The neural correlates of logical-mathematical symbol systems processing resemble those of spatial cognition more than language processing

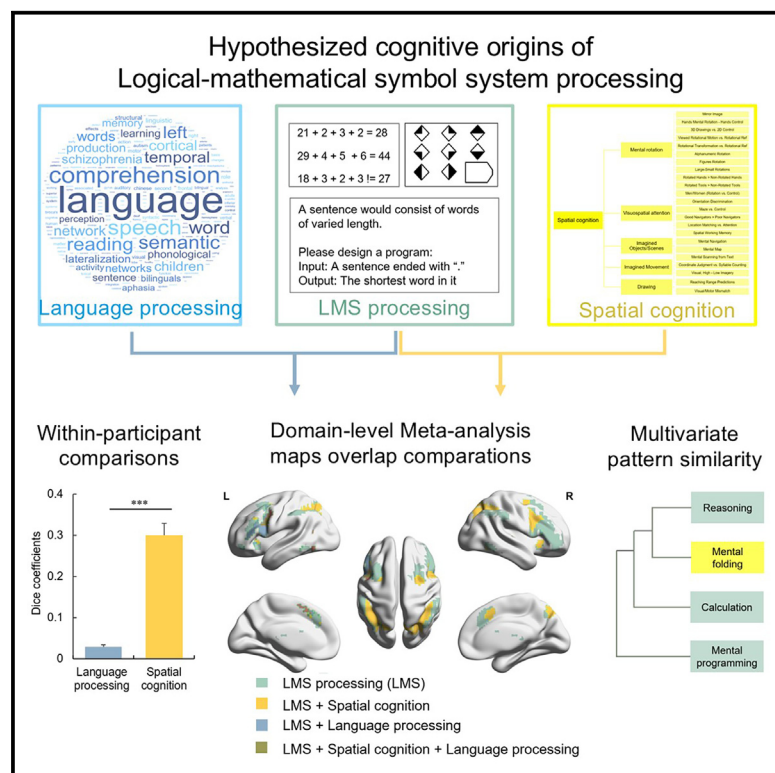## Graphical abstract

Authors

Yuannan Li, Shan Xu, Jia Liu

Correspondence

shan.xu@bnu.edu.cn

In brief

Neuroscience; Cognitive neuroscience

Highlights

- A direct domain-level comparison between two candidate origins of LMS processing

- Integrating meta-analyses and inhouse fMRI study of representative LMS tasks

- Greater activation overlap and pattern similarity between LMS and spatial cognition

- Spatial cognition, rather than language, likely underpins LMS processing

CellPress

Article

The neural correlates of logical-mathematical symbol systems processing resemble those of spatial cognition more than language processing

Yuannan Li,[1] Shan Xu,[2,3,*] and Jia Liu[1]
[1]Tsinghua Laboratory of Brain and Intelligence, Department of Psychological and Cognitive Sciences, Tsinghua University, Beijing 100086, China
[2]Faculty of Psychology, Beijing Normal University, Beijing 100875, China
[3]Lead contact
*Correspondence: shan.xu@bnu.edu.cn
https://doi.org/10.1016/j.isci.2025.112016

SUMMARY

The ability to use logical-mathematical symbols (LMS), encompassing tasks such as calculation, reasoning, and programming, is special to humans with recent emergence. LMS processing was suggested to build upon fundamental cognitive systems through neuronal recycling, with natural language processing and spatial cognition as key candidates. This study used meta-analyses and synthesized neural maps of representative LMS tasks, including reasoning, calculation, and mental programming, to compare their neural correlates with those of the two systems. Our results revealed greater activation overlap and multivariate similarity between LMS and spatial cognition than with language processing. Hierarchical clustering further indicated that LMS tasks were indistinguishable from spatial tasks at the neural level, suggesting an inherent connection. Our findings support the hypothesis that spatial cognition is the basis of LMS processing, shedding light on the logical reasoning limitations of large language models, particularly those lacking explicit spatial representations.

INTRODUCTION

Throughout its relatively brief history, humans have devised and mastered numerous logical-mathematical symbol (LMS) systems such as logical reasoning and calculation, reflecting our cultural and technological evolution. This trend continues with the creation of novel LMS systems with emerging knowledge domains such as programming, showcasing the remarkable adaptability of the human mind to increasingly complex cognitive challenges. In addition, the use of LMS systems to tackle abstract problems is attracting further interest, particularly for its potential to bridge the intellectual gap between artificial intelligence (AI) and human cognition. Despite its diverse applications and functionalities, LMS processing is theorized to originate from a single foundational system, namely the logical-mathematical intelligence, LM intelligence. As delineated by Gardner,[1] LM intelligence encompasses the ability to develop equations and proofs, make calculations, and solve abstract problems.[2] This capacity enables individuals to identify patterns, reason logically, and engage in analytical thinking, forming the basis for activities ranging from solving puzzles to advanced programming and conducting scientific analyses.[3–5] Accordingly, three functional domains are identified as the core cognitive units of LM intelligence, namely calculation, logical reasoning, and problem-solving. These domains are considered the pillars of the diverse tasks that recruit LMS intelligence.[1,5–8] However, the exact

cognitive origin of this foundational system remains largely elusive. In particular, considering that known LMS-related behaviors have emerged relatively recently, within the last 6000 years,[9] it is unlikely that a completely new set of neural facilities evolved specifically for this purpose. Accordingly, LMS processing may have evolved by repurposing phylogenetically more ancient cognitive systems through neuronal recycling.[10–13]

The primary systems thought to underlie LMS processing are natural language processing and spatial cognition.[11,12] Language processing and spatial cognition are both essential cognitive modules of human intelligence that support distinct cognitive functions. They involve distinct cognitive processes, a good manifestation being the separate subsidiary systems (i.e., the phonological loop and the visuospatial sketchpad, alongside the central attention controller) in working memory[14] dedicated to each of them. Language processing is primarily associated with left-hemispheric regions, including the precentral and the inferior frontal gyri, the supplementary motion area, and the superior and middle temporal gyri,[15,16] while spatial cognition is largely bilateral, particularly in the dorsal frontoparietal regions.[17] Despite these functional and neural distinctions, the two systems are also highly interconnected, as shown in the binding of verbal and spatial information in working memory tasks.[18,19] As such, isolating their respective contributions can be challenging. In terms of form, LMS systems bear an ostensible resemblance with natural language.[12,20] This similarity is especially evident

in the syntax-like operations of LMS, which are thought to parallel those in language processing.[12] For example, the verbal reasoning approach, which views deductive reasoning, a representative LMS task, as a transformation of verbal information, relies on processes similar to those in language comprehension and generation.[21,22] However, empirical evidence regarding the association between language and LMS processing is inconclusive, both in terms of correlation in behavioral performance[8,23–25] and identification of common neural correlates,[26–36] implying the complexity of the relationship between LMS and language systems. In addition, the comparison is limited to a handful of representative tasks, such as prose cloze versus code function judgment, sentence reading versus mental programming, and syllable concatenation versus number addition,[37–39] and the specifics of corresponding task settings risk confounding the test of commonality (but see Castelhano et al.[27] for a comparison between program understanding with the meta-analysis results of reading).

An alternative hypothesis proposes that LMS processing is built upon spatial cognition, which views the processing of the properties and relations among non-spatial LMS elements dependent on spatial abilities.[11,40] For instance, the mental model theory argues that reasoning is achieved by manipulating reasoning components in a visuospatial workspace, rather than a linguistic medium.[41] Similarly, the mental simulation theory posits that the ability to navigate physical environments underpins the simulation processes used to arrive at solutions in a logical-mathematical space.[42,43] Supporting these theoretical models is behavioral and neural evidence linking LMS processing abilities, such as numerical, arithmetical, and geometric abilities,[30,44,45] logical reasoning,[30,46,47] programming,[48–50] computational thinking,[37,51] and other STEM domains[52–54] to spatial ability. For instance, neural activation partially overlaps during spatial working memory and numerical working memory tasks.[43,55,56] This suggests that spatial cognition may play a critical role in facilitating LMS processing, offering an important perspective alongside the language-based approach. However, similar to the situation with the language hypothesis, only a handful of specific LMS and spatial tasks were utilized in the comparison.[38,39,49,57] For example, in our previous study,[37] to infer the cognitive substrates of computational thinking, we broke up the neural correlates of mental programming according to their varied pattern of neural resemblance with three other LMS tasks and a representative visuospatial task (mental folding), and we observed that the part showing resemblance between mental programming and other LMS tasks does not necessarily show high resemblance with mental folding. This leaves open a question of whether the hypothetical neural resemblance between LMS and spatial processing exists above task variation, being evident at the domain level, and whether such neural resemblance is constantly larger than the LMS-language resemblance across tasks.

To examine these two hypotheses at the domain level, here we compared the involvement of cortical regions associated with language processing and spatial cognition in LMS processing, respectively. Our rationale, grounded in the concept of neuronal recycling,[10,11] is that the basis of LMS processing is more likely to be the cognitive function whose cortical regions are predominantly engaged in LMS processing. The present study presented an attempt at direct and quantitative domain-level examination of the two hypotheses.

Instead of comparing the activation of specific tasks, we ran large-scale automated meta-analyses to identify the cross-task neural correlates of language processing and spatial cognition, respectively. For LMS processing, with few dedicated studies, we operationalized it by its three core cognitive units,[1] namely, calculation, logical reasoning, and problem-solving, and compiled a synthesized activation map through a conjunctive analysis of whole-brain univariate activations based on inhouse data of three representative tasks of each core unit. The representative task for calculation is number addition, adopted from Fedorenko et al.[58] The representative task for logical reasoning is Raven's reasoning test, a visual abstract reasoning task adopted from Raven's Advanced Progressive Matrices,[59] which had been widely used in assessing reasoning ability.[60,61] The representative task for problem-solving is mental programming, which focuses on problem-solving and algorithm generation[37] and is adopted from elementary program training problems.

With these meta-analysis results, we quantified and compared the domain-level overlap of LMS processing with language processing and spatial cognition, respectively. We indeed observed greater domain-level resemblance of LMS with spatial processing than with language processing, the cross-task homogeneity of which was further validated by multivariate activation pattern analysis. Finally, we used clustering analysis to explore the hierarchical structure among the neural correlates of LMS processing and its functional bases. These analyses collectively revealed a notable link between LMS processing and spatial cognition, supporting the hypothesis that the root of LMS processing lies in spatial cognition.

RESULTS

Univariate analysis revealed more overlap of LMS processing with spatial cognition

We first investigated whether LMS processing shared neural activation to a greater extent with language processing or spatial cognition. To identify the neural correlates intrinsic to LMS processing, distinct from those linked to task-specific properties, we conducted a conjunctive analysis using fMRI data from three representative tasks: reasoning, calculation, and mental programming (see Methods 4.2.2) (Figure 1A, middle). This synthesis allowed the generation of an activation map associated with LMS processing embedded across these three tasks. The resulting map revealed bilateral activation, in the frontal lobe (the middle frontal gyrus, the inferior frontal gyrus), the supplementary motion area spanning to the cingulum, the insula, the posterior parietal cortex, the precuneus, and the inferior temporal gyrus (Figure S1; Table S1).

