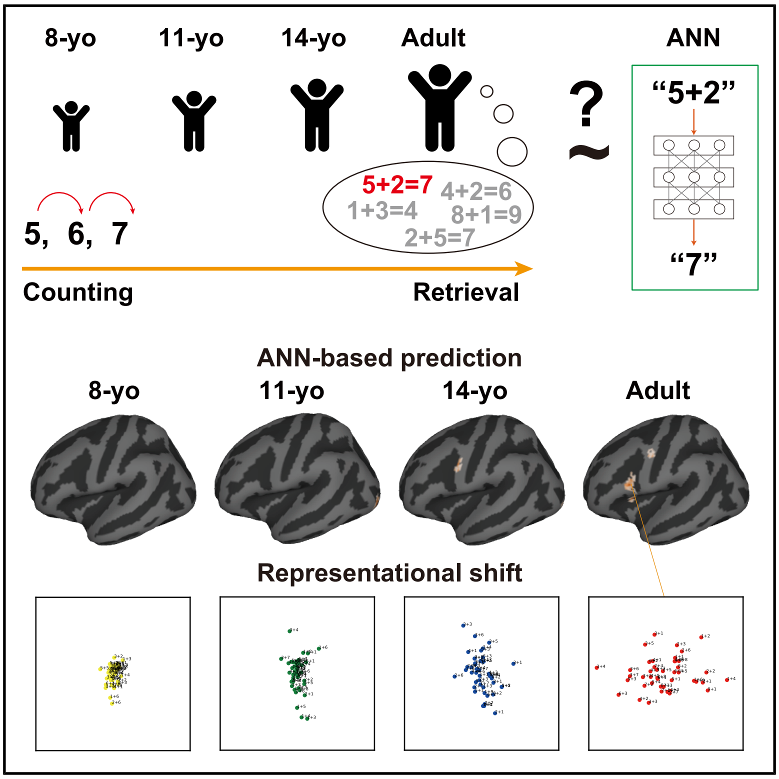
iScience

Article

Developmental changes in similarity between neural representations of mental arithmetic and artificial neural networks

Graphical abstract



Authors

Tomoya Nakai, Jérôme Prado

Correspondence

tnakai@g.ecc.u-tokyo.ac.jp (T.N.),
jerome.prado@cnrs.fr (J.P.)

In brief

Developmental neuroscience; Cognitive neuroscience; Biocomputational method; Computer modeling

Highlights

- ANN-derived features increasingly predict addition-related brain activity with age
- This increase was restricted to the left precentral sulcus
- Smaller arithmetic problems were better predicted in adults
- Some neural representations of arithmetic became more discrete with age

Nakai & Prado, 2026, iScience *29*, 117180
September 18, 2026
https://doi.org/10.1016/j.isci.2026.117180

CellPress

iScience

CellPress
OPEN ACCESS

Article

Developmental changes in similarity between neural representations of mental arithmetic and artificial neural networks

Tomoya Nakai[1,3,*] and Jérôme Prado[2,*]

[1]Graduate School of Information Science and Technology, the University of Tokyo, Tokyo, Japan

[2]Lyon Neuroscience Research Center (CRNL), INSERM U1028 - CNRS UMR5292, University of Lyon, Bron, France

[3]Lead contact

*Correspondence: tnakai@g.ecc.u-tokyo.ac.jp (T.N.), jerome.prado@cnrs.fr (J.P.)

https://doi.org/10.1016/j.isci.2026.117180

SUMMARY

The ability to learn simple arithmetic is often attributed to the associative nature of human memory, where repeated exposure strengthens the relations between operands and outcomes. Because artificial neural networks (ANNs) also learn input-output associations, we tested whether ANN-derived features would increasingly capture arithmetic-related brain activity with development. We analyzed fMRI responses to addition problems in 104 participants across four age groups (8-, 11-, 14-year-olds, and adults). An ANN-based encoding model better predicted activity in older than younger participants in the left precentral sulcus. In adults, prediction accuracy was higher for smaller than larger problems. ANN prediction patterns were consistent with increasingly discrete representations of individual addition problems with age. These findings suggest that some neural representations of arithmetic problems become discretely organized with development. However, ANN-derived features captured this change only in the left precentral sulcus, suggesting that associative mechanisms may account for only part of arithmetic processing.

INTRODUCTION

Our capacity to learn is largely thought to result from our ability to form and strengthen associations between elements. The idea that human memory is associative in nature has profound historical roots in both philosophy and psychology,[1] supporting cognitive models that posit the existence of networks of associations in the human mind.[2] These accounts not only assume that associative bonds can form between perceived elements, but also that bonds can form between elements and the outcomes of mental computations that are performed on them. The domain of simple arithmetic learning provides a compelling illustration of this principle. For example, it has been argued that there is a shift in the type of overt strategy children use to solve simple arithmetic problems, such as 5 + 2, with learning and experience.[3–6] Up until the age of 9 or 10, children tend to solve these problems by using counting strategies, first with manipulatives (e.g., fingers) and then without.[7] As children gain experience with these problems, overt counting strategies become less frequent, and answers to these problems are progressively given quickly without any conscious report of a specific strategy.[8,9] This has been traditionally interpreted as evidence of direct recollection from long term-memory, which would be made possible by the progressive forming and strengthening of associations between operands (e.g., 5 + 2) and outcomes (e.g., 7) over the course of multiple exposures during learning.[3,10,11] In sum, mental arithmetic would be a representative example of a domain in which the associative nature of human memory is thought to support the shift from relatively slow and inefficient computations to relatively fast and efficient retrieval of associations from memory.

