used a whole-brain RSA approach to identify brain regions showing unique math-space neural similarity during auditory processing of stimuli describing everyday math and spatial situations. We identified unique math-space neural similarity by controlling for patterns associated with auditory stimuli about reading and general auditory attention. Results identified five significant clusters showing greater

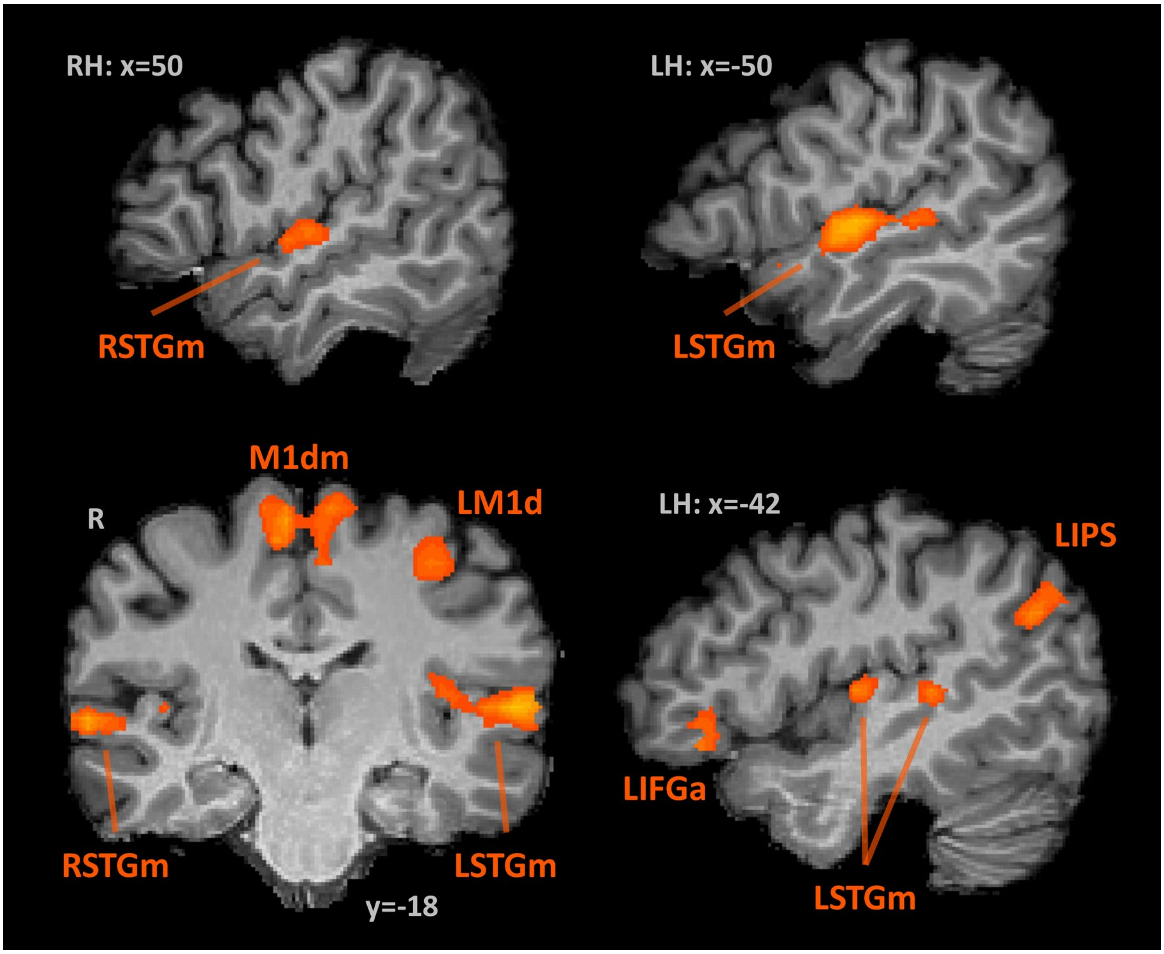

Fig. 1 Visualizes significant task verification brain areas: (Math Sentences > Tones) ∩ (Spatial Sentences > Tones) ∩ (Reading Sentences > Tones). For specific region details and abbreviations, see Table 1

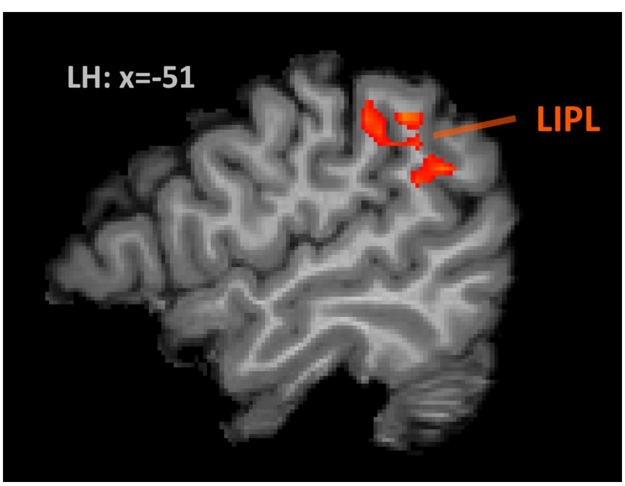

Fig. 2 Lone brain area showing significant univariate overlap for math and spatial semantic content: (Math Sentences > Reading Sentences) ∩ (Spatial Sentences > Reading Sentences). Region coordinates: − 51.3, − 41.4, 37.8 (volume: 891 mm^3)

unique math-space similarity than math-reading and space-reading similarity: bilateral anterior hippocampi (HPCa; mean *y*-coordinates were − 14 and − 13 for left and right hemispheres, respectively), bilateral Heschl's gyri (HG), and a cluster spanning bilateral caudate (head) and left nucleus accumbens (labeled simple 'striatum', or STRI for short). Regions are visualized in Fig. 3. Region details are shown in Table 2. Figure 4 shows average unique similarity values (averaged across participants) for each region. Figure 4 indicates that the planned contrast used in this analysis was not overly biased by just one of the similarity effects involving reading and that average unique math-space similarity values were positive in all 5 regions.