To map the neural substrates of language processing and spatial cognition, we leveraged large-scale automated meta-analyses to generate cross-task activation maps for each (see Methods 4.1) (Figure 1A, left and right). Language processing activates a left-lateralized network, including the precentral and the inferior frontal gyri, the supplementary motion area, and the superior and middle temporal gyri (Figure S2; Table S2). In contrast,
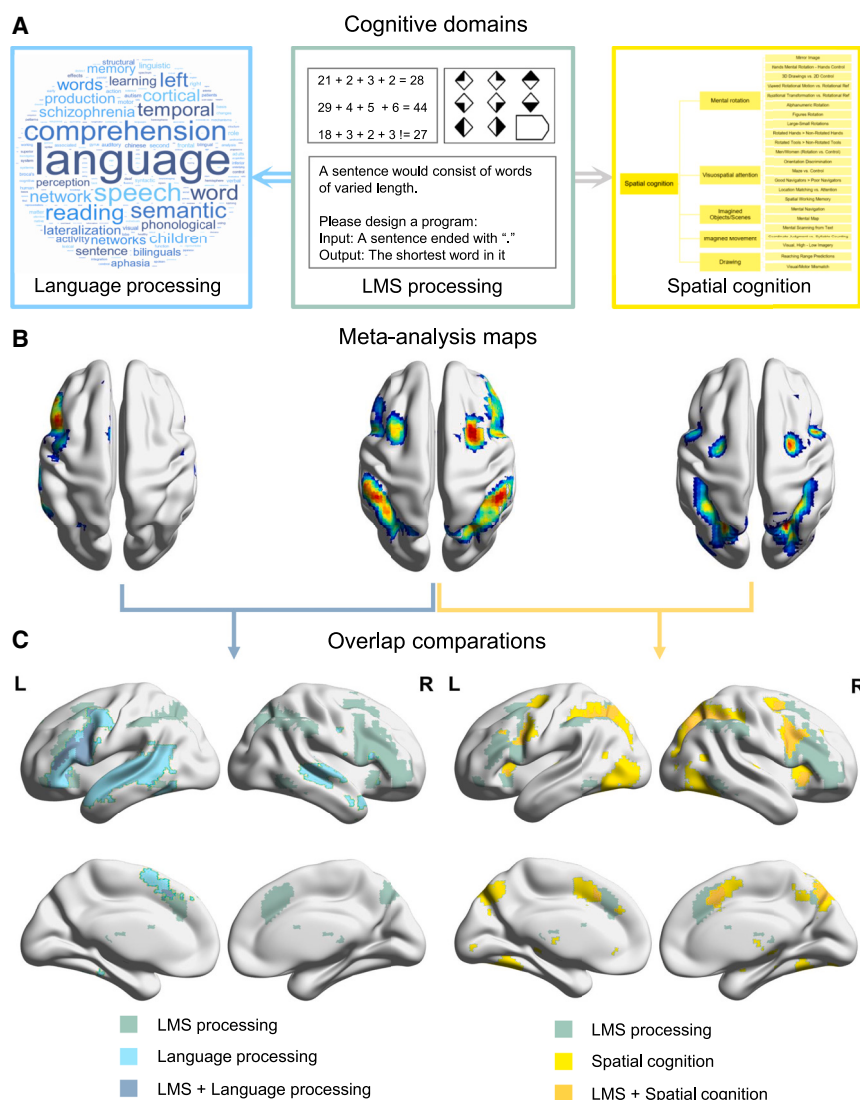
Figure 1. The comparison between neural correlates of LMS processing, language processing, and spatial cognition

(A) Alternatively, each panel or group of panels can be described separately. The generation of the domain-level neural correlate maps. LMS processing (middle): Image-based meta-analysis by representative tasks, i.e., reasoning, calculation, and mental programming. Language processing (left): automated meta-analysis by Neurosynth based on large-scale text statistics. The word cloud showed the composition of the words from the titles of the included papers, excluding neuroimaging terminology such as "fMRI". Spatial cognition (right): automated meta-analysis by Brainmap based on the established categorization. The tree graph displayed the experimental paradigms and contrasts included.

(B) The dorsal view of the resultant domain-level maps. The warmer color indicated a stronger likelihood of activation in the meta-analysis. All the maps were FDR corrected at $p = 0.01$.

(C) The overlap between the domain-level maps of LMS processing and either language processing or spatial cognition.

extent, between LMS processing and language processing. In the neural activation of LMS processing, the voxels overlapping with spatial cognition ($n = 1694$) amount to 2.7 times the number of voxels overlapping with language processing ($n = 634$). This difference is also reflected in the dice coefficient, where the coefficient between spatial cognition and LMS processing (0.255) was larger than that between language processing and LMS processing (0.113), suggesting a greater similarity in the univariate activation maps of spatial cognition and LMS processing. We replicated this pattern with the meta-analysis maps of the three representative tasks constructed from online databases (Figure S4).

To fully represent the within-domain variations in LMS, we conducted a thorough search in the 400 topic-based meta-analyses from Neurosynth's database-v5(2018) for topics where any of the top 5 keywords were related to the basic LMS abilities identified in the literature, namely number representation, calculation, reasoning, problem-solving, rule abstraction, and categorization.[5–7] Five topics were identified (Figure 2; Table S5). Of them, four consistently demonstrated greater overlap with spatial processing than with language processing tasks, and further inspection of their associated terms and papers suggested that these topics pertain to number representation, calculation, logical reasoning, problem-solving, categorization, and rule abstraction. The only exception is topic 331, which was related to sign-language processing.

Furthermore, for an in-depth examination of language-LMS association, we decomposed language processing into

spatial cognition recruits the bilateral superior frontal gyrus, the opercular part of the inferior frontal gyrus, the supplementary motion area, the insula, the superior parietal lobule spanning to the middle occipital gyrus, the precuneus, and the inferior occipital gyrus spanning to the fusiform (Figure S3; Table S3). These activation patterns are consistent with findings from previous reviews.[16,17]

Figure 1C shows the activation overlap among LMS processing, language processing, and spatial cognition. For a full list of the overlapping clusters, refer to Table S4. Visual inspection shows that the activation map of LMS processing overlapped with that of language processing mainly in the left inferior frontal gyrus and the left supplementary motion area. In contrast, overlap with spatial cognition is mainly in the bilateral posterior parietal cortex, the inferior and the middle frontal gyri, the right supplementary motion area, the right insula, and the right precuneus. Critically, there is substantial co-activation between LMS processing and spatial cognition, and to a somewhat lesser
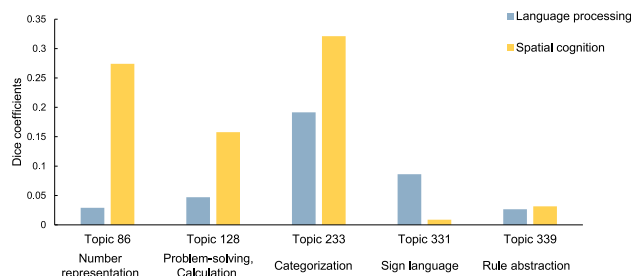
Figure 2. Dice coefficients on the meta-analysis maps of the basic abilities in LMS processing

The meta-analysis maps for LMS basic abilities were derived from topic-based automated meta-analysis maps from the Neurosynth-v5 database. The dice coefficients were calculated with the same domain-level maps of language processing and spatial cognition shown in Figure 1.

sub-domains, and compared them separately with LMS. Specifically, in the 400 topic-based meta-analyses based on Neurosynth database-v5 (2018), we identified nine topics related to language processing (Figure 3; Table S6) by searching for those where any of the top 5 keywords were related to major aspects of language processing identified in the literature.[16,62,63] We compared these meta-analysis maps with the LMS meta-map, and found that their corresponding overlap with LMS was consistently lower than the overlap between LMS and spatial cognition. Intriguingly, we observed a pattern in which the neural correlates of higher-level language processing (e.g., topic 82, sentence and discourse comprehension, and topics 274 and 201, language production) show even less overlap with LMS than lower-level language processing (e.g., topic 23, orthography).

To replicate these domain-level findings and statistically compare the resemblance of LMS processing to either language processing or spatial cognition, we calculated the subject-wise overlap between each subject's LMS synthesized map and their fMRI activation in representative language and spatial tasks, sentence processing, and mental folding, respectively (see methods 4.2.2) (Figure 4A, left and right). All the subject-wise maps were thresholded by a voxel-level threshold at $p = 0.001$. Consistent with our observations from the domain-level analysis, the dice coefficients between spatial cognition and LMS processing were significantly higher than those for language pro-

cessing and LMS processing ($t(19) = 8.825$, $p < 0.001$, *Cohen's d* = 2.025) (Figure 4B, left). This result illustrates the greater similarity in neural activation of LMS processing and spatial cognition compared to that between LMS processing and language processing.

Given that our synthesized LMS activation map was derived from three distinct representative LMS tasks, we further asked whether our findings were consistent across each task. To this end, we compared the group-level activation map of each LMS task (corrected by FDR at $p = 0.01$) with the domain-level language and spatial processing maps, and further statistically tested the difference between the task-wise LMS activations and those of sentence processing and mental folding.

In the reasoning task, overlap with spatial cognition was mainly in the bilateral posterior parietal cortex, the inferior frontal gyrus, the right middle frontal gyrus, the right insula, and the left precuneus, while overlap with language processing was mainly in the left inferior and middle frontal gyri, and the left supplementary motion area (Figure S5). The overlap of reasoning with spatial cognition doubled the size of its overlap with language processing. Consistently, the dice coefficient for reasoning with spatial cognition (0.280) was larger than that with language processing (0.170). This pattern was confirmed by a significant difference between corresponding subjective dice coefficients ($t(19) = 5.770$, $p < 0.001$, *Cohen's d* = 1.324) (Figure 4B, right).

Similar patterns were observed for calculation and mental programming tasks. In the calculation task, overlap with spatial cognition was mainly in the bilateral inferior parietal lobule, the supplementary motion area, the left middle frontal gyrus, the opercular part of the right inferior frontal gyrus, and the right insula, while that with language processing was relatively limited, predominantly in the opercular part of the left inferior frontal gyrus (Figure S6). Overlap between calculation and spatial cognition was 1.8 times larger than its overlap with language processing, consistent with the comparison of dice coefficients (with spatial cognition: 0.149 versus with language processing: 0.053), indicating a more substantial resemblance between the neural correlates of calculation and spatial cognition. This pattern was confirmed by a significant difference between corresponding subjective dice coefficients ($t(19) = 3.477$, $p < 0.01$, *Cohen's d* = 0.798) (Figure 4B, right).

In the mental programming task, overlap with spatial cognition was mainly in the left inferior parietal lobule and the left
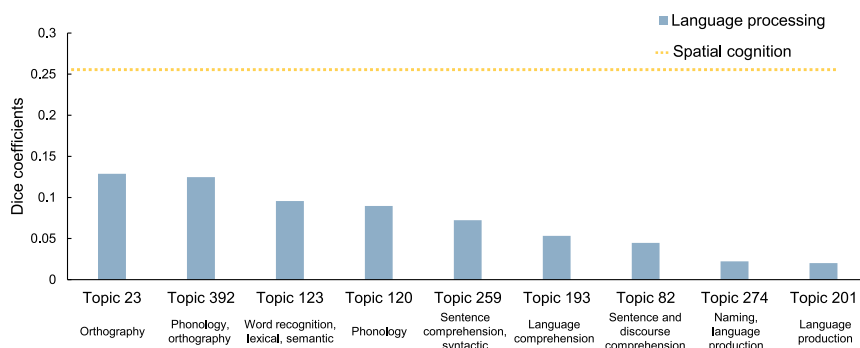
Figure 3. Dice coefficients between language-related topic-based automated meta-analysis maps from the Neurosynth-v5 database and the LMS domain-level map shown in Figure 1

The topics were sorted by the dice coefficients. The orange dashed line indicated the dice coefficients between the domain-level maps of LMS and spatial cognition shown in Figure 1. Visual inspection suggests a trend that from low-level to high-level language processing, the neural-correlates overlap with LMS seemingly decreases.