The hypothesized process of strengthening associations between operands and answers when children learn simple arithmetic may, to some extent, be related to the way learning is achieved in large-language artificial neural networks (ANNs). ANNs are computational models inspired by biological neural networks (BNNs), usually organized in a directed network that may include multiple layers of neuronal units.[12,13] Each layer aggregates the outputs of preceding units, applies (typically nonlinear) transformations, and transmits the result to the next layer, enabling the network to learn complex input-output relationships. Much like natural language, mathematics involves processing sequences of symbols.[14] Building on the recently developed Transformer model that can handle long-distance dependency in natural languages,[15] several models have now been proposed to process mathematical expressions.[16–20] Trained on large text corpora that include mathematical content, such ANNs output symbols based on statistical associations with the input symbol sequence (i.e., a math expression). Although ANNs can capture structural information in math expressions,[21] their math ability may partly depend on the memorization of training samples, reflecting their reliance on statistical associations learned from data.[22,23] In this sense, ANNs process math expressions by leveraging these associations and may therefore

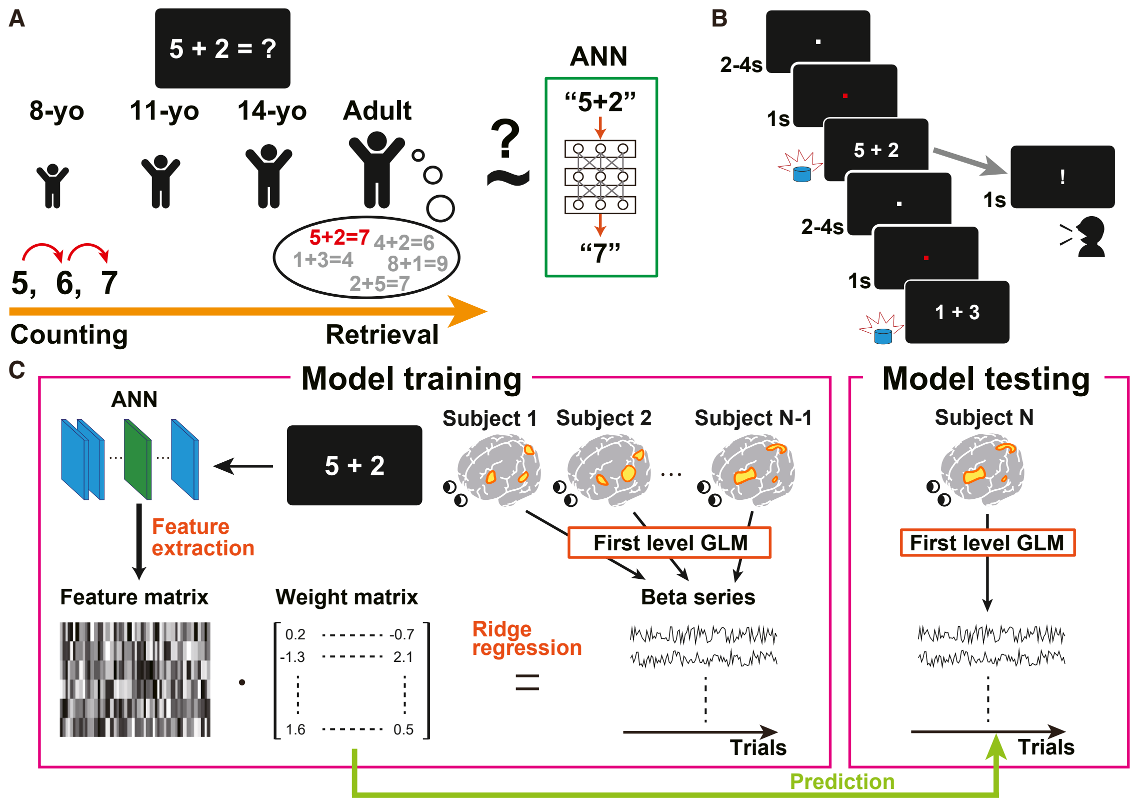


Figure 1. Hypothesis, experimental design, and analyses

(A) Associative models of arithmetic learning assume that the decrease in overt counting strategies observed across development is explained by an increase in associations between operands of problems and answers, which would then be retrieved (see main text). Therefore, these models predict that solving arithmetic problems in older children and adults should be comparable to the way artificial neural networks (ANNs) provide the answer to an arithmetic problem.

(B) In the scanner, participants were presented with single-digit addition problems and pressed a response button when they came up with an answer in their head (timing was self-paced). In some trials, participants said the answer aloud after being presented with an exclamation point.

(C) Trial-wise brain signals were estimated for each subject using a general linear model (GLM). Vertex-wise encoding models were trained using feature matrices extracted from addition problems via ANNs and trial-wise brain signal estimates concatenated across participants. Prediction performance was evaluated with the data of the left-out subject.

serve as a useful computational tool to characterize representational changes related to the retrieval of associations between problem elements and their outcomes.

Recent neuroimaging studies have observed similarities between the adult human brain and ANNs. For example, there is a correspondence between the hierarchy of human visual pathways and that of ANN layers in processing natural images,[24–26] where shallower layers correspond to low-level features (e.g., local contrast, edge, and so forth) and deeper layers correspond to more complex information (e.g., entire shape). Similar hierarchical correspondence has been reported for auditory[27] and language processing.[28] In the math domain, a Transformer-based ANN model has been shown to predict brain activity of adult participants when they solve a variety of arithmetic problems.[29] Much like with complex visual shapes, deeper ANN layers showed a better fit to the activity patterns of mathematical operators, suggesting that the latent features underlying mathematical operations reflect higher-order information integrated through cumulative transformations of stimulus input. However, this study only included adult participants, who represent the later point of arithmetic learning. It remains unknown whether an ANN model may also predict arithmetic-related activity in children who are still in the earlier process of learning arithmetic, and may therefore rely to a greater extent on counting strategies. Clearly, associationist models of arithmetic learning[3,10,11] would hypothesize that the performance of an ANN model in predicting arithmetic-related activity would improve with age and degree of exposure to simple problems, as associations between operands and answers would strengthen (Figure 1A).