Primary analysis 2: similarity correlations with math behavioral outcomes

Our second goal was to test how math-space neural similarity predicts math outcomes at the behavioral level. Note that the current dataset is a relatively small sample for

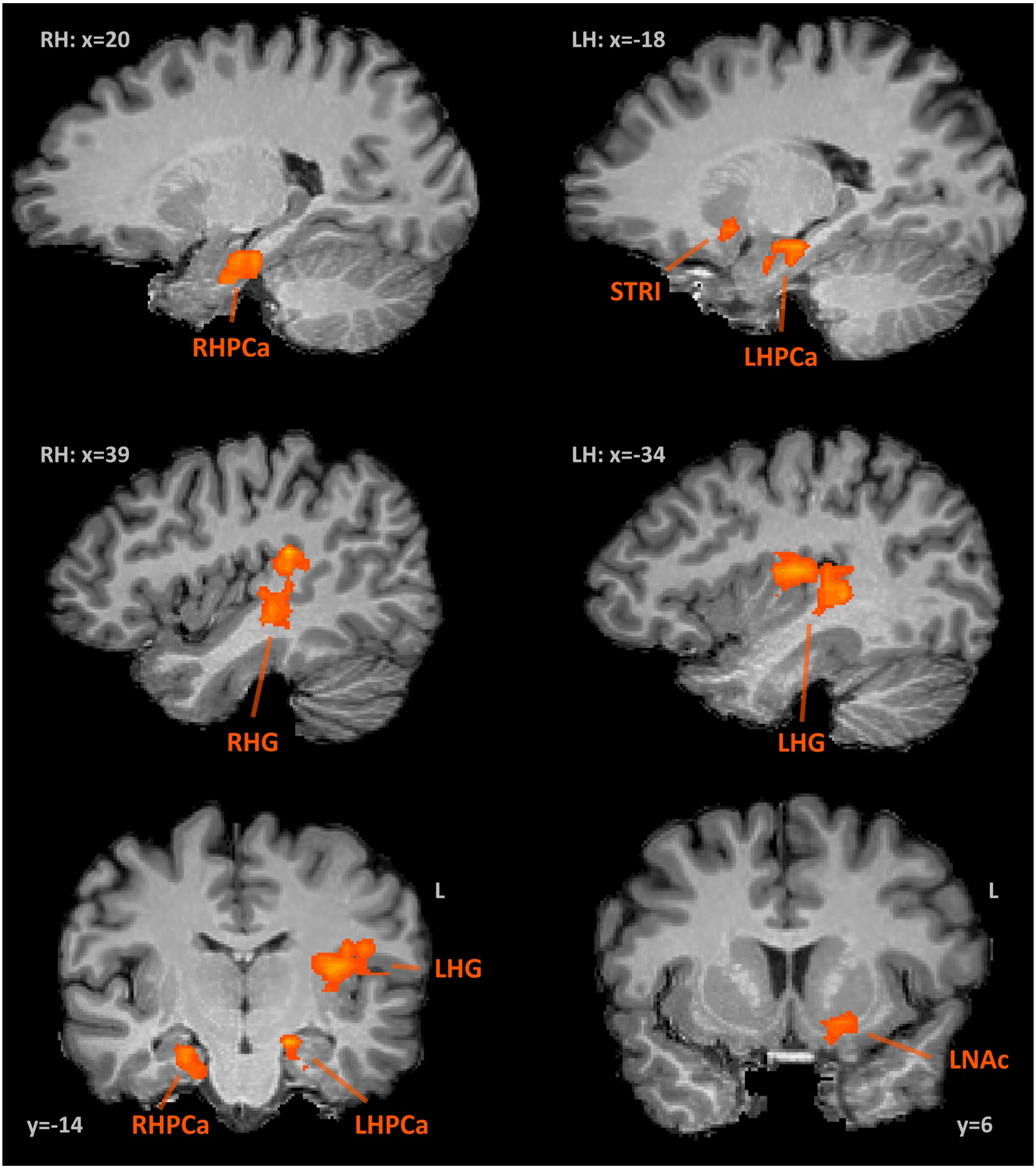

Fig. 3 Visualizes significant brain areas from whole-brain searchlight RSA. Regions show significantly greater math-space similarity than math-reading and space-reading similarity. For specific region details and abbreviations, see Table 2. For exact similarity values, see Fig. 4. Note that LNAc (left nucleus accumbens) is part of the larger STRI ROI

conducting correlation analyses, so sensitivity power analysis may provide useful cautionary context in interpreting the results that follow. Assuming a power of 0.80 and an alpha level of 0.05, the current sample ($N = 53$) was sufficient to detect zero-order correlation effects > 0.374, and partial correlation effects (with 7 covariates) > 0.401. See Appendix

Table 2 ROI anatomical details

	Mean x	Mean y	Mean z	Volume (mm^3)
LHPCa	− 17.7	− 14.0	− 14.6	856
RHPCa	19.4	− 13.3	− 18.4	1328
Striatum	− 8.1	14.0	− 4.0	3242
LHG	− 35.8	− 19.3	12.9	5907
RHG	39.7	− 29.3	13.4	1384

L/RHPCa left/right anterior hippocampus, *L/RHG* left/right Heschl's gyri, *STRI* striatum, *LNAc* left nucleus accumbens

Table 6 for the full correlation matrix of all behavioral variables, including covariates.

Our primary theoretical objective concerned math outcomes, so here we focus on math anxiety and math ability. We examined whether greater space-math similarity was associated with better (higher math performance, lower math anxiety) or worse (lower math performance, higher math anxiety) math outcomes. Figure 5a shows zero-order correlations between (z-transformed) math-space similarity values and math anxiety and math ability in each of the 5 regions from the whole-brain analysis above. Results showed that greater math-space similarity in bilateral anterior hippocampi (HPCa) predicted *higher* math anxiety and *lower* math ability.