A   Task paradigms

B   Within-participant comparisons
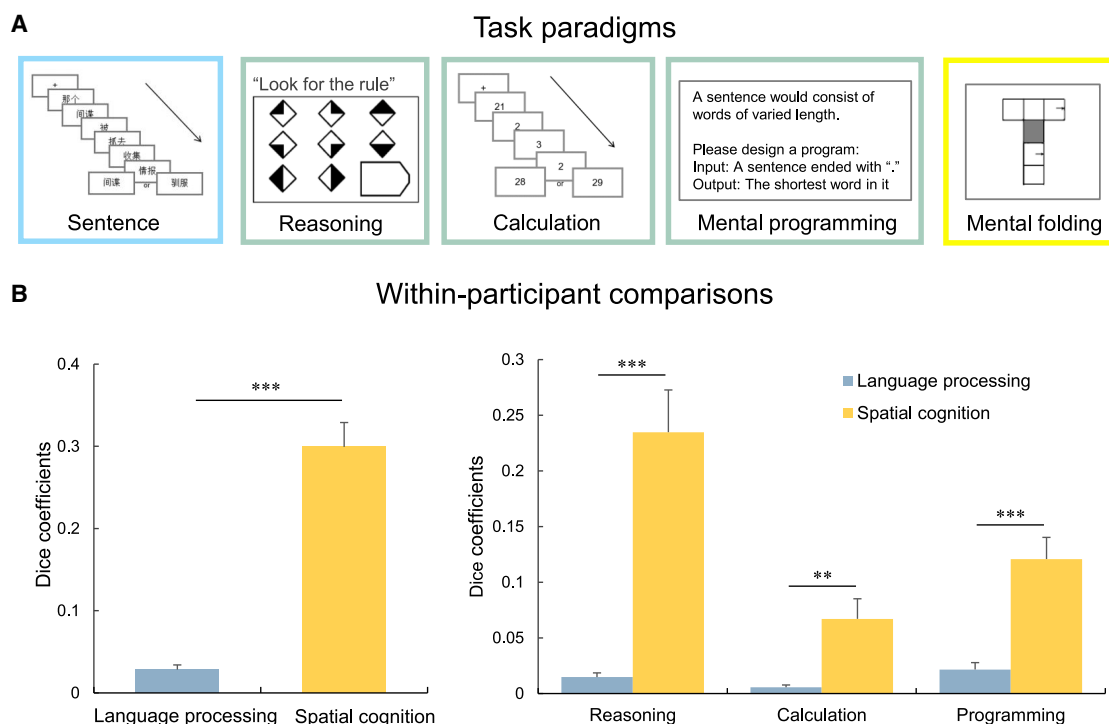
Figure 4. Task-level comparison between the univariate activation of LMS processing, language processing, and spatial cognition

(A) The representative tasks for each domain. The LMS processing tasks (middle): reasoning, calculation, and mental programming; the language processing task (left): sentence processing; the spatial cognition task (right): mental folding.

(B) The subject-wise LMS-spatial and LMS-language dice coefficients (left) and the corresponding dice coefficients after breaking LMS into separate representative tasks (right). All the subject-level activation maps were voxel-level thresholded at $p = 0.001$. The error bar denotes the standard error (SE). The asterisks denote the significance of the difference between the dice coefficients. Paired two-tailed t-tests, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

precuneus, while that with language processing was mainly in the left precentral gyrus and the triangular part of the left inferior frontal gyrus (Figure S7). The overlap between mental programming and spatial cognition is more than 10 times the size of the overlap between mental programming and language processing, consistent with the larger programming-spatial than programming-language dice coefficients at both the group level (with spatial cognition: 0.065 versus language processing: 0.044) and subject level ($t(19) = 5.215$, $p < 0.001$, *Cohen's d* = 1.196) (Figure 4B, right), suggesting that the neural activation in mental programming was more similar to spatial cognition than language processing.

In summary, we found substantially greater overlap in activation maps of LMS processing with spatial processing, especially in the posterior parietal cortex and inferior frontal gyrus, supporting the spatial origin hypothesis.

Similarity in activation patterns between LMS processing and spatial cognition

While the univariate analysis above revealed that LMS processing shared more similarities with spatial cognition in terms of activation maps overlap, it is important to note that a single cortical region can support distinct cognitive processes.[64–66] Particularly, the frontal and parietal cortex were known to be functionally heterogeneous and showed substantial cross-indi-

vidual variation.[31,67–71] Therefore, anatomical overlap in activation between LMS processing and spatial cognition in these regions does not necessarily imply functional similarity. To further test the hypothetical functional association between LMS and spatial processing, we examined whether the multivariate activation pattern of LMS processing more closely resembled that of spatial cognition than language processing. For an unbiased comparison, we deliberately chose cortical regions activated by both language processing and spatial cognition. This region of interest (ROI), delineated as the overlap between the corresponding meta-analysis activation maps, was left-lateralized, encompassing the precentral gyrus, the inferior frontal gyrus, the supplementary motor area, and the posterior parietal cortex, as well as the bilateral superior temporal gyrus (Figure S8; Table S7).

The similarity between domains was quantified by examining the likelihood of the multivariate activation patterns of LMS processing being classified as those of the representative spatial task rather than those of the representative language task in the ROI. To do this, we trained a support vector machine (SVM) for each subject, using it to classify the multivariate patterns of language processing, indexed by the sentence processing task, from those of spatial cognition, indexed by the mental folding task (Figure 5A) with leave-one-out cross-validation. Having established the boundary surface that divides
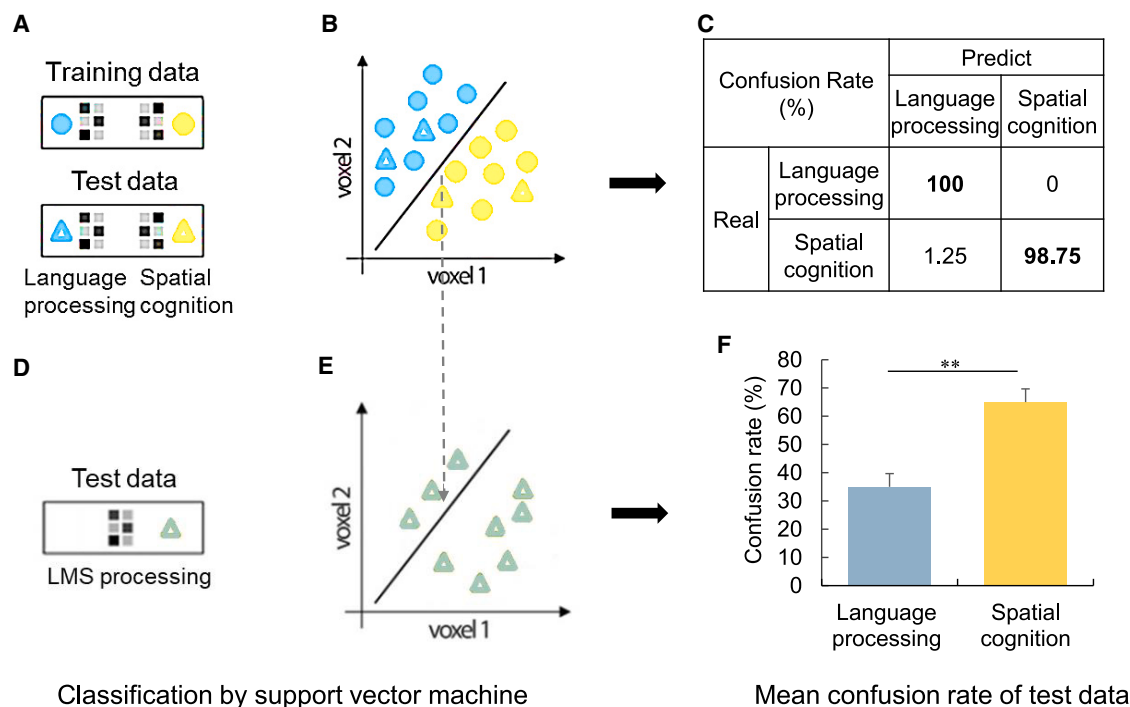
Figure 5. ROI analysis of the multivariate pattern similarity between LMS processing and spatial cognition (or language processing)
(A–C) The schematic illustration of the SVM trained to classify language processing and spatial cognition.
(A) illustrates the composition of the training (dots) and the testing (triangles) data. Blue denotes language processing and yellow spatial cognition. The gray matrices in the box are an illustration of the training and the testing data, i.e., the individual activation patterns during corresponding task runs in the ROI.
(B) illustrates the boundary surface for the two target classes (the line). The dots above the line would be classified as language processing, and the dots below the line as spatial cognition.
(C) shows the resultant confusion matrix averaged across participants, whose rows correspond to the real class of the data, and the columns represent the predicted class by the SVM.
(D–F) The schematic illustration of the classification of LMS processing activation patterns. The data of LMS processing (green triangles, D) was classified by the identical SVM (E) from B of the same subject. The dots above the line would be classified as language processing, and the dots below the line as spatial cognition.
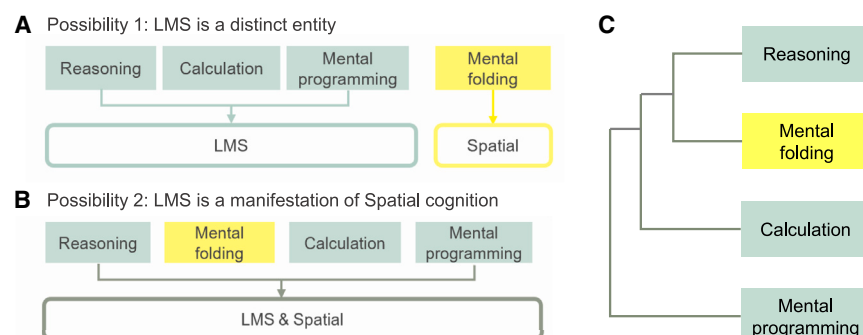(F) shows the subject-wise confusion rate of LMS processing to language processing and spatial cognition, respectively. The error bar denotes the standard error (SE). The asterisks denote the significance of the difference between the confusion rates. Paired two-tailed t-tests, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

multivariate patterns into two distinct classes, language processing and spatial cognition (Figure 5B), we then applied this classifier to the untrained multivariate patterns of LMS processing, indexed by the three representative tasks (Figure 5D). The underlying rationale is that if the ROI represented LMS in a manner more similar to spatial codes than to linguistic elements, then the classifier would more frequently classify LMS processing as spatial cognition than language processing (Figure 5E).

As expected, the spatial-linguistic SVM classifier demonstrated an accuracy close to 100% in distinguishing the multivariate patterns of language processing from those of spatial cognition across all subjects (Figure 5C), demonstrating the effectiveness of the classifier. Critically, when applying this classifier to the multivariate patterns of LMS processing in the ROI, the likelihood of LMS processing being classified as spatial cognition was 65%, significantly surpassing the 35% likelihood of being classified as language processing ($t(19)$ = 3.127, p = 0.006, Cohen's d = 0.717) (Figure 5F). That is, in regions recruited by both language processing and spatial cognition, the activation patterns elicited by LMS processing were more similar to those of spatial cognition than to language processing.

The observed anatomical and functional similarities between LMS processing and spatial cognition lead to an interesting question: Are the LMS system and the spatial cognition system distinct entities, with the latter supporting the former (Figure 6A)? Alternatively, could it be that the LMS processing is essentially one of the many manifestations of spatial cognition, rendering them indistinguishable in terms of neural correlates (Figure 6B)? To test these two hypotheses, we calculated the intra-LMS similarity among LMS tasks (i.e., reasoning, calculation, and mental programming) and the LMS-spatial similarities between each LMS task and the representative spatial cognition task (i.e., mental folding) based on their multivariate pattern in the ROI. The underlying rationale is that if the LMS processing is indeed synonymous with spatial cognition, LMS-spatial similarity should fall within the range of intra-LMS similarity. We also calculated LMS-language similarity between each LMS task and the representative language task (i.e., sentence processing) in the ROI, and expected it to be lower than both intra-LMS and LMS-spatial similarity.

As expected, the LMS-language similarity is significantly lower than both the intra-LMS similarity ($t(59)$ = 6.070, $p < 0.001$,

A Possibility 1: LMS is a distinct entity

B Possibility 2: LMS is a manifestation of Spatial cognition

C

Figure 6. The multivariate pattern similarity between individual LMS processing tasks and spatial cognition

(A) Possibility 1: the LMS processing and the spatial cognition system are distinct entities.

(B) Possibility 2: the various LMS tasks and spatial tasks are essentially manifestations of the same system.

(C) The hierarchical cluster of the multivariate activation patterns in the ROI. The longer horizontal line illustrates a higher distance and later inclusion in the cluster.

Cohen's d = 0.790) and the LMS-spatial similarity (*t(59)* = 7.255, *p* < 0.001, *Cohen's d* = 0.945). Critically, the intra-LMS similarity did not significantly exceed the LMS-spatial similarity (one-tail *t(59)* = −1.911, *p* = 0.061, *Cohen's d* = −0.249), challenging the hypothesis that LMS processing and the spatial cognition are distinct entities. This point is further supported by the hierarchical clustering analysis on the multivariate activation patterns of the LMS and spatial cognition tasks. As shown in Figure 6C, the LMS tasks did not form a distinct cluster separate from the spatial cognition task, revealing an intermingled relation among these tasks. This finding supports the idea that LMS processing likely operates via the same underlying mechanisms as spatial cognition.