To test this hypothesis, we measured fMRI activity of 104 participants in four age groups (8-year-olds [yo], 11-yo, 14-yo, and adults) while they solved single-digit addition problems (Figure 1B). We then compared arithmetic-related activity with ANN features using vertex-wise encoding models.[30] Encoding models, which are widely used in studies comparing brain activity and various ANN models, predict brain activity by a weighted linear combination of features extracted from experimental stimuli.[24,25,29,31,32] To mitigate effects of differences in processing times between different age groups, we first obtained trial-by-trial beta estimates of brain activity by means of a general linear

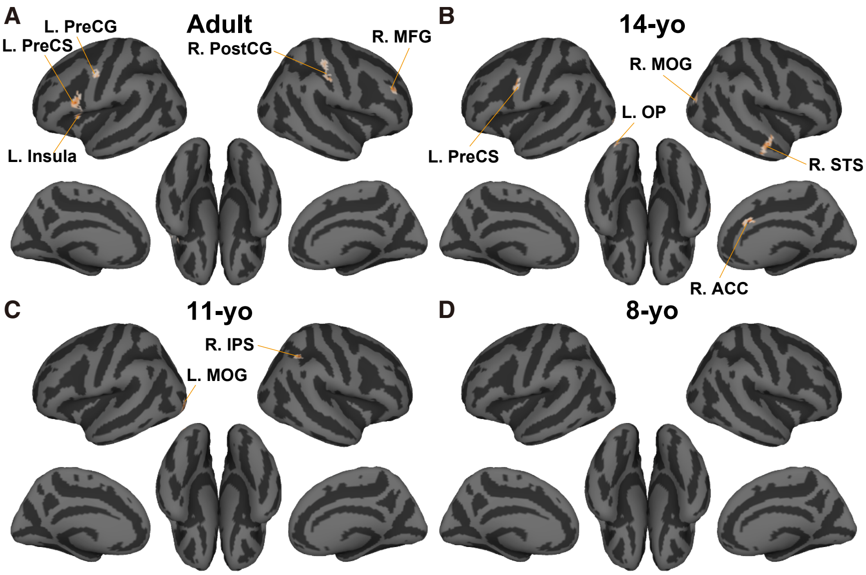


Figure 2. Prediction performance of ANN encoding model for each age group
Prediction accuracy of ANN encoding model using the 12th layer mapped onto the cortical surface for (A) adults, (B) 14-yo, (C) 11-yo, and (D) 8-yo. Only significantly predicted vertices are shown (Wilcoxon signed-rank test: peak-level, uncorrected $p < 0.001$; cluster-level, false-discovery rate [FDR] corrected $p < 0.05$).

model for each participant. We then built a series of vertex-wise encoding models predicting brain activity (trial beta estimates) from latent features extracted from a pre-trained large language ANN model (MathBERT)[17] (Figure 1C). The encoding model was trained for each age group and tested using leave-one-subject-out cross-validation.

This computational approach enabled us to test two main hypotheses. First, if operands become increasingly associated with answers of problems over the course of learning and exposure to arithmetic problems, then neural representations associated with simple addition problem-solving may be increasingly captured by ANN-derived features. Therefore, the ANN-based encoding model should show higher prediction performance for older participants (i.e., 14-yo and adults) than younger participants (i.e., 11-yo and 8-yo). Second, associationist accounts often predict that the strength of associations between operands and answers depends on the frequency of exposure,[3,33] which is believed to be higher for problems with a small sum than problems with a large sum (the latter being found less frequently in formal schooling[34]). Therefore, the ANN-based encoding model should show better prediction performance for relatively small addition problems (i.e., problems with operands ≤ 4) than problems that are larger in comparison (i.e., problems with operands ≥ 5).

RESULTS

ANN better predicts arithmetic-related activity in older than younger participants in the left frontal cortex

To test whether ANN features may better predict brain activity during addition problem solving in older than younger participants, we independently constructed an encoding model (hereafter referred to as "ANN model") and quantified model performance for each of the four age groups. Based on the observation in our previous study,[29] we expected that deeper ANN layers would capture higher-order information related to mathematical operations than shallower layers and thus extracted ANN features from the last layer of the MathBERT. In adults, the ANN model significantly predicted activity in several brain regions, including the left precentral sulcus (L. PreCS), left precentral gyrus (L. PreCG), left insula, right middle frontal gyrus (R. MFG), and right postcentral gyrus (R. PostCG) (Wilcoxon signed-rank test: peak level, uncorrected $p < 0.001$; cluster level, corrected $p < 0.05$) (Figure 2A). In 14-yo, prediction accuracy was significant in the L. PreCS, as well as in the left occipital pole (L. OP), right anterior cingulate cortex (R. ACC), right superior temporal sulcus (R. STS), and left middle occipital gyrus (R. MOG) (Figure 2B). In 11-yo, the ANN model predicted activity in the L. MOG and right intraparietal sulcus (R. IPS) (Figure 2C). Finally, in 8-yo, no brain regions showed significant prediction accuracy (Figure 2D). Critically, the number of significant vertices increased with age (adults, 121 vertices; 14-yo, 92 vertices; 11-yo, 43 vertices; 8-yo, 0 vertices), even without cluster thresholding (adults, 391 vertices; 14-yo, 335 vertices; 11-yo, 219 vertices; 8-yo, 236 vertices). Therefore, the ANN model appeared to predict activity in larger brain regions in older participants.