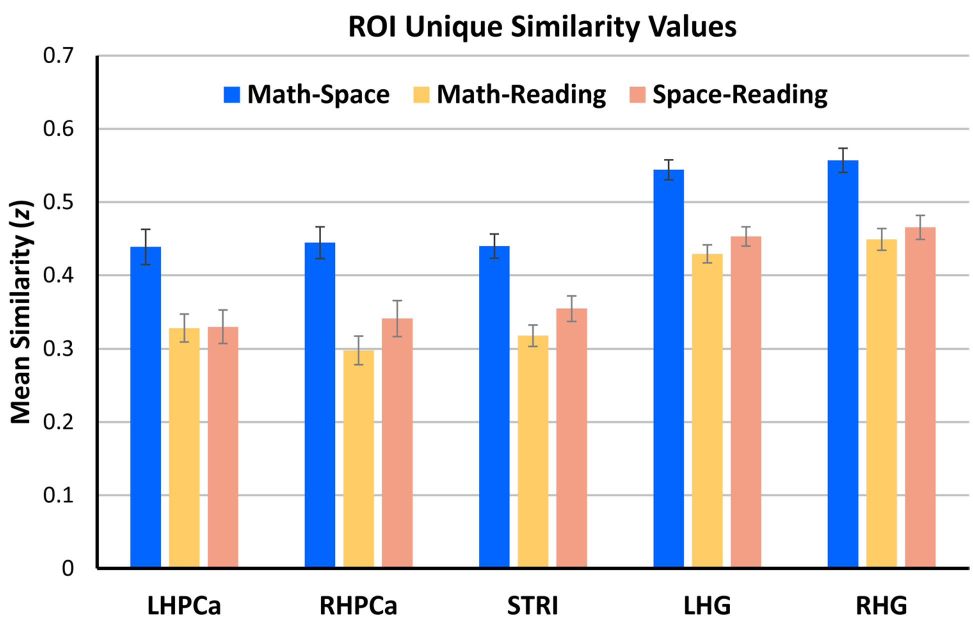

Fig. 4 Mean similarity values for each ROI. All similarity values represent unique relations: e.g., the math-space similarity value controls for the effect of reading (and tones). Similarity values were z-transformed (Fisher-z) and averaged across participants to generate the values shown above. Error bars reflect standard errors of the means. See Table 2 for abbreviations

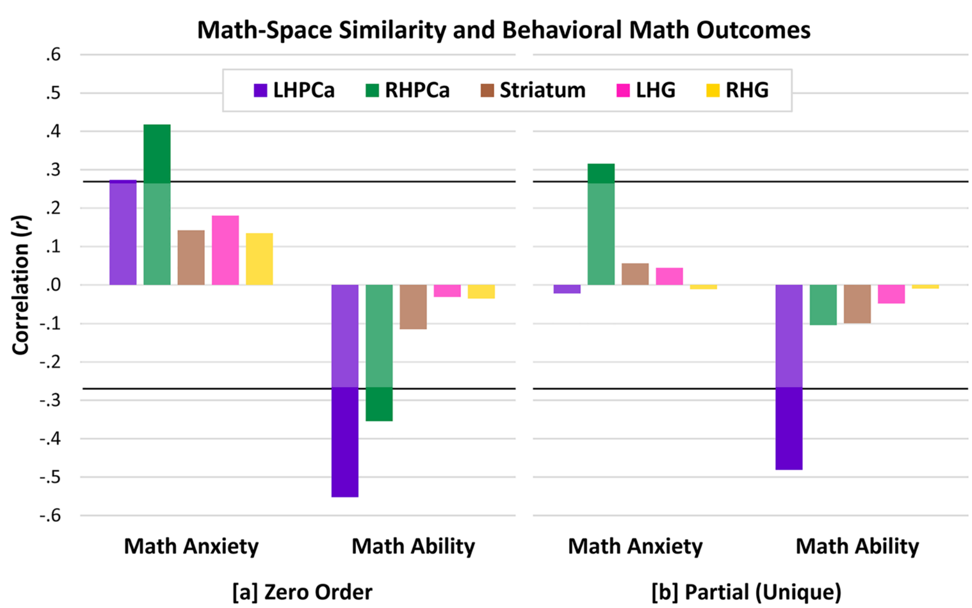

Fig. 5 Behavioral math outcomes correlated with math-space similarity values in each of the 5 regions from the whole-brain analysis. **a** Zero-order correlations; **b** partial correlations controlling for the mutual influence of math anxiety and math ability on one another, as well as spatial anxiety, spatial ability, trait anxiety, working memory, gender, and accuracy on the post-scan memory task. Horizontal black lines correspond to the ± r-value at $p = 0.05$. See Table 2 for ROI abbreviations. See Appendix Table 7 for exact r- and p-values

To improve specificity, we then tested whether these results were specific to math anxiety and math ability, *after controlling for one another*. In addition, we included several covariates: spatial anxiety, spatial ability, trait anxiety, working memory, gender, and accuracy on the post-scan memory task. Spatial anxiety and general anxiety controlled for non-math anxiety. Spatial ability and working memory controlled for non-math cognitive ability. We included gender as a covariate because prior work has shown gender differences in math anxiety (e.g., Sokolowski et al., 2019). Finally, we included post-scan memory accuracy to reduce the influence of individual variation in memory ability. Figure 5b shows a robust and unique relation between math-space similarity and math anxiety (positive effect) in the *right* HPCa and a robust and unique relation between math-space similarity and math ability (negative effect) in the *left* HPCa.

As a final check, we also tested whether math-space similarity was associated with behavioral variables in the spatial domain. Results showed no significant effects for spatial behavioral variables (ability or anxiety) in any of the 5 regions identified in the whole-brain RSA.

Discussion

Neural and behavioral evidence supports the idea that mathematical and spatial abilities are positively correlated and activate overlapping brain networks. Much of the literature exploring this relationship has examined behavioral variables or univariate neural activity during completion of standard lab tasks focused on cognitive performance. Together, this has led researchers to the general conclusion that math and spatial processing primarily overlap in the parietal cortex, and that interactions between math and space are beneficial for learning and performance (e.g., Hubbard et al., 2005; Newcombe et al., 2019). However, these lab-based tasks are representative of only a subset of contexts in which math and spatial thinking may be engaged, and therefore likely capture only a subset of the cognitive and affective processes that comprise our overall experiences with math and spatial thinking. In particular, lab-based tasks may be ill-equipped to measure and characterize other contexts in which math and spatial thinking are engaged, such as educational and everyday contexts.

To address this gap, we had participants engage in a domain-specific verbal memory task in which they listened to sentences describing a range of educational and everyday situations. Each sentence was designed to evoke domain-specific content, primarily related to either math or spatial content. We then used a neural similarity-based approach to assess the degree of unique integration between activity elicited while participants passively listened to and encoded these sentences in memory. While univariate results provided partial agreement with prior literature, RSA results implicated a different set of regions, including bilateral anterior hippocampi, bilateral Heschl's gyri, and the striatum. Also in contrast to previous work, behavioral data indicated that, in bilateral anterior hippocampi specifically, greater math-space similarity was related to *poorer* math performance and *higher* math anxiety. These results provide evidence for two things. First, there is, in fact, neural integration of math and space-related content during the encoding phase of a recognition memory task, although this similarity was generally observed in brain areas different from those previously reported. This suggests that use of an alternative type of task—one that in this case emphasized everyday math and spatial experience—captured a different subset of overlapping processes between math and spatial thinking than has been previously reported. Second, also in contrast to evidence from these lab-based measures, greater math-space neural integration during encoding of educational and everyday contexts was associated with *negative* math outcomes. Together, these results broaden our understanding of the neural integration between two key cognitive domains, and encourage the use of behavioral data to better characterize and validate neural findings. Additionally, these results underscore the value of broadening the scope of our investigations to include tasks which capture a wider variety of cognitive processes and contexts than those which are captured via the use of common lab tasks. By doing so, our results indicated that the implications of math-space integration may depend on task-specific contexts and mechanisms to a greater degree than previously realized.