DISCUSSION

The ability to process LMSs underpins everyday human activities such as calculation, reasoning, and programming. Based on the neuronal recycling hypothesis,[11] we investigated the functional origin of this processing by comparing the neural correlates of LMS processing with those of language processing and spatial cognition, respectively. The comparison of spatial extents in meta-analysis maps revealed a more substantial overlap between LMS processing and spatial cognition than that between LMS and language processing. Furthermore, a pattern similarity analysis revealed that the multivariate activation patterns of LMS processing were more similar to that of spatial cognition than language processing in regions activated by both. Taken together, our findings support the hypothesis that the ability of LMS processing is likely built upon spatial cognition rather than language processing.

The present study presented a direct comparison of the cross-task neural correlates of LMS processing with its two hypothetical origins. The present study differed from previous studies on neural commonality between LMS and spatial processing tasks[37–39,43,49] in overcoming task idiosyncrasy by meta-analysis and directly comparing competing hypotheses to address the theoretical question. First, by utilizing meta-analysis we overcome the impact of task idiosyncrasy which limited domain-level conclusions via task-wise comparison to mitigate the confounding factors. LMS tasks are integrative and idiosyncratic, given the large range of cognitive challenges they target. For instance, our previous study[37] observed dissociable components within

the cortical regions activated during mental programming, a representative LMS task, each of which showed a varied combination of neural resemblance with another three representative LMS tasks. Also, their resemblance with that of mental folding varied. This highlighted the substantial task idiosyncrasy in LMS processing, which leads to difficulty in evaluating LMS's neural resemblances with other domains, let alone comparing its relative resemblance with multiple other domains. Out of this consideration, the present study compared the cross-task neural correlates of the domains in question via either meta-analysis or synthesized activation map to mitigate the confounding factor of the task-specific properties in identifying the neural correlates.

Specifically, the cross-task map of LMS processing was generated by synthesizing in-house fMRI data of representative LMS tasks with image-based meta-analysis.[72,73] The resulting map revealed bilateral activation, albeit slightly right-lateralized, in the frontal lobe (the middle frontal gyrus, the inferior frontal gyrus), the supplementary motion area spanning to the cingulum, the insula, the posterior parietal cortex, the precuneus, and the inferior temporal gyrus. This map covers most major regions previously reported in LMS tasks, including findings from meta-analyses in mathematics,[43,55,74] reasoning,[36,75] problem-solving,[76,77] and programming,[27] among other representative LMS tasks.[29,30,32,33,49,78] By comparing this map with the meta-analysis map of language and spatial processing, we observed that the overlap between LMS processing and spatial cognition mainly resided in a bilaterally distributed set of frontoparietal regions, which are implicated in tasks such as arithmetic calculation, formal logical inference, mental programming, mental rotation, mental folding, and spatial working memory.[31,32,37,38,55] It appears that these regions might be the sites of the reuse of mechanisms for processing objects and their spatial relations. For instance, the frontoparietal dorsal attention network[79,80] may contribute to maintaining[81,82] and representing[83,84] both spatial codes and LMS, the right insula may be involved in decision-making and error monitoring in both domains,[27,85] and the inferior parietal lobule may be involved in retaining[86] and reactivating[87] ordered information in both spatial and non-spatial domains, especially in processing programming algorithms or abstract rule sets.[78,88–90] In contrast, the overlap in neural activation between LMS and language processing was significantly smaller, mainly localized in

the left inferior frontal gyrus, including Broca's area (BA45 and BA44). The involvement of these regions in LMS processing may result from task characteristics, such as the text-based visual representation of stimuli presented in LMS tasks.[27] In addition, while traditionally regarded as language-specific regions, recent studies have shown that the BA44 also engages in non-verbal working memory,[91–94] and the BA45 in retrieval and selection of non-linguistic information.[95,96] The finding of limited LMS-language overlap is consistent with observations that LMS systems emphasize repetition and symmetry, contrasting natural languages that tend to avoid repetition and rely on anti-symmetry. That is, despite both LMS and language utilizing discrete symbol systems, they differ fundamentally in their structural principles.[20] Taken together, the observed neural similarities in this study are in line with the idea that LMS processing likely originates from spatial cognition, possibly through a process of neuronal recycling.

A second distinct feature of the present study is that we achieved a direct comparison and placed the two hypothetical origins in a fair playground by focusing on the regions where both spatial and language processing activate in analyzing multivariate neural resemblance. The results further supported the spatial-origin hypothesis by showing that all tested LMS tasks were carried out in a way more like spatial than language processing when both are plausible. Note that for inferential statistics we leveraged the fMRI data from the same sample of participants in a representative language (sentence processing) task and a spatial (mental folding) task, and calculated the subject-level pairwise similarity among tasks to statistically compare intra-LMS similarity with LMS-spatial and LMS-language similarity. This reanalysis of previously reported data[37] was however distinct from previous analyses in the research question, the methodology, the anatomical scope, and the range of tasks included. The previous analysis was conducted within the cortical regions activated by mental programming and broke up the regions in question via clustering analysis based on neural resemblance to a set of non-verbal tasks, while the present study conducted ROI-wise multivariate pattern similarity analysis to compare the relative resemblance of LMS tasks with language and spatial tasks, which could not be predicted by the findings of the previous study.

Though our analysis centered on LMS processing, we indeed speculate that spatial cognition might be a key cognitive cornerstone extending far beyond LMS processing, potentially influencing a wide array of cognitive abilities. Previous studies revealed the involvement of cortical regions associated with spatial processing in various non-spatial domains.[97–101] In some cases, this involvement is attributable to the use of spatial information in the task, as in the case of spatial working memory. In some other cases, as we hypothesized for LMS, such involvement may reflect a functional evolution of spatial processing extending into the non-spatial domain, as in many LMS tasks. Theories like the mental model[41] and mental simulation[42,43] suggest that relation processing typically utilizes a visuospatial workspace or simulations to navigate representational spaces, and map-like representations are proposed to be a universal coding mechanism that organizes spatial, non-spatial, and even abstract information, using shared neural substrates.[98,102,103] For example, the inferior parietal cortex, also identified in this study, is known to encode non-spatial abstract information[78,88–90,97,104] in addition to spatial information. This suggests a shared neural basis for spatial and non-spatial information mapping. Indeed, the spatial navigation system has been proposed to provide critical computational components for the neural implementation of the "language of thought" (LoT).[13,105] LoT, advanced by Jerry Fodor,[106] describes not an actual language but a formal symbol system that stores and manipulates concepts symbolically, adhering to compositional rules. This system enables the construction of complex thoughts from primitive symbolic concepts and thus gives rise to the productivity and creativity of the human mind, advocating LoT as a candidate for human singularity in cognition.[20,107]

However, the proposal that spatial cognition serves as a cross-domain computational foundation in the human mind leads to a puzzle: why does spatial cognition, an ability shared with many species, generate sophisticated non-spatial abilities in humans, such as formal operations of LMS systems? One possible answer is that spatial cognition may not be the exclusive origin of LMS processing. Language processing, although not the primary source, may also contribute to LMS processing.[27,36,108] For example, language processing might be essential in the initial stage of reasoning tasks, in which verbally presented information is encoded as non-verbal representations for later mental operations that are dissociated with natural language,[109] and language capacity might be one of the key environmental and educational factors affecting reasoning competency and could affect one's subjective bias in manipulating symbols[110] and fluency in using symbols,[111] particularly in using linguistic symbols for calculation in second language context.[112] Additionally, language processing is proposed to influence the early development of computational thinking by supporting the verbalization of algorithmic processes as a cognitive scaffolding strategy. Consistent with this idea, we observed a pattern in which the neural correlates of higher-level language processing show even less overlap with LMS than lower-level language processing. Also, we observed a larger overlap with language than with spatial cognition in some aspects of LMS, e.g., reasoning with sign language (see Figure 2). These findings suggest an intricate and task-sensitive involvement of language processing in LMS. For a specific task, the relative resemblance to language versus spatial cognition may be influenced by a variety of factors, including low-level factors such as task format and experimental administration, as well as higher-level factors such as the cognitive nature of the task itself. For instance, reasoning tasks that involve semantic meaning and code comprehension or tasks that rely heavily on linguistic symbols may show substantial activation in cortical regions associated with language processing due to the extensive involvement of text-based materials, even though they primarily engage reasoning and problem-solving processes. Future investigation is needed to further characterize the intricate involvement of language in LMS processing, which might advance the understanding of how language functions as a cognitive scaffolding supporting LMS and other high-level cognitive functions in general.

Limitations of the study

Our study relies predominantly on neural overlaps in activation maps. This can result in an oversight of functional similarities between different anatomical regions. According to the multiple parallel circuit hypothesis,[20] anatomically separate regions may instead undertake similar functionalities in different symbol systems, which were all based on the recursive composition of corresponding primitives to encode nested repetitions with variations. Thus, future studies should explore the potential computational commonalities between language and LMS processing to further understand these intricate cognitive interactions.

It should be noted that our in-house fMRI study recruited only participants with programming experience. There has been report of expertise-dependent neural correlates of programming.[29] Compared to non-programmers, programmers exhibit more efficient neural processing when solving easy tasks, indicated by more restricted cortical recruitment in non-spatial regions.[48] Consequently, the exclusive recruitment of experienced participants might lead to an underestimation of the overlap between programming and language processing. However, our follow-up analysis, using the meta-analysis map of problem-solving from Neurosynth (Figure S4), replicated the finding of a larger overlap with spatial cognition than with the language network, supporting our original findings. However, we do agree that individuals with diverse life histories and educational backgrounds may approach LMS tasks with different cognitive strategies, potentially varying in their reliance on spatial and language networks. Future studies focusing on the cultural and experiential sensitivity of language-vs-spatial reliance may advance our understanding of the flexible manifestations of human intelligence. Also, our in-house fMRI study did not examine the influence of sex or age, and all participants identified as Han Chinese. While the observed activation pattern similarities were consistent with those from large-scale meta-analyses, the generalizability of these findings to more diverse populations may be limited. Future studies exploring the effects of sex, age, and culture may further advance understanding of the potential influence of these factors on the neural correlates of LMS processing.

Another limitation of our study is its reliance on correlational evidence. Future research is needed to establish a causal link between LMS processing and spatial cognition. For instance, neural manipulation techniques such as transcranial magnetic stimulation could be used to test the causal relationship between spatial processing and LMS. An intriguing direction of mechanistic investigation is to leverage the recent advance in artificial neural network. Specifically, incorporating a spatial module and 're-running' the evolution process could provide insights into how spatial cognition scaffolds the emergence of LMS processing. Supporting this intuition, state-of-the-art LLMs, proficient in natural language tasks, show brittle performance in LMS tasks.[113,114] Wong et al.[115] suggested that this might be because LLMs work upon the context-aware mapping between language and meanings, lacking a world model based on probabilistic LoT. Based on the link between LMS processing and spatial cognition observed in this study, it could be worth investigating the enhancement of LMS processing in LLMs through training on spatial cognition tasks, such as path integrating, ori-

enting, and map reading. Indeed, recent studies have demonstrated training navigation in abstract concept space using cognitive maps could improve the reasoning and problem-solving ability of LLMs.[116,117] Extending such investigation to training spatial cognition and endowing LLMs with spatial abilities might bridge the gap between disembodied information in LLM training and embodied reality that artificial general intelligence (AGI) ultimately aims to interact with, which could significantly contribute to the development of next-generation AGI.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Shan Xu (shan.xu@bnu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Original in-house fMRI images and data for analysis have been deposited at Mendeley data: https://data.mendeley.com/preview/bkn4xx4kc3?a=9161e868-1e4e-4819-a9da-5e77c12abff2 and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- Code can be obtained at Mendeley data: https://data.mendeley.com/preview/bkn4xx4kc3?a=9161e868-1e4e-4819-a9da-5e77c12abff2 and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- The online database for large-scale automated meta-analyses including Neurosynth database (version 0.5 released on February 23, 2015): https://github.com/neurosynth/neurosynth-data/releases/tag/0.5, Neurosynth database-v5 (the database as of July 2018): https://www.neurosynth.org/analyses/topics/v5-topics-400/ and BrainMap database (accessed in August 19, 2023): https://brainmap.org/index.html, are publicly available through the accession links listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