We then formally tested whether the prediction performance of the ANN model increased with age (and therefore frequency of exposure to arithmetic problems). We first examined whether some brain regions displayed different prediction accuracy between the four age groups using a Kruskal-Wallis test (peak level, uncorrected $p < 0.001$; cluster level, corrected $p < 0.05$). To ensure that any between-group effect would only be observed in brain regions that also showed significant prediction accuracy in at least one of the groups, we inclusively masked that contrast by brain regions showing significant prediction accuracy in at least one of the four age groups (i.e., the union of Figures 2A–2D). We found a significant main effect of age in the anterior and posterior L. PreCS (Figure 3A), indicating that prediction performance of the ANN model differed between groups in these two prefrontal clusters. Using post-hoc paired tests between groups, we then compared prediction accuracy by the ANN model for all combinations of age groups using functional regions-of-interest (ROIs) in the anterior and posterior PreCS. The two functional L. PreCS subclusters showed distinct patterns. In the anterior L. PreCS subcluster, which was located on the boundary of the opercular part of left inferior frontal gyrus (L. IFG), prediction accuracy was significantly larger in adults than in the other three groups (Wilcoxon rank-sum test: Adult vs. 14-yo, $p < 0.001$; $d = 1.85$; Adult vs. 11-yo, $p < 0.001$,

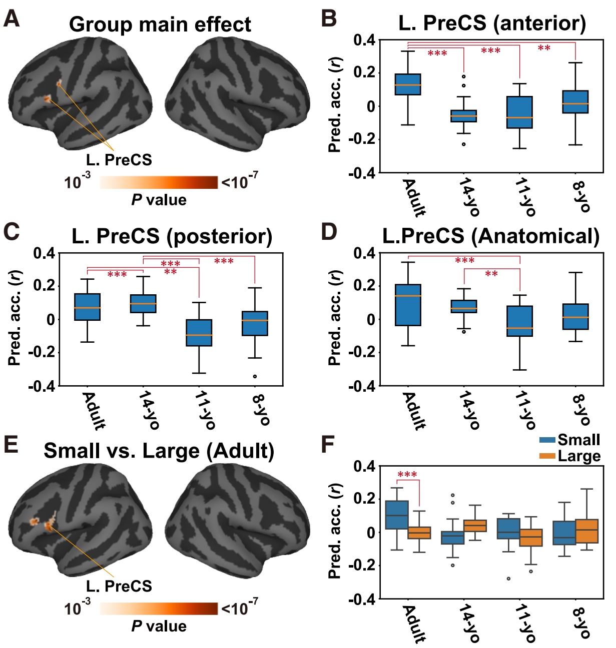


Figure 3. ANN prediction performance increases with age and decreases with problem size in the left PreCS

(A) A whole cortical map of the main effect of prediction accuracy across four age groups (Kruskal-Wallis test), within the inclusion mask of the union of Figures 2A–2D. Statistical significance is evaluated with a peak-level $p < 0.001$ and cluster-level $p < 0.05$ (FDR corrected).

(B–D) Boxplots show the prediction accuracy of four age groups using the ANN model, averaged in the (B) anterior and (C) posterior parts of the L. PreCS functional ROIs, which are extracted from Figure 3A, and (D) anatomical L. PreCS ROI.

(E) A whole cortical map of the direct comparison of prediction accuracy between small and large problems (Wilcoxon signed-rank test) for adults (see Figure S3 for other age groups).

(F) A boxplot shows prediction accuracy of four age groups, separately calculated for small $\leq$4) and large ($\geq$5) problems, averaged in the anatomical L. PreCS ROI. Boxplots show the median and interquartile range, with whiskers extending to the most extreme values within 1.5 times the interquartile range. Individual points beyond the whiskers represent outliers. *, $p < 0.05$; **, $p < 0.01$; and ***, $p < 0.001$. Age-group comparisons used Wilcoxon rank-sum tests, and small-versus-large comparisons used Wilcoxon signed-rank tests.

$d = 1.60$; Adult vs. 8-yo, $p = 0.001$, $d = 0.90$; Figure 3B). In the posterior L. PreCS subcluster, both adults and 14-yo showed larger prediction accuracy than 11-yo and 8-yo (Adult vs. 11-yo, $p < 0.001$, $d = 1.48$; Adult vs. 8-yo, $p = 0.003$, $d = 0.82$; 14-yo vs. 11-yo, $p < 0.001$, $d = 1.90$; 14-yo vs. 8-yo, $p < 0.001$, $d = 1.15$; Figure 3C). Finally, a L. PreCS ROI defined anatomically showed a significant main effect of age (Kruskal-Wallis test, $p = 0.002$, $\eta^2 = 0.161$) and larger prediction accuracy in Adult and 14-yo than 11-yo (Adult vs. 11-yo, $p < 0.001$, $d = 1.03$; 14-yo vs. 11-yo, $p < 0.003$, $d = 0.98$; Figure 3D; also see Figure S1 for other anatomical ROIs). We did not observe significant activation differences in analyses based on beta contrasts, either across the cortex or in ROI-based analyses (all $p > 0.066$; Figure S2). These results suggest that the ANN model better predicts arithmetic-related activity in older than younger participants in the L. PreCS.

As a control analysis, we performed the same group-level comparison using features extracted from a ResNet-18 model, a convolutional neural network that captures visual properties of the stimuli and has an architecture distinct from the Transformer-based MathBERT model. As with the MathBERT model, this analysis was restricted to regions showing significant prediction accuracy in at least one of the four age groups using an inclusive mask. In contrast to the MathBERT model, this analysis did not reveal any significant main effect of age across the cortex (Kruskal-Wallis test, peak-level uncorrected $p < 0.001$; cluster-level corrected $p < 0.05$). This suggests that the age-related increase in prediction accuracy observed in the L. PreCS is not driven by low-level visual features of the stimuli.