Whole-brain search-light analyses demonstrated unique similarity between semantic comprehension of stimuli that evoke math and spatial content in educational and everyday situations. This similarity constituted partial correlations representing unique variance shared by spatial and mathematical semantic content, after controlling for auditory attention and content from a third domain (reading). Further, we showed not just the presence of unique (positive) math-space associations, but that the magnitude of these associations was significantly greater than the unique positive relations between math and reading content, and space and reading content. A close analogy would be a multiple regression analysis with, say, math skills as the DV, and spatial skills, reading skills and auditory attention as competing IVs. In the analogy, we showed the neural equivalent that (1) the unique spatial effect remained significant, and (2) this effect was significantly greater than the unique reading effect. Results revealed several brain areas that passed this stringent test at the whole-brain level: a large striatal activation, bilateral Heschl's gyri, and bilateral anterior hippocampi. The latter of these—hippocampus—is perhaps especially intriguing. On the one hand, it has been implicated in both mathematical and spatial processing (Bird & Burgess, 2008; Burgess

Table 3 Complete list of math sentence stimuli presented in the scanner

Type	Duration (ms)	Sentence
Math	2302	You receive a mathematics textbook
Math	3804	You watch your teacher work through an algebra problem on the board
Math	2978	You wait in line to sign up for a mathematics course
Math	4362	You listen to another student explain a complicated mathematical formula
Math	2284	You walk to a mathematics class
Math	2464	You study for a final mathematics test
Math	3936	You take the mathematical problem solving part of an achievement test
Math	3622	You read your cash register receipt after making a purchase
Math	2408	You have to take a mathematics quiz tomorrow
Math	2987	You take a final examination in a mathematics course
Math	3354	You are given a set of addition problems to solve on paper
Math	3644	You are presented with a set of subtraction problems to solve on paper
Math	3335	You are given a set of multiplication problems to solve on paper
Math	3479	You are presented with a set of division problems to solve on paper
Math	3780	You pick up your mathematics textbook to begin working on a homework assignment
Math	5206	You are given a homework assignment with many difficult mathematics problems, which is due next class
Math	2453	You have a mathematics test in one week
Math	3374	You have a mathematics test tomorrow, for which you feel unprepared
Math	2445	You have a mathematics test in one hour
Math	3956	You realize you must take several more mathematics courses in order to graduate
Math	3583	You pick up a mathematics textbook to begin a difficult assignment
Math	2969	You receive your final mathematics grade on your report card
Math	4098	You open your calculus homework assignment and see a page full of problems
Math	2617	You prepare to study for a mathematics test
Math	2905	You discover you have a "pop" quiz in mathematics class
Math	2592	You think about having to take a mathematics test
Math	2799	You decide whether to take a mathematics course
Math	2406	You receive your grade on a mathematics test
Math	2663	You imagine your grade in a mathematics course
Math	3672	You need to take one more mathematics course to graduate university

et al., 2002; Menon, 2016), though prior work using standard lab tasks has not shown this to be a major region showing consistent overlap in univariate math and space activation (Hawes et al., 2019). On the other hand, the hippocampus is associated with a range of cognitive and affective functions beyond math and spatial processing (Strange et al., 2014), making the risk of unwarranted reverse inference rather high (Poldrack, 2008). We, therefore, turned to behavioral data to provide a degree of external validation and thus adjudicate between potential functional inferences.

One interpretation of the hippocampal results in the present data pertains to cognitive demands of math processing. We know from previous research that the hippocampus is involved in multiple aspects of spatial thinking (Bird & Burgess, 2008; Burgess et al., 2002), as well as memory encoding and retrieval (Bird & Burgess, 2008; Squire, 1986). The hippocampus has also been implicated in multiple aspects of math learning (Fias et al., 2021; Menon, 2016; Van Opstal et al., 2008). If, in fact, the math-space neural integration observed here in the hippocampus reflects one or more of these cognitive components of math processing, then we would expect there to be evidence of a relationship between individual levels of math-space integration and individual levels of math ability. Further, one can surmise two potential relationships. Greater math-space integration in the hippocampus could be associated with higher or lower levels of math ability. Results indicated a *negative* association with math ability in bilateral anterior hippocampi, with the most robust effect obtained in the left hemisphere. These results indeed suggest hippocampal math-space integration in the current study has a cognitive basis, such that more similar neural responses to stimuli about everyday math and spatial

Table 4 Complete list of space sentence stimuli presented in the scanner

Type	Duration (ms)	Sentence
Space	2664	You receive directions without a map
Space	3918	You imagine a 3-D structure using a 2D drawing
Space	3170	You describe a radio announcer you have not seen
Space	5365	You are asked to give a detailed description of a person's face that you have only seen one time
Space	1919	You walk to an engineering class
Space	2419	You memorize maps for a geography test
Space	4055	You take a test where you create a drawing that reproduces a photo
Space	3528	You must imagine and mentally rotate a 3D figure
Space	2418	You must recreate a signature from memory
Space	3429	You imagine how the orbit of a comet changes over time
Space	3159	You are asked to do the navigational planning for a car trip
Space	3596	You are given a 2D drawing to determine how pulleys interact
Space	4354	You are asked to recall details of a person's tie that you saw yesterday
Space	3783	You must find your way to a meeting in a city you have never been to
Space	3527	You imagine the motion of a mechanical system given a static picture
Space	4480	You are trying to get to a place you have never been in the middle of an unfamiliar city
Space	2512	You try to take a short cut without a map
Space	4703	You must recall the details of a friend's face, whom you have not seen in years
Space	2260	You have an art history test in one hour
Space	4534	You memorize a picture then take a test where you must point out differences on a new picture
Space	4335	You imagine the 3D structure of a human brain from a 2D image
Space	3548	You imagine how gravity interacts with passing light beams
Space	4158	You are tested on how a 3D landscape would look from a different point of view
Space	2843	You must describe in detail the cover of a book
Space	3299	You must find your way back after becoming lost while driving
Space	2548	You think about having to take a navigation test
Space	2869	You decide whether to take an art history course
Space	2801	You receive your grade on a learner's driving test
Space	2752	You imagine your grade in an engineering course
Space	3467	You try to find your way to a class that is in an unfamiliar building

content in—*especially left*—hippocampus are associated with *poorer* math skills.