AUTHOR CONTRIBUTIONS

Conceptualization, J.L., S.X., and Y.L.; methodology, S.X. and Y.L.; Investigation, S.X. and Y.L.; writing–original draft, S.X. and Y.L.; writing—review and editing, J.L., S.X., and Y.L.; funding acquisition, J.L. and S.X.; resources, J.L., S.X., and Y.L.; supervision, J.L. and S.X.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE

- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
- METHOD DETAILS
  - Meta-analyses of language processing and spatial cognition
  - FMRI studies
  - Language processing and spatial cognition
  - Dice coefficients of meta-analysis maps and subject-level maps
  - Multivariate pattern analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

REFERENCES

1. Gardner, H.E. (1983). Frames of Mind: The Theory of Multiple Intelligences (Basic books).
2. Davis, K., Christodoulou, J., Seider, S., and Gardner, H.E. (2011). The Theory of Multiple Intelligences. In The Cambridge handbook of intelligence, R.J. Sternberg and S.B. Kaufman, eds. (Cambridge University Press). https://doi.org/10.1017/CBO9780511977244.025.
3. Milsan, A.L., and Wewe, M. (2019). Hubungan antara kecerdasan logis matematis dengan hasil belajar matematika. J. Educ. Technol. 2, 65–99.
4. Tajuddin, A.T., Sujadi, I., Slamet, I., and Hendriyanto, A. (2023). Mathematical Critical Thinking: Analysis of Middle School Students' Thinking Processes in Solving Trigonometry Problems. Mosharafa J. Pendidik. Mat. 12, 703–720.
5. Arum, D.P., Kusmayadi, T.A., and Pramudya, I. (2018). Students' logical-mathematical intelligence profile. J. Phys. Conf. Ser. 1008, 012071. https://doi.org/10.1088/1742-6596/1008/1/012071.
6. Nurlaelah, E., Usdiyana, D., and Fadilah, N. (2024). The Relationship Between Computational Thinking Ability and Logical Mathematical Intelligence. Mosharafa J. Pendidik. Mat. 13, 87–96. https://doi.org/10.31980/mosharafa.v13i1.1978.
7. Maftoon, P., and Sarem, S.N. (2012). The realization of Gardner's multiple intelligences (MI) theory in second language acquisition (SLA). J. Lang. Teach. Res. 3, 1233. https://doi.org/10.4304/jltr.3.6.1233-1241.
8. Shearer, C.B., and Karanian, J.M. (2017). The neuroscience of intelligence: Empirical support for the theory of multiple intelligences? Trends Neurosci. Educ. 6, 211–223. https://doi.org/10.1016/j.tine.2017.02.002.
9. Pande, N.A. (2010). Numeral systems of great ancient human civilizations. J. Sci. Arts 2, 209.
10. Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The "neuronal recycling" hypothesis. In From Monkey Brain to Human Brain, S. Dehaene, J.-R. Duhamel, M.D. Hauser, and G. Rizzolatti, eds. (MIT Press), pp. 133–157. https://doi.org/10.7551/mitpress/3136.003.0012.
11. Dehaene, S., and Cohen, L. (2007). Cultural Recycling of Cortical Maps. Neuron 56, 384–398. https://doi.org/10.1016/j.neuron.2007.10.004.
12. Fedorenko, E., Ivanova, A., Dhamala, R., and Bers, M.U. (2019). The Language of Programming: A Cognitive Perspective. Trends Cognit. Sci. 23, 525–528. https://doi.org/10.1016/j.tics.2019.04.010.
13. Kazanina, N., and Poeppel, D. (2023). The neural ingredients for a language of thought are available. Trends Cognit. Sci. 27, 996–1007. https://doi.org/10.1016/j.tics.2023.07.012.
14. Baddeley, A. (1992). Working Memory. Science 255, 556–559. https://doi.org/10.1126/science.1736359.
15. Lipkin, B., Tuckute, G., Affourtit, J., Small, H., Mineroff, Z., Kean, H., Jouravlev, O., Rakocevic, L., Pritchett, B., Siegelman, M., et al. (2022). Probabilistic atlas for the language network based on precision fMRI data from >800 individuals. Sci. Data 9, 529. https://doi.org/10.1038/s41597-022-01645-3.
16. Price, C.J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62, 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062.
17. Cona, G., and Scarpazza, C. (2019). Where is the "where" in the brain? A meta-analysis of neuroimaging studies on spatial cognition. Hum. Brain Mapp. 40, 1867–1886. https://doi.org/10.1002/hbm.24496.
18. Campo, P., Maestú, F., Ortiz, T., Capilla, A., Santiuste, M., Fernández, A., and Amo, C. (2005). Time Modulated Prefrontal and Parietal Activity during the Maintenance of Integrated Information as Revealed by Magnetoencephalography. Cerebr. Cortex 15, 123–130. https://doi.org/10.1093/cercor/bhh115.
19. Wu, X., Chen, X., Li, Z., Han, S., and Zhang, D. (2007). Binding of verbal and spatial information in human working memory involves large-scale neural synchronization at theta frequency. Neuroimage 35, 1654–1662. https://doi.org/10.1016/j.neuroimage.2007.02.011.
20. Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., and Sablé-Meyer, M. (2022). Symbols and mental programs: a hypothesis about human singularity. Trends Cognit. Sci. 26, 751–766. https://doi.org/10.1016/j.tics.2022.06.010.
21. Krumnack, A., Bucher, L., Nejasmic, J., Nebel, B., and Knauff, M. (2011). A model for relational reasoning as verbal reasoning. Cogn. Syst. Res. 12, 377–392. https://doi.org/10.1016/j.cogsys.2010.11.001.
22. Polk, T.A., and Newell, A. (1995). Deduction as verbal reasoning. Psychol. Rev. 102, 533–566.
23. Fedorenko, E., and Varley, R. (2016). Language and thought are not the same thing: evidence from neuroimaging and neurological patients. Ann. N. Y. Acad. Sci. 1369, 132–153. https://doi.org/10.1111/nyas.13046.
24. LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B.L., Bisanz, J., Kamawar, D., and Penner-Wilger, M. (2010). Pathways to Mathematics: Longitudinal Predictors of Performance. Child Dev. 81, 1753–1767. https://doi.org/10.1111/j.1467-8624.2010.01508.x.
25. Prat, C.S., Madhyastha, T.M., Mottarella, M.J., and Kuo, C.-H. (2020). Relating Natural Language Aptitude to Individual Differences in Learning Programming Languages. Sci. Rep. 10, 3817. https://doi.org/10.1038/s41598-020-60661-8.
26. Amalric, M., and Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. Proc. Natl. Acad. Sci. USA 113, 4909–4917. https://doi.org/10.1073/pnas.1603205113.
27. Castelhano, J., Duarte, I.C., Duraes, J., Madeira, H., and Castelo-Branco, M. (2021). Reading and Calculation Neural Systems and Their Weighted Adaptive Use for Programming Skills. Neural Plast. 2021, e5596145. https://doi.org/10.1155/2021/5596145.
28. Coetzee, J.P., Johnson, M.A., Lee, Y., Wu, A.D., Iacoboni, M., and Monti, M.M. (2022). Dissociating Language and Thought in Human Reasoning. Brain Sci. 13, 67. https://doi.org/10.3390/brainsci13010067.
29. Floyd, B., Santander, T., and Weimer, W. (2017). Decoding the Representation of Code in the Brain: An fMRI Study of Code Review and Expertise. In 2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE), S. Uchitel, A. Orso, and M. Robillard, eds. (IEEE), pp. 175–186. https://doi.org/10.1109/ICSE.2017.24.
30. Houdé, O., and Tzourio-Mazoyer, N. (2003). Neural foundations of logical and mathematical cognition. Nat. Rev. Neurosci. 4, 507–514. https://doi.org/10.1038/nrn1117.
31. Ivanova, A.A., Srikant, S., Sueoka, Y., Kean, H.H., Dhamala, R., O'Reilly, U.-M., Bers, M.U., and Fedorenko, E. (2020). Comprehension of computer code relies primarily on domain-general executive brain regions. Elife 9, e58906. https://doi.org/10.7554/eLife.58906.

32. Liu, Y.-F., Kim, J., Wilson, C., and Bedny, M. (2020). Computer code comprehension shares neural resources with formal logical inference in the fronto-parietal network. Elife 9, e59340. https://doi.org/10.7554/eLife.59340.

33. Maruyama, M., Pallier, C., Jobert, A., Sigman, M., and Dehaene, S. (2012). The cortical representation of simple mathematical expressions. Neuroimage 61, 1444–1460. https://doi.org/10.1016/j.neuroimage.2012.04.020.

34. Monti, M.M., Parsons, L.M., and Osherson, D.N. (2009). The boundaries of language and thought in deductive inference. Proc. Natl. Acad. Sci. USA 106, 12554–12559. https://doi.org/10.1073/pnas.0902422106.

35. Monti, M.M., Parsons, L.M., and Osherson, D.N. (2012). Thought Beyond Language: Neural Dissociation of Algebra and Natural Language. Psychol. Sci. 23, 914–922. https://doi.org/10.1177/0956797612437427.

36. Wang, L., Zhang, M., Zou, F., Wu, X., and Wang, Y. (2020). Deductive-reasoning brain networks: A coordinate-based meta-analysis of the neural signatures in deductive reasoning. Brain Behav. 10, e01853. https://doi.org/10.1002/brb3.1853.

37. Xu, S., Li, Y., and Liu, J. (2021). The Neural Correlates of Computational Thinking: Collaboration of Distinct Cognitive Components Revealed by fMRI. Cerebr. Cortex 31, 5579–5597. https://doi.org/10.1093/cercor/bhab182.

38. Zago, L., Petit, L., Turbelin, M.-R., Andersson, F., Vigneau, M., and Tzourio-Mazoyer, N. (2008). How verbal and spatial manipulation networks contribute to calculation: An fMRI study. Neuropsychologia 46, 2403–2414. https://doi.org/10.1016/j.neuropsychologia.2008.03.001.

39. Endres, M., Karas, Z., Hu, X., Kovelman, I., and Weimer, W. (2021). Relating Reading, Visualization, and Coding for New Programmers: A Neuroimaging Study. In 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), N. Juristo, A.v. Deursen, and T. Xie, eds. (IEEE), pp. 600–612. https://doi.org/10.1109/ICSE43902.2021.00062.

40. Lin, Y., and Dillon, M.R. (2024). We Are Wanderers: Abstract geometry reflects spatial navigation. J. Exp. Psychol. Gen. 153, 386–398. https://doi.org/10.1037/xge0001504.

41. Johnson-Laird, P.N. (2001). Mental models and deduction. Trends Cognit. Sci. 5, 434–442.

42. Barsalou, L.W. (2008). Grounded Cognition. Annu. Rev. Psychol. 59, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

43. Hawes, Z., and Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. Psychon. Bull. Rev. 27, 465–482.

44. Wei, W., Yuan, H., Chen, C., and Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. Br. J. Educ. Psychol. 82, 157–181. https://doi.org/10.1111/j.2044-8279.2011.02049.x.

45. Xie, F., Zhang, L., Chen, X., and Xin, Z. (2020). Is Spatial Ability Related to Mathematical Ability: a Meta-analysis. Educ. Psychol. Rev. 32, 113–155. https://doi.org/10.1007/s10648-019-09496-y.

46. Goel, V. (2007). Anatomy of deductive reasoning. Trends Cognit. Sci. 11, 435–441.

47. Knauff, M., Fangmeier, T., Ruff, C.C., and Johnson-Laird, P.N. (2003). Reasoning, models, and images: Behavioral measures and cortical activity. J. Cognit. Neurosci. 15, 559–573.

48. Helmlinger, B., Sommer, M., Feldhammer-Kahr, M., Wood, G., Arendasy, M.E., and Kober, S.E. (2020). Programming experience associated with neural efficiency during figural reasoning. Sci. Rep. 10, 13351. https://doi.org/10.1038/s41598-020-70360-z.

49. Huang, Y., Liu, X., Krueger, R., Santander, T., Hu, X., Leach, K., and Weimer, W. (2019). Distilling Neural Representations of Data Structure Manipulation using fMRI and fNIRS. In 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE), J. Atlee, T. Bultan, and J. Whittle, eds. (IEEE), pp. 396–407. https://doi.org/10.1109/ICSE.2019.00053.

50. Sharafi, Z., Huang, Y., Leach, K., and Weimer, W. (2021). Toward an Objective Measure of Developers' Cognitive Activities. ACM Trans. Software Eng. Methodol. 30, 1–40. https://doi.org/10.1145/3434643.