One concern in evaluating between-age differences is the potential impact of differences in signal-to-noise ratio (SNR). Indeed, the higher prediction accuracy in the older group might be attributed to their higher SNR. To exclude this possibility, we calculated noise ceiling (i.e., maximum prediction accuracy) and SNR values for each subject and compared them between different age groups[35] (Figure S3). In either index, we did not find any significant group differences for the whole-cortex comparison or the anatomical ROI analysis in the L. PreCS (noise ceiling, $p = 0.849$, $\eta^2 = 0.009$; SNR, $p = 0.845$, $\eta^2 = 0.007$), suggesting that the higher prediction accuracy of older groups was not caused by a reduced noise level.

ANN better predicts arithmetic-related activity for smaller than larger problems in the left frontal cortex of adult participants

Next, we tested whether prediction performance was larger for small problems (where the operands are $\leq$4) than for large problems (where the operands are $\geq$5). Across the whole cortex, the contrast revealed significantly better prediction accuracy for small than large problems in the L. PreCS and left inferior frontal sulcus in adult participants, but not in any other groups (Figures 3E and S4). An anatomical ROI analysis in the L. PreCS further replicated this better prediction accuracy for small than large problems in adults (Wilcoxon signed-rank test, $p < 0.001$, $d = 1.26$; Figure 3F). The difference between small and large problems was greater in adults than in other groups (Adult vs. 14-yo, $p < 0.001$, $d = 1.52$; Adult vs. 11-yo, $p = 0.017$, $d = 0.63$; Adult vs. 8-yo, $p < 0.001$, $d = 1.07$). We did not observe significant beta-value differences between small and large problems in any age group, either at the whole-brain level or in ROI-based analyses (all $p > 0.079$; Figure S5). These results suggest that the ANN model could better predict activity associated with small problems compared to activity associated with large problems, but only for adults who were the most exposed to arithmetic problems.

ANN performance correlates with that of a model in which representations of individual addition problems are discrete in older participants

The ANN model consists of a large number of parameters, and its internal representation is often difficult to interpret.[36,37] Specifically, it remains unclear how information from input stimuli is transformed into ANN-derived features that both support accurate prediction of brain activity and correspond to interpretable

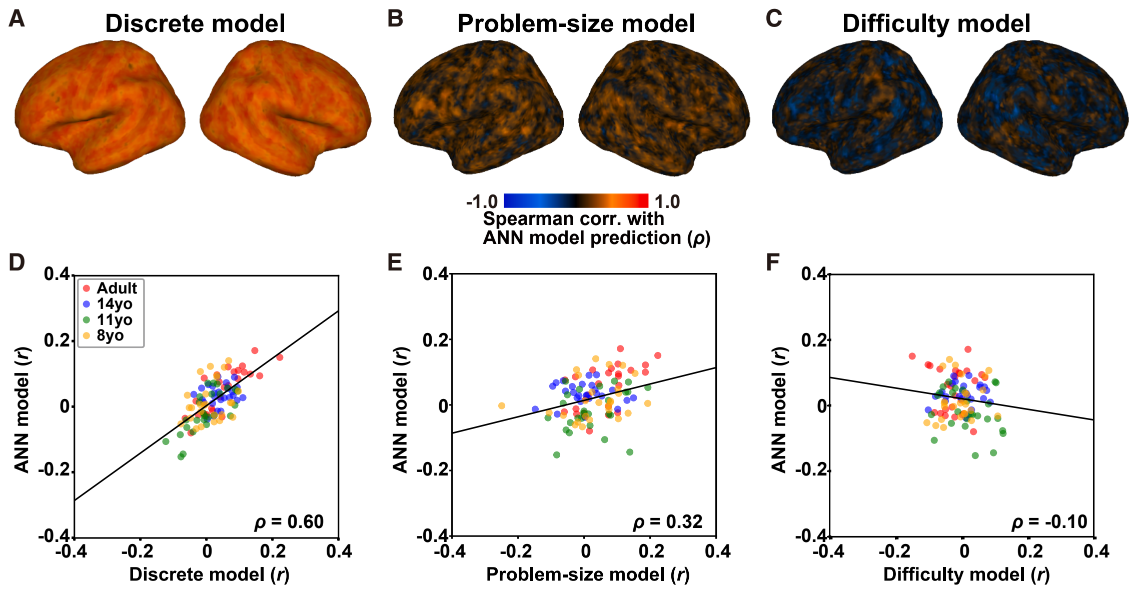


Figure 4. The discrete model exhibited prediction patterns that were similar to the ANN model

(A–C) Whole cortical maps show Spearman's correlation coefficients of prediction accuracy (A) between the discrete and ANN models, (B) between the problem-size and ANN models, and (C) between the difficulty and ANN models.

(D–F) Scatterplots show correlations of prediction accuracies (D) between the discrete and ANN models, (E) between the problem-size and ANN models, and (F) between the difficulty and ANN models, averaged in the left PreCS (anatomical ROI). The regression line and Spearman's correlation coefficient (ρ, indicated at the bottom of each plot) are calculated based on the concatenated data across adults (red dots), 14-yo (blue dots), 11-yo (green dots), and 8-yo (yellow dots).

cognitive functions. To shed some light on the reasons for an age-dependent increase in prediction accuracy of the ANN model, we performed several additional analyses. The first of these analyses assessed the correlation of prediction patterns across participants between ANN and different features. Specifically, we constructed additional encoding models using different features and compared their prediction patterns against those of ANNs. We reasoned that, if the ANN model exhibits similar prediction patterns as encoding models of other features, the ANN should represent information in a way that is similar to these features.