Another potential explanation for the integration of space and math in bilateral anterior hippocampi concerns the role of this brain area in affective processing. The hippocampus contributes to the regulation of the hypothalamic–pituitary–adrenal (HPA) axis (Jankord & Herman, 2008). In response to stressful situations, the HPA axis produces a cascade of effects which ultimately end in the production and release of glucocorticoids, or stress hormones, such as cortisol (Jacobson & Sapolsky, 1991). The hippocampus is densely packed with glucocorticoid receptors which detect the concentration of these stress hormones, in which case, the anterior hippocampus plays a key role in inhibiting the HPA axis, thereby regulating further release of these hormones (Jankord & Herman, 2008).

If math-space integration observed here in the anterior hippocampi reflects that structure's role in affective processing, then there should be evidence for a relationship between individual levels of math-space integration and affective measures, particularly individual levels of math anxiety. Here again, though, one can imagine two potential relationships. Greater math-space integration in the anterior hippocampus could be associated with either higher or lower levels of math anxiety. We thus once again turned to behavioral data to adjudicate between these possibilities. Results indicated a *positive* association with math anxiety in bilateral anterior hippocampi, with the most robust effect obtained in the right hemisphere. These results indeed suggest hippocampal math-space integration has an affective (as well as a cognitive) basis, such that more similar neural responses to stimuli about everyday math and spatial content in—*especially right*—anterior hippocampus is associated with *higher* math anxiety. These results are consistent with research showing that strong associations between number and space are predictive of higher levels of math anxiety (Georges et al., 2016).

Limitations

A key limitation of this work is that we did not measure brain activity while participants were actually engaging in educational or everyday activities; to do so using fMRI would be largely impossible. Instead, we approximated these experiences by measuring brain activity while participants encoded in memory sentences evoking either math or spatial content. These stimuli were presented within the context of a standard verbal memory task, which raises the question of whether the neural results reflect semantic understanding of the sentences' content. To this end, note that participants' memory accuracy for the sentences was relatively high, and all participants affirmed they were actively attending to the sentences, suggesting they accessed the sentence content. With respect to correlations with math outcomes, note that we controlled for accuracy on the memory task, indicating the brain-behavior results were unlikely to have been driven exclusively by memory-encoding processes. That said, it would be too strong to rule out memory-related processes altogether, especially in the case of the whole-brain similarity results.

It is also worth noting a second limitation of this work in that we were unable to control for participants' expertise or familiarity with the situations described in the task sentences. It is possible that semantic processing of such sentences may differ depending on how much expertise or familiarity a participant has with the scenario described in the sentence (e.g., "You have an art history test in one hour"). However, all participants were randomly sampled, which we believe mitigates this concern for the majority of our results (with perhaps an important caveat made for the

Table 5 Complete list of reading sentence stimuli presented in the scanner

Type	Duration (ms)	Sentence
Reading	2009	You receive a literature textbook
Reading	3583	You watch your teacher work through a reading comprehension problem
Reading	2961	You wait in line to sign up for a literature course
Reading	3991	You listen to another student explain a complicated English literary passage
Reading	1933	You walk to a literature class
Reading	2379	You study for a final literature test
Reading	4207	You take the English reading comprehension section of an achievement test
Reading	3704	You read all your emails and letters after a long vacation
Reading	2402	You have to take a literature quiz tomorrow
Reading	3013	You take a final examination in a literature course
Reading	3106	You are presented with a set of novels to read and analyze
Reading	3946	You are given a list of short fictional stories to read and analyze
Reading	3718	You are given a book of narrative poetry to read and analyze
Reading	3651	You are presented with a set of English essays to read and analyze
Reading	3679	You pick up your English poetry books to begin working on a homework assignment
Reading	4766	You are given a homework assignment with some difficult literary passages, which is due next class
Reading	2231	You have a literature test in one week
Reading	3506	You have a literature test tomorrow, for which you feel unprepared
Reading	2200	You have a literature test in one hour
Reading	4230	You realize you must take several more English essay courses in order to graduate
Reading	3815	You pick up an English grammar textbook to begin a difficult assignment
Reading	2860	You receive your English literature grade on your report card
Reading	4197	You open your English assignment and see a page full of difficult questions
Reading	2426	You prepare to study for a literature test
Reading	3037	You discover you have a "pop" quiz in literature class
Reading	2368	You think about having to take a literature test
Reading	2335	You decide whether to take a literature course
Reading	2348	You receive your grade on a literature test
Reading	2568	You imagine your grade in a literature course
Reading	3664	You need to take one more English essay course to graduate university

Table 6 Full zero-order correlation matrix between all behavioral variables (and gender)

	Math ability	Math Anxiety	Spatial ability	Spatial anxiety	Working memory	General anxiety	Memory task	Gender
Math ability		− 0.394 (0.003)	0.294 (0.033)	0.034 (0.806)	0.182 (0.192)	− 0.026 (0.853)	0.013 (0.930)	− 0.270 (0.051)
Math anxiety	− 0.394 (0.003)		− 0.196 (0.160)	0.448 (8E− 04)	0.141 (0.313)	0.497 (2E−04)	0.013 (0.924)	0.392 (0.004)
Spatial ability	0.294 (0.033)	− 0.196 (0.160)		− 0.299 (0.030)	0.277 (0.045)	0.013 (0.924)	0.178 (0.208)	− 0.388 (0.004)
Spatial anxiety	0.034 (0.806)	0.448 (8E−04)	− 0.299 (0.030)		− 0.113 (0.420)	0.313 (0.023)	0.006 (0.965)	0.542 (3E−05)
Working memory	0.182 (0.192)	0.141 (0.313)	0.277 (0.045)	− 0.113 (0.420)		− 0.070 (0.617)	0.078 (0.582)	− 0.275 (0.047)
General anxiety	− 0.026 (0.853)	0.497 (2E−04)	0.013 (0.924)	0.313 (0.023)	− 0.070 (0.617)		0.160 (0.256)	0.372 (0.006)
Memory task	0.013 (0.930)	0.013 (0.924)	0.178 (0.208)	0.006 (0.965)	0.078 (0.582)	0.160 (0.256)		− 0.045 (0.753)
Gender	− 0.270 (0.051)	0.392 (0.004)	− 0.388 (0.004)	0.542 (3E− 05)	− 0.275 (0.047)	0.372 (0.006)	− 0.045 (0.753)	

Values in parentheses are p-values. Gender is coded as female = 1, male = 0

individual differences results shown in Fig. 5). Altogether, despite the above limitations, we believe these data nevertheless extend our understanding of math-space relations to include comprehension of linguistic stimuli depicting educational and everyday content.