51. Román-González, M., Pérez-González, J.-C., and Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. Comput. Hum. Behav. 72, 678–691. https://doi.org/10.1016/j.chb.2016.08.047.

52. Wai, J., Lubinski, D., and Benbow, C.P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. J. Educ. Psychol. 101, 817–835. https://doi.org/10.1037/a0016127.

53. Khine, M.S. (2017). Spatial Cognition: Key to STEM Success. In Visual-spatial Ability in STEM Education: Transforming Research into Practice, M.S. Khine, ed. (Springer International Publishing), pp. 3–8. https://doi.org/10.1007/978-3-319-44385-0_1.

54. Shea, D.L., Lubinski, D., and Benbow, C.P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. J. Educ. Psychol. 93, 604–614. https://doi.org/10.1037/0022-0663.93.3.604.

55. Hawes, Z., Sokolowski, H.M., Ononye, C.B., and 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. Neurosci. Biobehav. Rev. 103, 316–336. https://doi.org/10.1016/j.neubiorev.2019.05.007.

56. Toomarian, E.Y., and Hubbard, E.M. (2018). On the genesis of spatial-numerical associations: Evolutionary and cultural factors co-construct the mental number line. Neurosci. Biobehav. Rev. 90, 184–199. https://doi.org/10.1016/j.neubiorev.2018.04.010.

57. Endres, M., Fansher, M., Shah, P., and Weimer, W. (2021). To read or to rotate? comparing the effects of technical reading training and spatial skills training on novice programming ability. In Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, D. Spinellis, G. Gousios, M. Chechik, and M. Di Penta, eds. (ACM), pp. 754–766. https://doi.org/10.1145/3468264.3468583.

58. Xia, M., Wang, J., and He, Y. (2013). BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. PLoS One 8, e68910. https://doi.org/10.1371/journal.pone.0068910.

59. Raven, J. (2003). Raven Progressive Matrices. In Handbook of Nonverbal Assessment, R.S. McCallum, ed. (Springer US), pp. 223–237. https://doi.org/10.1007/978-1-4615-0153-4_11.

60. Gonthier, C. (2022). Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: an integrative review and a model for matrices tasks. Cogn. Res. Princ. Implic. 7, 11. https://doi.org/10.1186/s41235-021-00350-w.

61. Zhao, S., You, H., Zhang, R.-Y., Si, B., Zhen, Z., Wan, X., and Wang, D.-H. (2023). An Interpretable Neuro-symbolic Model for Raven's Progressive Matrices Reasoning. Cogn. Comput. 15, 1703–1724. https://doi.org/10.1007/s12559-023-10154-3.

62. Price, C.J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. Ann. N. Y. Acad. Sci. 1191, 62–88. https://doi.org/10.1111/j.1749-6632.2010.05444.x.

63. Weiss-Croft, L.J., and Baldeweg, T. (2015). Maturation of language networks in children: A systematic review of 22years of functional MRI. Neuroimage 123, 269–281. https://doi.org/10.1016/j.neuroimage.2015.07.046.

64. Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cognit. Sci. 14, 172–179.

65. Woolgar, A., Thompson, R., Bor, D., and Duncan, J. (2011). Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. Neuroimage 56, 744–752. https://doi.org/10.1016/j.neuroimage.2010.04.035.

66. Woolgar, A., Jackson, J., and Duncan, J. (2016). Coding of Visual, Auditory, Rule, and Response Information in the Brain: 10 Years of Multivoxel Pattern Analysis. J. Cognit. Neurosci. *28*, 1433–1454. https://doi.org/10.1162/jocn_a_00981.

67. Blank, I.A., and Fedorenko, E. (2017). Domain-General Brain Regions Do Not Track Linguistic Input as Closely as Language-Selective Regions. J. Neurosci. *37*, 9999–10011. https://doi.org/10.1523/JNEUROSCI.3642-16.2017.

68. Braga, R.M., Van Dijk, K.R.A., Polimeni, J.R., Eldaief, M.C., and Buckner, R.L. (2019). Parallel distributed networks resolved at high resolution reveal close juxtaposition of distinct regions. J. Neurophysiol. *121*, 1513–1534. https://doi.org/10.1152/jn.00808.2018.

69. Frost, M.A., and Goebel, R. (2012). Measuring structural–functional correspondence: Spatial variability of specialised brain regions after macro-anatomical alignment. Neuroimage *59*, 1369–1381. https://doi.org/10.1016/j.neuroimage.2011.08.035.

70. Tahmasebi, A.M., Davis, M.H., Wild, C.J., Rodd, J.M., Hakyemez, H., Abolmaesumi, P., and Johnsrude, I.S. (2012). Is the Link between Anatomical Structure and Function Equally Strong at All Cognitive Levels of Processing? Cerebr. Cortex *22*, 1593–1603. https://doi.org/10.1093/cercor/bhr205.

71. Vázquez-Rodríguez, B., Suárez, L.E., Markello, R.D., Shafiei, G., Paquola, C., Hagmann, P., van den Heuvel, M.P., Bernhardt, B.C., Spreng, R.N., and Misic, B. (2019). Gradients of structure–function tethering across neocortex. Proc. Natl. Acad. Sci. USA *116*, 21219–21227. https://doi.org/10.1073/pnas.1903403116.

72. Salimi-Khorshidi, G., Smith, S.M., Keltner, J.R., Wager, T.D., and Nichols, T.E. (2009). Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. Neuroimage *45*, 810–823. https://doi.org/10.1016/j.neuroimage.2008.12.039.

73. Soares, J.M., Magalhães, R., Moreira, P.S., Sousa, A., Ganz, E., Sampaio, A., Alves, V., Marques, P., and Sousa, N. (2016). A Hitchhiker's Guide to Functional Magnetic Resonance Imaging. Front. Neurosci. *10*, 00515. https://doi.org/10.3389/fnins.2016.00515.

74. Tablante, J., Krossa, L., Azimi, T., and Chen, L. (2023). Dysfunctions associated with the intraparietal sulcus and a distributed network in individuals with math learning difficulties: An ALE meta-analysis. Hum. Brain Mapp. *44*, 2726–2740. https://doi.org/10.1002/hbm.26240.

75. Hobeika, L., Diard-Detoeuf, C., Garcin, B., Levy, R., and Volle, E. (2016). General and specialized brain correlates for analogical reasoning: A meta-analysis of functional imaging studies. Hum. Brain Mapp. *37*, 1953–1969. https://doi.org/10.1002/hbm.23149.

76. Bartley, J.E., Boeving, E.R., Riedel, M.C., Bottenhorn, K.L., Salo, T., Eickhoff, S.B., Brewe, E., Sutherland, M.T., and Laird, A.R. (2018). Meta-analytic evidence for a core problem solving network across multiple representational domains. Neurosci. Biobehav. Rev. *92*, 318–337. https://doi.org/10.1016/j.neubiorev.2018.06.009.

77. Feng, W., Wang, W., Liu, J., Wang, Z., Tian, L., and Fan, L. (2021). Neural Correlates of Causal Inferences in Discourse Understanding and Logical Problem-Solving: A Meta-Analysis Study. Front. Hum. Neurosci. *15*, 666179. https://doi.org/10.3389/fnhum.2021.666179.

78. Ikutani, Y., Kubo, T., Nishida, S., Hata, H., Matsumoto, K., Ikeda, K., and Nishimoto, S. (2021). Expert programmers have fine-tuned cortical representations of source code. Eneuro *8*, ENEURO.0405-20.2020.

79. Corbetta, M., and Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. *3*, 201–215. https://doi.org/10.1038/nrn755.

80. Kincade, J.M., Abrams, R.A., Astafiev, S.V., Shulman, G.L., and Corbetta, M. (2005). An Event-Related Functional Magnetic Resonance Imaging Study of Voluntary and Stimulus-Driven Orienting of Attention. J. Neurosci. *25*, 4593–4604. https://doi.org/10.1523/JNEUROSCI.0236-05.2005.

81. Ravizza, S.M., Delgado, M.R., Chein, J.M., Becker, J.T., and Fiez, J.A. (2004). Functional dissociations within the inferior parietal cortex in verbal working memory. Neuroimage *22*, 562–573. https://doi.org/10.1016/j.neuroimage.2004.01.039.

82. Sack, A.T. (2009). Parietal cortex and spatial cognition. Behav. Brain Res. *202*, 153–161. https://doi.org/10.1016/j.bbr.2009.03.012.

83. Hanakawa, T., Honda, M., Sawamoto, N., Okada, T., Yonekura, Y., Fukuyama, H., and Shibasaki, H. (2002). The role of rostral Brodmann area 6 in mental-operation tasks: an integrative neuroimaging approach. Cerebr. Cortex *12*, 1157–1170.

84. Pinel, P., Dehaene, S., Rivière, D., and LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. Neuroimage *14*, 1013–1026. https://doi.org/10.1006/nimg.2001.0913.

85. Uddin, L.Q. (2015). Salience processing and insular cortical function and dysfunction. Nat. Rev. Neurosci. *16*, 55–61.

86. Marshuetz, C., Smith, E.E., Jonides, J., DeGutis, J., and Chenevert, T.L. (2000). Order Information in Working Memory: fMRI Evidence for Parietal and Prefrontal Mechanisms. J. Cognit. Neurosci. *12*, 130–144. https://doi.org/10.1162/08989290051137459.

87. Corbetta, M., Kincade, J.M., and Shulman, G.L. (2002). Neural Systems for Visual Orienting and Their Relationships to Spatial Working Memory. J. Cognit. Neurosci. *14*, 508–523. https://doi.org/10.1162/089892902317362029.

88. Crittenden, B.M., Mitchell, D.J., and Duncan, J. (2016). Task Encoding across the Multiple Demand Cortex Is Consistent with a Frontoparietal and Cingulo-Opercular Dual Networks Distinction. J. Neurosci. *36*, 6147–6155. https://doi.org/10.1523/JNEUROSCI.4590-15.2016.

89. Pischedda, D., Görgen, K., Haynes, J.-D., and Reverberi, C. (2017). Neural Representations of Hierarchical Rule Sets: The Human Control System Represents Rules Irrespective of the Hierarchical Level to Which They Belong. J. Neurosci. *37*, 12281–12296. https://doi.org/10.1523/JNEUROSCI.3088-16.2017.

90. Srikant, S., Lipkin, B., Ivanova, A.A., Fedorenko, E., and O'Reilly, U.-M. (2022). Convergent representations of computer programs in human and artificial neural networks. In Proceedings of the 36th International Conference on Neural Information Processing Systems NIPS '22, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds. (Curran Associates Inc.), pp. 18834–18849. https://doi.org/10.5555/3600270.3601638.

91. Gould, R.L., Brown, R.G., Owen, A.M., ffytche, D.H., and Howard, R.J. (2003). fMRI BOLD response to increasing task difficulty during successful paired associates learning. Neuroimage *20*, 1006–1019. https://doi.org/10.1016/S1053-8119(03)00365-3.

92. Honey, G.D., Bullmore, E.T., and Sharma, T. (2000). Prolonged reaction time to a verbal working memory task predicts increased power of posterior parietal cortical activation. Neuroimage *12*, 495–503.

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

94. Maguire, E.A., Frackowiak, R.S., and Frith, C.D. (1997). Recalling routes around London: activation of the right hippocampus in taxi drivers. J. Neurosci. *17*, 7103–7110.

95. Liakakis, G., Nickel, J., and Seitz, R.J. (2011). Diversity of the inferior frontal gyrus—A meta-analysis of neuroimaging studies. Behav. Brain Res. *225*, 341–347. https://doi.org/10.1016/j.bbr.2011.06.022.

96. Thompson-Schill, S.L., D'Esposito, M., and Kan, I.P. (1999). Effects of Repetition and Competition on Activity in Left Prefrontal Cortex during Word Generation. Neuron *23*, 513–522. https://doi.org/10.1016/S0896-6273(00)80804-1.

97. Gottlieb, J., and Snyder, L.H. (2010). Spatial and non-spatial functions of the parietal cortex. Curr. Opin. Neurobiol. *20*, 731–740.