We defined three models. The "discrete" model included features consisting of 45 one-hot vectors (i.e., 45 dimensions) representing all 45 addition problems as discrete instances. The "problem-size" model included a one-dimensional feature corresponding to the sum of addition problems. Finally, the "difficulty" model included a one-dimensional feature corresponding to reaction times of addition problems (averaged across all runs for each participant).

We found similar prediction patterns between the ANN model and the discrete model. The discrete model significantly predicted brain activity in the left PreCS for adults and 14-yo (Figures S6A and SB), but only the left occipital cortex was significant for 11-yo (Figure S6C). No significant regions were found for 8-yo (Figure S6D). Participants showing higher prediction performance with the ANN model also exhibited higher prediction performance with the discrete model across the cortex (Spearman's correlation coefficient, 0.621 ± 0.070; Figure 4A) and in the L. PreCS (ρ = 0.60; Figure 4D).

In contrast, the problem-size and difficulty models displayed distinct patterns. Notably, Figures S7 and S8 show prediction accuracy of these models themselves, rather than their correlation with the ANN model. The mean prediction accuracies in the L. PreCS displayed a smaller correlation between the ANN and problem-size models (across the cortex, 0.148 ± 0.109; in L. PreCS, ρ = 0.32; Figures 4B and E) and between the ANN and difficulty models (across the cortex, 0.007 ± 0.111; in L. PreCS, ρ = −0.10; Figures 4C and F). Direct comparison revealed that the correlation between ANN and discrete models was higher than that between ANN and problem-size models in 99.6% of vertices, and higher than that between ANN and difficulty models in all vertices (permutation tests, FDR-corrected $p < 0.05$), indicating that problem-size information and cognitive load cannot explain prediction patterns of the ANN model. Overall, these results suggest that, in the ANN model, addition problems are more likely represented as discrete categories.

Addition problems are more discretely represented in older participants

A second analysis aiming to shed some light on the reasons for an age-dependent increase in prediction accuracy of the ANN model involved a visualization of 45 addition problems using ANN model weights. Specifically, we visualized the relation among 45 addition problems using principal component analysis (PCA). First, we applied PCA to the discrete model weights (concatenated across four age groups) in the L. PreCS, and mapped all addition problems on the two-dimensional space using the first and second PC loadings.

We found that addition problems are most widely scattered in this space for adults (Figure 5A). Such representational dispersion decreased in younger participants: 14-yo, 11-yo, and 8-yo

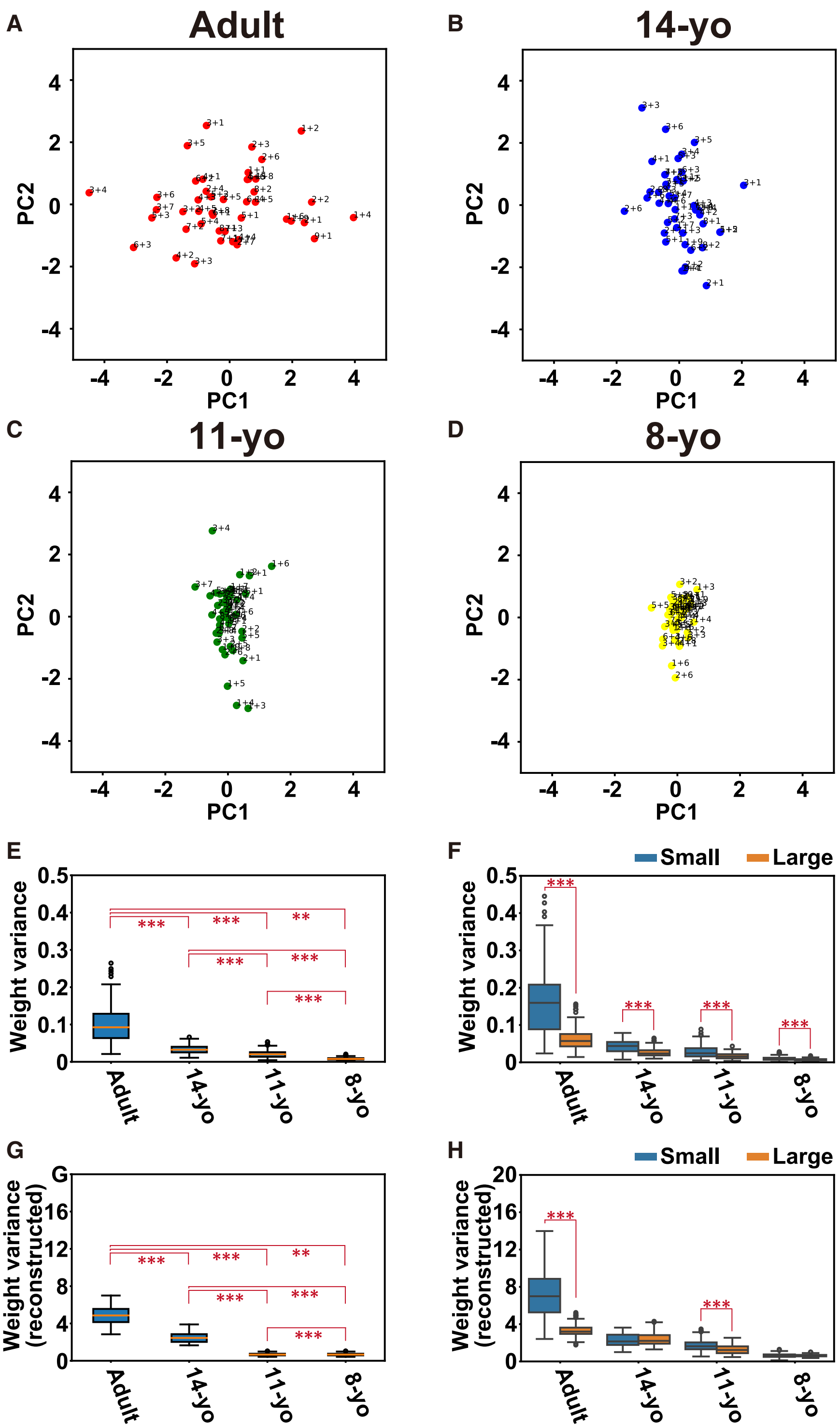


Figure 5. Representations of addition problems become discrete in older participants