Broader conclusions

Here we take our discussion of the results in a somewhat more speculative direction, but given their surprising nature (greater math-space similarity associated with poorer math outcomes), such speculation may be warranted. If we were to choose a single, unifying interpretation, it would be that the neural similarity invoked by comprehending the everyday and educational sentences reflects simulated episodic experiences. Recent work using intracranial EEG demonstrated Heschl's gyrus responds systematically to continuous natural speech stories (Khalighinejad et al., 2021). The hippocampus is centrally involved in construction of simulated episodic experiences, much as one might do when constructing semantic understanding of a sentence describing a quotidian situation (Addis et al., 2011; Schacter et al., 2008; Thakral et al., 2017, 2020). Here it may be useful to emphasize that in the current study all sentences were phrased in the second-person—e.g., "You study for a final mathematics test," or "You memorize maps for a geography test." To us therefore, the most convincing interpretation of the current data is thus that simulating everyday episodic experiences of everyday math-related contexts and space-related contexts is underlain by similar neural processes, as revealed by the whole-brain similarity results in Fig. 3. Furthermore, the more similar these simulations are, at least in the

Table 7 Exact *r*- and *p*-values, along with 95% confidence intervals for *r*-values, for the results shown in Fig. 5a, b

Math outcome: math anxiety					Math outcome: math ability				
Region	r	p	CI+	CI−	Region	r	p	CI +	CI−
Zero-order correlations (from Fig. 5a)									
LHPCa	0.273	0.048	0.003	0.506	LHPCa	− 0.552	2E−05	0.003	− 0.331
RHPCa	0.418	0.002	0.166	0.618	RHPCa	− 0.354	0.009	0.166	− 0.093
Striatum	0.142	0.309	− 0.133	0.397	Striatum	− 0.115	0.413	− 0.133	0.160
LHG	0.181	0.196	− 0.094	0.430	LHG	− 0.031	0.826	− 0.094	0.241
RHG	0.135	0.336	− 0.141	0.391	RHG	− 0.036	0.801	− 0.141	0.237
Partial correlations (from Fig. 5b)									
LHPCa	− 0.022	0.885	− 0.310	0.270	LHPCa	− 0.481	7E−04	− 0.310	− 0.222
RHPCa	0.316	0.033	0.028	0.555	RHPCa	− 0.104	0.492	0.028	0.192
Striatum	0.056	0.711	− 0.238	0.341	Striatum	− 0.099	0.512	− 0.238	0.197
LHG	0.044	0.771	− 0.249	0.330	LHG	− 0.049	0.749	− 0.249	0.245
RHG	− 0.011	0.940	− 0.301	0.280	RHG	− 0.010	0.948	− 0.301	0.281

hippocampus, the higher one's math anxiety and the lower one's math ability. Episodic simulation integrates information from a variety of neural sources, which in turn may help explain why hippocampal math-space neural similarity is related to both affective (math anxiety) and cognitive (math ability) aspects of math.

Taking this view a step further, one might ask how it might come to pass. Episodic simulation is thought to be a form of memory-based simulation, in which case the current results would depend in part on an individual's past experiences with math and spatial contexts. The fact that greater math-space similarity was associated with negative math outcomes may be a reflection of the amount of experience an individual has with math. The more experience one has with something, the more distinct their memories of that thing are likely to be. Therefore, if an individual has more experience with math, they may simply have more distinct memories related to math than to space. Of course, though bolstered by brain behavior correlations, this interpretation is at its core based on reverse inference, and so must remain speculative barring further evidence.

Another interpretation is that the RSA results reflect core cognitive and affective aspects of math and spatial processing. On the one hand, this would seem to contradict prior lab-based studies that showed math-space overlap in other brain areas, such as prefrontal and parietal cortex (Hawes et al., 2019). Here is where we point out that prior work relied not just on lab-based tasks, but on univariate overlap. When we used a similar technique (Preliminary Analysis 2, Fig. 2), we too found co-activation of left parietal cortex. However, as noted above, univariate overlap is a weak test of shared neural mechanism. Hence, an open question is whether future work using traditional lab-based studies, but a technique allowing for stronger inferences about shared neural mechanisms—such as the RSA approach used here—would find evidence for math-space similarity in regions more in line with what we showed here.

With regard to bolstering inferences about underlying mechanism, had we not used behavioral data to inform our interpretation of our neural results, we might have been biased to confirm expectations based on previous literature. In particular, prior work relying primarily on standard lab-based tasks has demonstrated math-space integration is generally positive for math outcomes (e.g., Hubbard et al., 2005; Newcombe et al., 2019). On its surface, then, the significant whole-brain math-space similarity results might have been interpreted as simply extending assumptions about the benefits of math-space integration to stimuli depicting educational and everyday contexts as well. However, our brain-behavior results indicate the opposite. Whether greater math-space integration is predictive of positive or negative math outcomes may depend crucially on whether one is examining processes that subserve task-based performance, or processes that subserve the generation of episodic experience.

In sum, prior work examining standard lab-based measures of math and space indicates greater integration in those contexts may be beneficial for math outcomes; here, our results suggest that greater integration when processing more educational or everyday content is predictive of poorer math outcomes. For this reason, it is imperative for future researchers to use behavioral data to reduce the risk of confirmation bias when interpreting neural data. Moreover, interventions explicitly aimed at integrating math and spatial processing might do well to consider implications of their intervention that go beyond the lab and perhaps impact more quotidian experiences.

Appendix 1

See Tables 3, 4 and 5.

Appendix 2

See Tables 6 and 7.