98. Igarashi, K.M., Lee, J.Y., and Jun, H. (2022). Reconciling neuronal representations of schema, abstract task structure, and categorization under

cognitive maps in the entorhinal-hippocampal-frontal circuits. Curr. Opin. Neurobiol. *77*, 102641.

99. Lumaca, M., Keller, P.E., Baggio, G., Pando-Naude, V., Bajada, C.J., Martinez, M.A., Hansen, J.H., Ravignani, A., Joe, N., Vuust, P., et al. (2024). Frontoparietal network topology as a neural marker of musical perceptual abilities. Nat. Commun. *15*, 8160. https://doi.org/10.1038/s41467-024-52479-z.

100. Marek, S., and Dosenbach, N.U.F. (2018). The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. Dialogues Clin. Neuroscience (San Diego, CA, U. S.) *20*, 133–140.

101. Ptak, R., Schnider, A., and Fellrath, J. (2017). The Dorsal Frontoparietal Network: A Core System for Emulated Action. Trends Cognit. Sci. *21*, 589–599. https://doi.org/10.1016/j.tics.2017.05.002.

102. Behrens, T.E.J., Muller, T.H., Whittington, J.C.R., Mark, S., Baram, A.B., Stachenfeld, K.L., and Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. Neuron *100*, 490–509. https://doi.org/10.1016/j.neuron.2018.10.002.

103. Theves, S., Fernandez, G., and Doeller, C.F. (2019). The hippocampus encodes distances in multidimensional feature space. Curr. Biol. *29*, 1226–1231.

104. Viswanathan, P., and Nieder, A. (2020). Spatial neuronal integration supports a global representation of visual numerosity in primate association cortices. J. Cognit. Neurosci. *32*, 1184–1197.

105. Frankland, S.M., and Greene, J.D. (2020). Concepts and Compositionality: In Search of the Brain's Language of Thought. Annu. Rev. Psychol. *71*, 273–303. https://doi.org/10.1146/annurev-psych-122216-011829.

106. Fodor, J.A. (1975). The Language of Thought (Harvard University Press).

107. Sablé-Meyer, M., Fagot, J., Caparos, S., Van Kerkoerle, T., Amalric, M., and Dehaene, S. (2021). Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity. Proc. Natl. Acad. Sci. USA *118*, e2023123118. https://doi.org/10.1073/pnas.2023123118.

108. Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistic Influences on Mathematical Development: How Important Is the Transparency of the Counting System? Philos. Psychol. *21*, 523–538. https://doi.org/10.1080/09515080802285511.

109. Monti, M.M., and Osherson, D.N. (2012). Logic, language and the brain. Brain Res. *1428*, 33–42. https://doi.org/10.1016/j.brainres.2011.05.061.

110. Lampinen, A.K., Dasgupta, I., Chan, S.C.Y., Sheahan, H.R., Creswell, A., Kumaran, D., McClelland, J.L., and Hill, F. (2024). Language models, like humans, show content effects on reasoning tasks. PNAS Nexus *3*, pgae233. https://doi.org/10.1093/pnasnexus/pgae233.

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

112. Wang, Y., Lin, L., Kuhl, P., and Hirsch, J. (2007). Mathematical and Linguistic Processing Differs Between Native and Second Languages: An fMRI Study. Brain Imaging Behav. *1*, 68–82. https://doi.org/10.1007/s11682-007-9007-y.

113. López Espejel, J., Ettifouri, E., Yahaya Alassan, M., Chouham, E.M., and Dahhane, W. (2023). GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts. Nat. Lang. Proc. J. *5*, 100032. https://doi.org/10.1016/j.nlp.2023.100032.

114. Frieder, S., Pinchetti, L., Chevalier, A., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P., and Berner, J. (2023). Mathematical capabilities of ChatGPT. In Proceedings of the 37th International Conference on Neural Information Processing Systems NIPS '23, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, eds. (Curran Associates Inc.), pp. 27699–27744.

115. Wong, L., Grand, G., Lew, A.K., Goodman, N.D., Mansinghka, V.K., Andreas, J., and Tenenbaum, J.B. (2023). From Word Models to World Models: Translating from Natural Language to the Probabilistic Language of Thought. Preprint at arXiv. https://doi.org/10.48550/arxiv.2306.12672.

116. Stöckl, C., Yang, Y., and Maass, W. (2024). Local prediction-learning in high-dimensional spaces enables neural networks to plan. Nat. Commun. *15*, 2344. https://doi.org/10.1038/s41467-024-46586-0.

117. Sun, J., Xu, C., Tang, L., Wang, S., Lin, C., Gong, Y., Ni, L.M., Shum, H.-Y., and Guo, J. (2024). Think-on-Graph: Deep and Responsible Reasoning of Large Language Model on Knowledge Graph. ICLR *2024*.

118. Fedorenko, E., Duncan, J., and Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. Curr. Biol. *22*, 2059–2062.

119. Zhou, X., Li, M., Li, L., Zhang, Y., Cui, J., Liu, J., and Chen, C. (2018). The semantic system is involved in mathematical problem solving. Neuroimage *166*, 360–370. https://doi.org/10.1016/j.neuroimage.2017.11.017.

120. Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., and Wager, T.D. (2011). Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods *8*, 665–670. https://doi.org/10.1038/nmeth.1635.

121. Fox, P.T., Laird, A.R., Fox, S.P., Fox, P.M., Uecker, A.M., Crank, M., Koenig, S.F., and Lancaster, J.L. (2005). Brainmap taxonomy of experimental design: Description and evaluation. Hum. Brain Mapp. *25*, 185–198. https://doi.org/10.1002/hbm.20141.

122. Milivojevic, B., Johnson, B.W., Hamm, J.P., and Corballis, M.C. (2003). Non-identical neural mechanisms for two types of mental transformation: event-related potentials during mental rotation and mental paper folding. Neuropsychologia *41*, 1345–1356. https://doi.org/10.1016/S0028-3932(03)00060-5.

123. Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., and Kanwisher, N. (2010). New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects. J. Neurophysiol. *104*, 1177–1194. https://doi.org/10.1152/jn.00032.2010.

124. Yan, C.-G., Wang, X.-D., Zuo, X.-N., and Zang, Y.-F. (2016). DPABI: Data Processing & Analysis for (Resting-State) Brain Imaging. Neuroinformatics *14*, 339–351. https://doi.org/10.1007/s12021-016-9299-4.

125. Salo, T., Yarkoni, T., Nichols, T.E., Poline, J.-B., Kent, J.D., Gorgolewski, K.J., Glerean, E., Bottenhorn, K.L., Bilgel, M., Wright, J., et al. (2023). Neurostuff/NiMARE: 0.0.13 (Zenodo). https://doi.org/10.5281/zenodo.7545355.

126. Stouffer, S.A., Suchman, E.A., Devinney, L.C., Star, S.A., and Williams, J.,R.M. (1949). The American soldier: Adjustment during army life. In Studies in social psychology in World War II (Princeton Univ. Press).

127. Poldrack, R.A., Mumford, J.A., Schonberg, T., Kalar, D., Barman, B., and Yarkoni, T. (2012). Discovering Relations Between Mind, Brain, and Mental Disorders Using Topic Mapping. PLoS Comput. Biol. *8*, e1002707. https://doi.org/10.1371/journal.pcbi.1002707.

128. Hebart, M.N., Görgen, K., and Haynes, J.-D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. Front. Neuroinformatics *8*, 88.

129. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw fMRI data and other data used Code for analyses	This paper; Mendeley repository	https://data.mendeley.com/preview/bkn4xx4kc3?a=9161e868-1e4e-4819-a9da-5e77c12abff2
Neurosynth database (version 0.5 released on February 23, 2015)	Tal Yarkoni	https://github.com/neurosynth/neurosynth-data/releases/tag/0.5
Neurosynth database-v5 (the database as of July 2018)	Tal Yarkoni	https://www.neurosynth.org/analyses/topics/v5-topics-400/
BrainMap database (accessed in August 19, 2023)	Research Imaging Institute	https://brainmap.org/index.html
Software and algorithms		
MATLAB_R2022b	Math Works, Natick, MA, USA	https://it.mathworks.com/products/matlab.html
Statistical Parametric Mapping (SPM 8)	Wellcome Department of Imaging Neuroscience, London, UK	https://www.fil.ion.ucl.ac.uk/spm/software/download/
DPABI	Chaogan-Yan	https://github.com/Chaogan-Yan/DPABI
The Decoding Toolbox (TDT)	Martin N Hebart, Kai Görgen and John-Dylan Haynes	https://sites.google.com/site/tdtdecodingtoolbox
BrainNet Viewer	Beijing Normal University	http://www.nitrc.org/projects/bnv/
Python	Python Software Foundation	https://www.python.org/
SciPy Python package	The SciPy community	https://docs.scipy.org/doc/scipy-1.5.4/reference/install_upgrade.html
NiMARE Python package	neurostuff/NiMARE: 0.0.13	https://doi.org/10.5281/zenodo.7545355
Neurosynth Python package	Tal Yarkoni	https://github.com/neurosynth/neurosynth
BrainMap Sleuth	Research Imaging Institute	https://brainmap.org/sleuth/
GingerALE	Research Imaging Institute	https://brainmap.org/ale/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

Twenty healthy participants (right-handed neurologically normal volunteers with normal or corrected-to-normal vision) completed the experiment. All participants were graduate or undergraduate students (around 20 years old) at Beijing Normal University and identified as Han Chinese. The participants were screened before recruitment to ensure they either had taken programming-related courses or had experience in programming. The targeted sample size (20 valid participants) was decided *a priori*, to be similar to previous studies using similar tasks (e.g., Fedorenko et al.,[118] n = 13–16; Zhou et al.,[119] n = 24; Amalric & Dehaene,[26] n = 15). The study was approved by the Institutional Review Board of BNU. Written informed consent was obtained from all participants before they took part in the experiment, and the participants received money for their time. For each participant, the fMRI data of three LMS processing tasks, one language processing task, and one spatial cognition task were recorded and analyzed in the present study.

METHOD DETAILS

Meta-analyses of language processing and spatial cognition

We generated meta-analysis activation maps based on large-scale automated meta-analyses of functional magnetic resonance imaging (fMRI) for language processing and spatial cognition.

Meta-analysis of language processing

To generate a map of the neural correlates of language processing, we queried the Neurosynth[120] database (version 0.5 released on February 23, 2015) using the term "language" (accessed on 2023.8.19). This yielded a collection of 823 fMRI studies that exhibited significant relevance to the term with the default frequency threshold of 0.001. Based on the coordinate-based meta-analysis technique of Neurosynth, we generated the map of brain regions that were preferentially engaged across the selected studies. Each voxel within the resulting map was labeled by the Z score resulting from a two-way ANOVA assessing the concomitant presence of term

loading and voxel activation. The map was further corrected for multiple comparisons employing the false discovery rate (FDR) approach with an anticipated FDR threshold of 0.01.

Meta-analysis of spatial cognition

As for spatial cognition, there are no uniform terms to search in Neurosynth. Therefore, we used the BrainMap database to generate the meta-analysis map of spatial cognition. Following the taxonomy outlined by Fox et al.,[121] we curated the experiments categorized under the Behavioral Domain "Cognition.Spatial" in the BrainMap functional database (accessed by Sleuth 3.0.4 in 2023.8.19), under the constraint of "activations only". This curation yielded 251 experimental contrasts from 63 relevant publications. The meta-analysis based on the experimental contrasts was implemented by the activation likelihood estimation (ALE), with BrainMap's GingerALE tool. Same as the language processing analysis, we applied FDR correction (at $p = 0.01$) to the resulting Z score map, adhering to the default GingerALE settings.

FMRI studies

The present study used fMRI data from a separate sample of participants ($n = 20$) to estimate the neural correlates of a set of representative LMS processing tasks, in addition to a representative language processing task and a representative spatial cognition task. The representative task for calculation is number addition, adopted from Fedorenko et al.[58] The representative task for logical reasoning is adopted from Raven's Advanced Progressive Matrices.[59] The representative task for problem-solving is mental programming requiring algorithm generation for programming problems, which were adopted from elementary training problems of programming competitions.[37] The representative task for language processing is sentence understanding and was adopted from Fedorenko et al.[58] The representative task for spatial processing is mental folding and was adopted from Milivojevic et al.[122]

Analyses based on these data for a different purpose have been reported in a previous study.[37]

Experimental paradigms

LMS processing. BOLD signals during three representative LMS processing tasks, namely mental programming, reasoning, and calculation, were collected.