(A–D) Principal component analysis based on the weight matrix of the discrete model. 45 addition problems are mapped onto the two-dimensional space of the first and second principal components (PC1 and PC2, respectively) based on their loadings, separately visualized for (A) adults, (B) 14-yo, (C) 11-yo, and (D) 8-yo participants.

(E) The variance of model weights across 45 addition problems is plotted using boxplots.

(F) The variance of model weights is separately calculated for small (problems with operands $\leq$4) and large (problems with operands $\geq$5) problems and plotted using boxplots.

(G) Discrete model weights are reconstructed using the ANN model, and the variance of the reconstructed weights across 45 addition problems is plotted using boxplots.

(H) The variance of reconstructed weights is separately calculated for small (problems with operands $\leq$4) and large (problems with operands $\geq$5) problems and plotted using boxplots. Boxplots show the median and interquartile range, with whiskers extending to the most extreme values within 1.5 times the interquartile range. Individual points beyond the whiskers represent outliers. *, $p < 0.05$; **, $p < 0.01$; and ***, $p < 0.001$. Age-group comparisons used Wilcoxon rank-sum tests, and small-versus-large comparisons used Wilcoxon signed-rank tests.

(Figures 5B–D, respectively). A quantitative assessment of weight variance showed a similar trend. The weight variance was significantly larger in older groups than in younger groups (Kruskal-Wallis test: $p < 0.001$, $\eta^2 = 0.639$; Wilcoxon rank-sum test: Adult vs. 14-yo, $p < 0.001$, $d = 1.81$; Adult vs. 11-yo, $p < 0.001$, $d = 2.14$; Adult vs. 8-yo, $p < 0.001$, $d = 2.52$; 14-yo vs. 11-yo, $p < 0.001$, $d = 1.11$; 14-yo vs. 8-yo, $p < 0.001$, $d = 2.94$; 11-yo vs. 8-yo, $p < 0.001$, $d = 1.66$; Figure 5E). Furthermore, the small problems (i.e., problems with operands $\leq$4) were more widely scattered compared to the large problems (Adult, $p < 0.001$, $d = 1.45$; 14-yo, $p < 0.001$, $d = 1.16$; 11-yo, $p < 0.001$, $d = 0.86$; 8-yo, $p < 0.001$, $d = 0.70$; Figure 5F), which is consistent with the prediction performance result (shown in Figure 3F). Similar patterns were also observed in other math-relevant regions, including the hippocampus and IPS (Figure S9). These results indicate that, for older groups, addition problems are represented as discrete categories in the brain. Such discrete representations could have contributed to prediction patterns in the discrete model (and hence in the ANN model too).

Next, we tested whether such discrete representations of addition problems spontaneously emerge using the ANN model without assuming discrete representations in the model. To this end, we reconstructed discrete model weights by multiplying the ANN features for each addition problem by the ANN model weights, resulting in highly similar reconstructed weights across

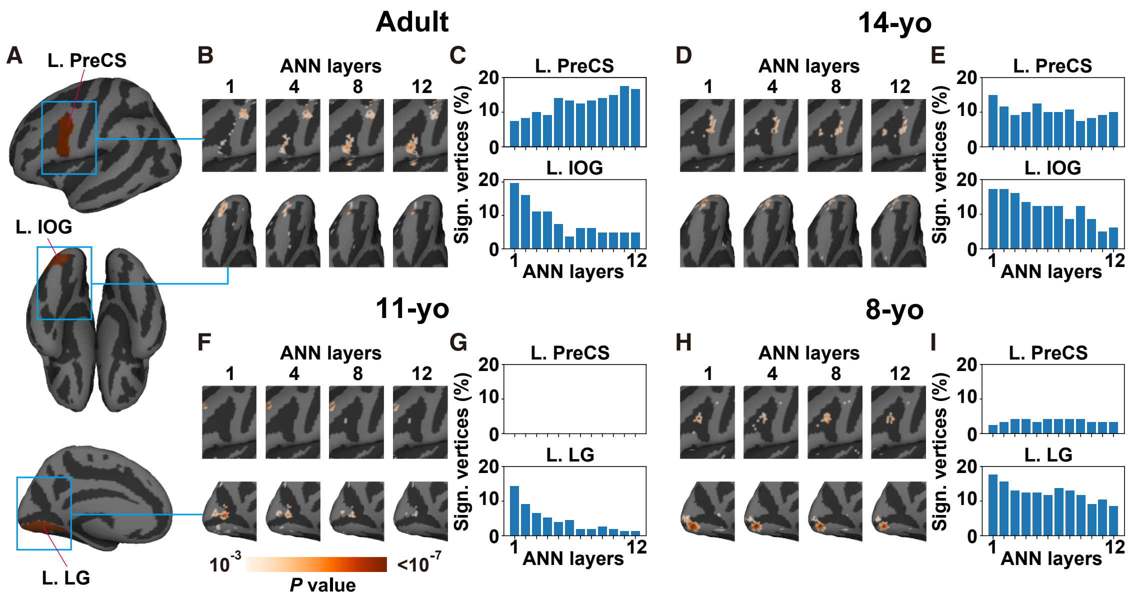


Figure 6. Topographic expansion and reduction of predictable brain regions with ANN layers across age groups

(A) The target anatomical regions-of-interest (in red) and the surrounding areas (in blue squares) of the left precentral sulcus (L. PreCS), left inferior occipital gyrus (L. IOG), and left lingual gyrus (L. LG) are shown on the cortical surface.