Acknowledgements This work was supported by the National Institute of Child Health and Development (1R01-HD100429-01A1) to Lyons, Faculty Start-up funds to Lyons (Georgetown University), Banting Postdoctoral Fellowship to Lyons (National Sciences and Engineering Research Council, NSERC, Canada), and University of Western Ontario internal research award to Lyons (Western Research and Faculty of Social Science). NSERC Alexander Graham Bell Canada Graduate Scholarships-Doctoral Program (CGS-D) to Sokolowski. National Science Foundation (NSF) – Education and Human Resources (EHR): CAREER-2041887 to Lyons.

Author contributions Conceptualization—I.M.L. Methodology—I.M.L and H.M.S. Formal analysis and investigation—I.M.L. Writing, original draft—R.N.M and I.M.L. Writing, review & editing—all authors funding acquisition—I.M.L and D.A. Resources—D.A. Supervision—I.M.L and D.A.

Data availability Data for this project can be found on the Open Science Framework (OSF) at https://osf.io/zubq8/.

Declarations

Conflict of interest The authors declare no competing interests.

References

Addis, D. R., Cheng, T., Roberts, R. P., & Schacter, D. L. (2011). Hippocampal contributions to the episodic simulation of specific and general future events. *Hippocampus, 21*(10), 1045–1052. https://doi.org/10.1002/hipo.20870

Alexander, L., & Martray, C. R. (1989). The development of an abbreviated version of the Mathematics Anxiety Rating Scale. *Measurement and Evaluation in Counseling and Development,* *22*(3), 143–150. https://doi.org/10.1080/07481756.1989.12022923

Arbuckle, S. A., Yokoi, A., Pruszynski, J. A., & Diedrichsen, J. (2019). Stability of representational geometry across a wide range of fMRI activity levels. *NeuroImage, 186*, 155–163. https://doi.org/10.1016/j.neuroimage.2018.11.002

Atit, K., Power, J. R., Pigott, T., Lee, J., Geer, E. A., Uttal, D. H., Ganley, C. M., & Sorby, S. A. (2022). Examining the relations between spatial skills and mathematical performance: A meta-analysis. *Psychonomic Bulletin & Review, 29*(3), 699–720. https://doi.org/10.3758/s13423-021-02012-w

Bird, C. M., & Burgess, N. (2008). The hippocampus and memory: Insights from spatial processing. *Nature Reviews Neuroscience, 9*(3), 182–194. https://doi.org/10.1038/nrn2335

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron, 35*(4), 625–641. https://doi.org/10.1016/S0896-6273(02)00830-9

Caissie, A. F., Vigneau, F., & Bors, D. A. (2009). What does the Mental Rotation Test measure? An analysis of item difficulty and item characteristics. *The Open Psychology Journal, 2*, 94–102. https://doi.org/10.2174/1874350100902010094

Calabria, M., & Rossetti, Y. (2005). Interference between number processing and line bisection: A methodology. *Neuropsychologia, 43*(5), 779–783. https://doi.org/10.1016/j.neuropsychologia.2004.06.027

Calzavarini, F., & Cevolani, G. (2022). Abductive reasoning in cognitive neuroscience: Weak and strong reverse inference. *Synthese, 200*, 70. https://doi.org/10.1007/s11229-022-03585-2

Chang, E. F., Raygor, K. P., & Berger, M. S. (2015). Contemporary model of language organization: An overview for neurosurgeons. *Journal of Neurosurgery, 122*(2), 250–261. https://doi.org/10.3171/2014.10.jns132647

Chatham, C. H., & Badre, D. (2019). How to test cognitive theory with fMRI. In D. Spieler & E. Schumacher (Eds.), *New methods in cognitive psychology* (pp. 86-127). Routledge.

Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development, 15*(1), 2–11. https://doi.org/10.1080/15248372.2012.725186

Coltheart, M. (2013). How can functional neuroimaging inform cognitive theories? *Perspectives on Psychological Science, 8*(1), 98–103. https://doi.org/10.1177/1745691612469208

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769–786. https://doi.org/10.3758/bf03196772

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

Dimsdale-Zucker, H. R., & Ranganath, C. (2018). Chapter 27—representational similarity analyses: A practical guide for functional MRI applications. In D. Manahan-Vaughan (Ed.), *Handbook of behavioral neuroscience* (Vol. 28, pp. 509–525). Elsevier. https://doi.org/10.1016/B978-0-12-812028-6.00027-6

Đokić, R., Koso-Drljević, M., & Đapo, N. (2018). Working memory span tasks: Group administration and omitting accuracy criterion do not change metric characteristics. *PLoS ONE, 13*(10), Article e0205169. https://doi.org/10.1371/journal.pone.0205169

Enge, A., Abdel Rahman, R., & Skeide, M. A. (2021). A meta-analysis of fMRI studies of semantic cognition in children. *NeuroImage, 241,* 118436. https://doi.org/10.1016/j.neuroimage.2021.118436

Ferré, P. (2003). Effects of level of processing on memory for affectively valenced words. *Cognition and Emotion, 17*(6), 859–880. https://doi.org/10.1080/02699930244000200

Fias, W., Sahan, M. I., Ansari, D., & Lyons, I. M. (2021). From counting to retrieving: Neural networks underlying alphabet arithmetic learning. *Journal of Cognitive Neuroscience, 34*(1), 16–33. https://doi.org/10.1162/jocn_a_01789

Fischer, M. H. (2001). Number processing induces spatial performance biases. *Neurology, 57*(5), 822–826. https://doi.org/10.1212/wnl.57.5.822

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience, 6*(6), 555–556. https://doi.org/10.1038/nn1066

Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition—from single digits to arithmetic. *The Quarterly Journal of Experimental Psychology, 67*(8), 1461–1483. https://doi.org/10.1080/17470218.2014.927515

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine, 33*(5), 636–647. https://doi.org/10.1002/mrm.1910330508

Galton, F. (1880). Visualised numerals. *Nature, 21*, 252–256. https://doi.org/10.1038/021252a0

Georges, C., Hoffmann, D., & Schiltz, C. (2016). How math anxiety relates to number-space associations. *Frontiers in Psychology, 7*, Article 1401. https://doi.org/10.3389/fpsyg.2016.01401

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology, 48*(5), 1229–1241. https://doi.org/10.1037/a0027433

Hawes, Z., & Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. *Psychonomic Bulletin & Review, 27*(3), 465–482. https://doi.org/10.3758/s13423-019-01694-7