The mental programming task engaged participants in mentally solving simple programming problems, with reporting familiarity with knowledge entries as the control condition. The programming problems were presented in Chinese as a one-to-three-line introductory statement, a programming probe ("Please design a program with:"), a one-line statement specifying the potential input, and a one-line statement describing the desired output. The knowledge entries in the control condition were matched with the materials of the programming condition in length, format, and the use of symbols. Each entry started with a one-to-three-line introductory statement, followed by a brief conjunctive clause (e.g., "in other words"), and two one-to-two-line elaborations of the concept introduced in the introductory statement. The mental programming trials started with a programming task probe presenting "report progress", and the participants were to read and consider the programming problem, and wait for the response prompt to choose from a four-point scale to report their progress in programming. The participants were instructed before the scanning that for programming they should consider the algorithm instead of specific code, and the solution of the problem means finding the algorithm instead of finishing the coding. The control trials started with the control task probe which read "report familiarity", and participants were to read and consider whether they were familiar with the content of the knowledge entry and choose from a four-point scale to report their familiarity with the content. The mental programming task was divided into six runs, each consisting of four trials in the mental programming condition and four in the control condition.

The reasoning task comprised a reasoning condition adopted from Raven's APM and a control condition. In both conditions, the participants first saw a problem with a complex main figure, with eight choices below. The problems of the reasoning trials were chosen from Raven's APM, and those of the control condition were composed by rearranging the choices of other Raven's APM items in the main figure and the choices in each problem. Above each problem, a task probe would indicate whether it is a reasoning trial ("Look for the rule") or a control trial ("Look for the choices"). The participants were to consider which choice complies with the rule of figure arrangement and completes the main pattern (the reasoning condition), or to consider which choice was presented in the main pattern (the control condition). Each participant completed five functional runs consisting of eight trials each.

The calculation task was adopted from Fedorenko et al.[58] with an additional control condition. In the calculation trials, the participants would first see a task probe ("Do the addition") and then see one number (11–30) followed by three sequentially presented addends. Then the participants had to choose the correct sum in a two-choice, forced-choice question. In the control trials, after the task probe ("Memorize the numbers") and the same sequential presentation of four numbers, the participants were to select, from two alternatives, the one among the four preceding numbers. Each participant completed four runs which each consisted of six calculation blocks and six control blocks.

Language processing and spatial cognition

The sentence processing task was used as the representative language processing task. It was adopted from Fedorenko et al.[123] but used sentences and pseudo-characters in Mandarin Chinese. Each experimental trial consisted of a sequential presentation of a string of seven Chinese real words which formed one sentence followed by a target word. The task was to decide whether the target word appeared in the preceding string. In control trials, the Chinese characters were replaced by pseudo-characters. The trials were grouped into six-trial blocks. Each participant underwent four block-design runs, each consisting of three experimental blocks and three control blocks.

Spatial cognition was measured in a mental folding task adopted from Milivojevic et al.[122] In each trial, the participants were shown a black outline of six squares joined together, representing the unfolded surface of a cube. Two sides belonging to different squares (mental folding condition) or the same square (control condition) were pointed by two small arrows. In the "match" trials, the pointed sides would meet if the squares were folded up into a cube. Trials in the mental folding and the control conditions differed in the total number of squares carried along for each fold for the match–mismatch decision to be made. The control trials presented 1-square-carried stimuli since the judgment of such stimuli can be made without mental folding. Every four trials of the same condition were grouped into a block. Each participant completed five or six block-design functional runs. Note that the first three participants completed six runs, and the rest participants underwent only five runs due to the limitation of the scanning time.

For a complete description of the experimental paradigms, see Xu et al.[37]

FMRI data acquisition and analysis

Image acquisition. The fMRI data were acquired using a Siemens 3T scanner with a 12-channel phased-array head coil. Task-state fMRI (ts-fMRI) was acquired using a T2*-weighted echo-planar-imaging (EPI) sequence with a whole-brain protocol (TR = 2000 ms, TE = 30 ms, flip angle = 90°, and in-plane resolution = 3.1 × 3.1 × 3.5 mm, 33 contiguous interleaved slices). For a complete description of the imaging parameters, see Xu et al.[37]

Preprocessing. The functional data were preprocessed with DPABI.[124] The main preprocessing procedure included slice timing, head motion correction, co-registration, segmentation, spatial normalization to the standard MNI space, and resampling to 3 × 3 × 3 mm isotropic voxels. The data were smoothed with a 4 mm FWHM kernel for univariate activation analysis. Smoothing was omitted for multivariate activation analysis. Runs characterized by excessive head movement were excluded from further analysis.

Univariate analyses. First-level analyses were performed for functional images of each participant in each task using the general linear model (GLM). The GLM incorporated regressors corresponding to the following events: the task condition, the control condition, and the presence of the response prompt and task probe. The regressors were convolved with the canonical hemodynamic response function (HRF). A 1/128 Hz high-pass filter was applied to remove low-frequency noise, with the AR (1) model used to account for serial correlations. For each experiment, subject-level and group-level analyses were conducted within a gray matter mask by SPM8 (Wellcome Department of Imaging Neuroscience, London; www.fil.ion.ucl.ac.uk/spm). The subject-level map corresponded to the contrasts of parameter estimates (COPE) between the respective task versus the control condition of each subject. The group-level map was estimated in the second-level analysis of the corresponding task-control contrasts. The activation maps were visualized with the BrainNet Viewer.[58]

Conjunctive analysis of LMS processing

To estimate the neural correlates across diverse LMS processing tasks, we generated a synthesized activation map of LMS processing based on the data of the three representative LMS tasks, i.e., reasoning, calculation, and mental programming tasks described above. We utilized Stouffer's method in the image-based meta-analysis. The unthresholded group-level maps of each LMS processing task were subjected to Stouffer's estimator using the NiMARE Python package version 0.0.13.[125] The estimator combined the group-level maps via Stouffer's method[126] and produced a fixed-effect estimate of the combined effect, to obtain the most sensitive pooling of the studies. Each voxel within the resultant synthesized map conveyed a third-level COPE, indicating the pooled activation across LMS processing tasks. The synthesized map was corrected with an expected FDR of 0.01.

Dice coefficients of meta-analysis maps and subject-level maps

To quantify the extent of overlap between LMS processing and language processing and that between LMS processing and spatial cognition, we computed dice coefficients between the activation maps for each pair, which was an index of the convergence between activation maps. For a given pair of maps, the dice coefficient was determined as the ratio between twice the count of shared voxels and the sum of voxels in both maps. The dice coefficients were calculated between the thresholded meta-analysis map of language processing and the thresholded synthesized map of LMS processing, and between spatial cognition and LMS processing. In addition, for statistical comparison, corresponding dice coefficients were calculated for each subject based on the corresponding subject-level maps (thresholded by a voxel-level threshold at $p = 0.001$).

Topic-based meta-analysis maps of LMS processing

To rule out the impact of task specifics, we retrieved meta-analysis maps of each core cognitive unit of LMS from Neurosynth by the terms calculation, reasoning, and (problem) solving, and similarly calculated their dice coefficients with the meta-analysis maps of language processing and spatial cognition (Figure S4).

Further, to rule out the possibility that the three core cognitive units are not sufficient to represent the within-domain variations in LMS, we conducted a thorough search among the 400 topic-based meta-analyses from Neurosynth's database-v5(2018). A Neurosynth topic represents a cluster in the latent conceptual structure of fMRI literature, generated by latent Dirichlet allocation based on a large corpus.[127] Each topic is defined by a distribution of words that frequently appear in research text, associated with a set of papers in which that set of words occurs with high frequency. A meta-analysis is automatically generated for each topic, reflecting, though with noise, neural correlates statistically associated with the cluster of research in question. Among all the topic-based meta-analyses, we searched for topics where any of the top 5 keywords were related to the basic LMS abilities identified in the literature, namely number representation, calculation, reasoning, problem-solving, rule abstraction, and categorization.[5–7] This search led to a list of five topics (see Table S5; Figure 2).

Topic-based language processing meta-analysis maps

For an in-depth examination of language-LMS association, we decomposed language processing into sub-domains, and calculated their dice coefficients separately with the LMS domain-level map shown in Figure 1. Specifically, in the 400 topic-based meta-analyses based on Neurosynth database-v5 (2018), we identified topics related to language processing by searching for those where any of the top 5 keywords were related to major aspects of language processing identified in the literature,[16,62,63] namely language production, naming, language comprehension, orthography, phonology, lexical, word recognition, semantic, syntactic, sentence comprehension, and discourse comprehension. This search led to a list of nine topics (see Table S6; Figure 3).

Multivariate pattern analysis

To compare the multivariate pattern similarity between LMS processing and those associated with language processing and spatial cognition, we conducted multivariate pattern analysis in an ROI-based manner. The region of interest (ROI) was defined as the common activation region between the meta-analysis maps of language processing and spatial cognition.

The comparison of multivariate activation pattern similarity was quantified by the confusion rate in multivariate activation pattern classification. Specifically, we trained a classifier for each subject to classify the multivariate activation pattern in the run-level beta map of the sentence processing task from that of the mental folding task. The run-level beta map (task versus fixation) was based on the GLM using non-smoothed data. The classifier was based on the support vector machine (SVM) with a linear kernel implemented by The Decoding Toolbox (TDT),[128] and trained with the leave-one-out cross-validation (LOOCV) to prevent overfitting. In each cross-validation step, one SVM was trained by three runs of sentence processing task and three runs of mental rotation task separated from one run of each task kept unknown. Then in the test session, the SVM would classify the left-out run from each training task for performance validation (indicated as accuracy). For quantitative balance, all tasks included the first four runs in a set. Note that the outcome appeared consistent regardless of which specific four runs were chosen in the examination. The LOOCV traversed all combinations of left-out runs and resulted in an averaged classifier of each participant. The classifier achieved an accuracy of nearly 100% in distinguishing the two tasks across 20 participants. In the test session, each run of LMS processing tasks was classified by the SVM into one of the training tasks. The confusion rate at which LMS processing was (mis)classified as a given training task quantified the pattern resemblance between the LMS processing task and the corresponding training task.

To test the relationship between LMS processing and spatial cognition, we calculated the intra-LMS similarity, LMS-spatial similarities, and LMS-language similarity based on the same run-level beta maps in the ROI. For each participant, the multivariate pattern for a representative task was represented as a one-dimensional vector concatenated sequentially from the first four runs' vectors. The pairwise multivariate pattern similarities were calculated using Spearman's r for all corresponding combinations of two tasks for each participant and statistically tested at the group level.

To further investigate the hierarchical multivariate pattern similarity between LMS processing and spatial cognition, agglomerative hierarchical clustering analysis was conducted on the multivariate patterns of the tasks, based on the same run-level beta maps in the ROI. For each representative task, the multivariate pattern was a one-dimensional vector formed by concatenating the multivariate pattern from each participant. The clustering analysis was conducted with the nearest point algorithm implemented by the SciPy Python package version 1.5.4.[129]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using the SciPy Python package version 1.5.4.[129] The paired t-test (two-tailed) was used to assess the significance of differences in subject-wise activation overlap and multivariate pattern similarity. Specifically, we compared the dice coefficients of LMS processing (using both synthesized and single representative maps, as shown in Figure 4B) with spatial cognition and language processing correspondingly, among the 20 participants ($n = 20$). We also compared the mean confusion rate of LMS processing tasks (Figure 5F) from each participant ($n = 20$). The two-sample t-test (two-tailed) was used to assess the significance of differences in intra- and inter-domain multivariate pattern similarities, using each corresponding task pair from each participant ($n_1 = 60$, $n_2 = 60$). Further, to assess whether intra-LMS similarity exceeded LMS-spatial similarity, we used one-tailed two-sample t-test ($n_1 = 60$, $n_2 = 60$). Additionally, we computed Cohen's d as a measure of effect size to quantify the magnitude of the observed differences. For paired samples, Cohen's d was calculated as the mean difference divided by the pooled standard deviation of the differences. For independent samples, Cohen's d was calculated by dividing the mean difference by the pooled standard deviation of the two groups. The details for data illustrations, the associated statistical analyses and sample sizes are indicated in individual results. P-values smaller than 0.05 were considered as significant. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.