Hawes, Z., Sokolowski, H. M., Ononye, C. B., & Ansari, D. (2019). Neural underpinnings of numerical and spatial cognition: An fMRI meta-analysis of brain regions associated with symbolic number, arithmetic, and mental rotation. *Neuroscience and Biobehavioral Reviews, 103*, 316–336. https://doi.org/10.1016/j.neubiorev.2019.05.007

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience, 6*(6), 435–448. https://doi.org/10.1038/nrn1684

Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *NeuroImage, 84*, 1061–1069. https://doi.org/10.1016/j.neuroimage.2012.12.075

Jacobson, L., & Sapolsky, R. (1991). The role of the hippocampus in feedback regulation of the hypothalamic-pituitary-adrenocortical axis. *Endocrine Reviews, 12*(2), 118–134. https://doi.org/10.1210/edrv-12-2-118

Jankord, R., & Herman, J. P. (2008). Limbic regulation of hypothalamo–pituitary–adrenocortical function during acute and chronic stress. *Annals of the New York Academy of Sciences, 1148*(1), 64–73. https://doi.org/10.1196/annals.1410.012

Karlsson, K., Sikström, S., & Willander, J. (2013). The semantic representation of event information depends on the cue modality: An instance of meaning-based retrieval. *PLoS ONE, 8*(10), e73378. https://doi.org/10.1371/journal.pone.0073378

Kaufmann, L., Vogel, S. E., Wood, G., Kremser, C., Schocke, M., Zimmerhackl, L.-B., & Koten, J. W. (2008). A developmental fMRI study of nonsymbolic numerical and spatial processing. *Cortex, 44*(4), 376–385. https://doi.org/10.1016/j.cortex.2007.08.003

Khalighinejad, B., Patel, P., Herrero, J. L., Bickel, S., Mehta, A. D., & Mesgarani, N. (2021). Functional characterization of human Heschl's gyrus in response to natural speech. *NeuroImage, 235*, 118003. https://doi.org/10.1016/j.neuroimage.2021.118003

Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the onion of brain representations. *Annual Review of Neuroscience, 42*, 407–432. https://doi.org/10.1146/annurev-neuro-080317-061906

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: Re-balancing the scale. *Social Cognitive and Affective Neuroscience, 4*(4), 423–428. https://doi.org/10.1093/scan/nsp052

Lyons, I. M., Ramirez, G., Maloney, E. A., Rendina, D. N., Levine, S. C., & Beilock, S. L. (2018). Spatial anxiety: A novel questionnaire with subscales for measuring three aspects of spatial anxiety. *Journal of Numerical Cognition, 4*(3), 526–553. https://doi.org/10.5964/jnc.v4i3.154

Menon, V. (2016). Chapter 7—Memory and cognitive control circuits in mathematical cognition and learning. In M. Cappelletti & W. Fias (Eds.), *Progress in Brain Research* (Vol. 227, pp. 159–186). Elsevier. https://doi.org/10.1016/bs.pbr.2016.04.026

Mix, K. S., & Cheng, Y.-L. (2012). Chapter 6—the relation between space and math: Developmental and educational implications. In J. B. Benson (Ed.), *Advances in child development and behavior* (Vol. 42, pp. 197–243). Elsevier Academic Press. https://doi.org/10.1016/b978-0-12-394388-0.00006-x

Newcombe, N. S., Booth, J. L., & Gunderson, E. A. (2019). 5—Spatial skills, reasoning, and mathematics. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 100–123). Cambridge University Press. https://doi.org/10.1017/9781108235631.006

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Poldrack, R. A. (2008). The role of fMRI in cognitive neuroscience: Where do we stand? *Current Opinion in Neurobiology, 18*(2), 223–227. https://doi.org/10.1016/j.conb.2008.07.006

Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature, 526*, 371–379. https://doi.org/10.1038/nature15692

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences, 1191*(1), 62–88. https://doi.org/10.1111/j.1749-6632.2010.05444.x

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3), 164–171. https://doi.org/10.1027/1015-5759/a000123

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An activation likelihood estimation meta-analysis. *Brain and Language, 141*, 89–102. https://doi.org/10.1016/j.bandl.2014.11.012

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2008). Episodic simulation of future events: Concepts, data, and applications. *Annals of the New York Academy of Sciences, 1124*(1), 39–60. https://doi.org/10.1196/annals.1440.001

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Simon, O., Kherif, F., Flandin, G., Poline, J.-B., Rivière, D., Mangin, J.-F., Le Bihan, D., & Dehaene, S. (2004). Automatized clustering and functional geometry of human parietofrontal networks for language, space, and number. *NeuroImage, 23*(3), 1192–1202. https://doi.org/10.1016/j.neuroimage.2004.09.023

Sokolowski, H. M., Hawes, Z., & Lyons, I. M. (2019). What explains sex differences in math anxiety? A closer look at the role of spatial processing. *Cognition, 182*, 193–212. https://doi.org/10.1016/j.cognition.2018.10.005

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.

Squire, L. R. (1986). Mechanisms of memory. *Science, 232*(4758), 1612–1619. https://doi.org/10.1126/science.3086978

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*(2), 245–251. https://doi.org/10.1037/0033-2909.87.2.245

Strange, B. A., Witter, M. P., Lein, E. S., & Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews Neuroscience, 15*(10), 655–669. https://doi.org/10.1038/nrn3785

Thakral, P. P., Benoit, R. G., & Schacter, D. L. (2017). Characterizing the role of the hippocampus during episodic simulation and encoding. *Hippocampus, 27*(12), 1275–1284. https://doi.org/10.1002/hipo.22796

Thakral, P. P., Madore, K. P., Addis, D. R., & Schacter, D. L. (2020). Reinstatement of event details during episodic simulation in the hippocampus. *Cerebral Cortex, 30*(4), 2321–2337. https://doi.org/10.1093/cercor/bhz242

Van Opstal, F., Verguts, T., Orban, G. A., & Fias, W. (2008). A hippocampal-parietal network for learning an ordered sequence. *NeuroImage, 40*(1), 333–341. https://doi.org/10.1016/j.neuroimage.2007.11.027

Woolrich, M. W. (2012). Bayesian inference in FMRI. *NeuroImage, 62*(2), 801–810. https://doi.org/10.1016/j.neuroimage.2011.10.047

Young, C. J., Levine, S. C., & Mix, K. S. (2018). The connection between spatial and mathematical ability across development. *Frontiers in Psychology, 9*, Article 755. https://doi.org/10.3389/fpsyg.2018.00755

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. https://doi.org/10.1016/j.neuropsychologia.2008.03.001