(B, D, F, and H) Prediction patterns are shown for the 1st, 4th, 8th, and 12th ANN layers. Only significantly predicted vertices are shown (peak-level, uncorrected $p < 0.001$; cluster-level, corrected $p < 0.05$). Areas around the L. PreCS and L. IOG are shown for (B) adults and (D) 14-yo, while those around the L. PreCS and L. LG are shown for (F) 11-yo and (H) 8-yo.

(C, E, G, and I) Bar graphs of the percent of significant vertices are plotted for all of the twelve ANN layers. Bar graphs of the L. PreCS and L. IOG are shown for (C) adults and (E) 14-yo, while those of the L. PreCS and L. LG are shown for (G) 11-yo and (I) 8-yo.

the cortex (Figures S10A and S10B). We then performed PCA using the reconstructed weight values. We again found that addition problems are most widely scattered in the two-dimensional space for adults than younger age groups (Figures S10C–S10F). The weight variance was significantly larger in older groups than in younger groups (Figure 5G), and the small problems were more widely scattered compared to the large problems (Figure 5H). These results indicate that addition problems are represented discretely within the ANN model, and such representations become more apparent with older age groups having more exposure to arithmetic problems.

ANN predicts arithmetic-related activity in a layer-dependent manner

A third analysis aiming to shed light on the reasons for an age-dependent increase in prediction accuracy of the ANN model involved a comparison of prediction performance across different intermediate ANN layers. Indeed, our previous study had revealed a stronger association between arithmetic-related activity and ANN features in deeper layers,[29] which is in line with a large number of studies investigating the association between ANNs and the human brain.[24–28] If arithmetic-related representations in older participants are more similar to the ANN model than in younger participants, such hierarchical correspondence would be more evident in older than younger participants.

To test this possibility, we extracted latent features from the shallowest layer (i.e., first hidden layer) to the deepest layer (twelfth hidden layer) and examined whether such layer-dependent prediction patterns depend on age groups. We observed a topographic expansion of significant vertices in the L. PreCS for adult participants (Figures 6A and 6B). That is, deeper ANN layers predicted larger brain regions for this age group. Such topographic expansion in the frontal cortex was absent in younger groups (Figures 6C–6I). In contrast, there was a topographic reduction of significant vertices in the left inferior occipital gyrus (L. IOG) in adults and 14-yo (Figures 6A–6E). The same pattern was observed in 11-yo and 8-yo, but in a different sub-portion of the occipital cortex (i.e., left lingual gyrus, L. LG) (Figures 6F–6I). In sum, prediction performance by the ANN model increased with intermediate layers in the left frontal region in adults and 14-yo and in the right parietal region in adults, 11-yo, and 8-yo. In contrast, it decreased in the occipital region in all age groups.

These observations suggest that between-age effect in the ANN model performance may depend on the target layers. Indeed, we found a significant main effect of age of prediction accuracy in the L. PreCS (Kruskal-Wallis test: peak level, uncorrected $p < 0.001$; cluster level, corrected $p < 0.05$) when using features extracted from the 8th layer (Figure S11), but no significant effect was found when using features extracted from the 4th and 1st layers. This suggests that age-related representational changes in arithmetic emerge primarily in deeper layers of the ANN model.

DISCUSSION

Models of arithmetic processing suggest that arithmetic learning involves an increase in memory-based retrieval of associations between operands of problems and their answers.[3] Because this process bears similarity with the way information is retrieved in ANNs (i.e., through retrieval of statistical associations between elements), we asked in the current study whether arithmetic-related brain activity would be increasingly captured by an ANN-based encoding model with age. Measuring fMRI activity associated with single-digit mental addition in participants from age 8 to adulthood, we found that this was the case in a region of the L. PreCS, particularly for the small problems. Further analyses showed that ANN predictions best aligned with a model in which representations of arithmetic problems were discrete, compared to models in which problems were categorized as varying with problem size or difficulty. PCA visualization revealed that addition problems were more discretely distributed in older than in younger participants, even when using discrete model weights reconstructed via ANN features. Finally, layer-dependent increase in prediction accuracy was primarily observed in the L. PreCS in adults. To some extent, these findings suggest that neural activity associated with mental arithmetic can be increasingly predicted by an ANN with learning and development. However, this was largely constrained to a specific area of the left frontal cortex, suggesting that ANN-derived features capture only a portion of the neural processes underlying mental arithmetic.

The developmental increase in similarity between arithmetic-related activity and predictions from an ANN in the L. PreCS is consistent with the idea that, with increasing age, arithmetic problem solving relies more on the retrieval of associations between operands and answers. This is in keeping with associative models of arithmetic learning. Indeed, these models assume that operands and answers of problems are progressively associated with one another through repeated exposure,[3,11] a property that can be captured by models based on statistical learning. Interestingly, because smaller problems tend to be more frequently encountered than larger problems,[34] associative models typically predict that associations between operands and answers decrease with problem size.[11] In line with this idea, we found that activity associated with the small problems (i.e., problems with operands $\leq$4) was better explained by the ANN-based model than large problems (i.e., problems with operands >4) in older participants.

On the one hand, the similarity between arithmetic-related activity and predictions from an ANN provides some support for the idea that solving single-digit addition problems in adults and older children may rely on retrieving associations between operands and answers from memory, much like an ANN may have learned to associate co-occurring elements in its training set. On the other hand, our ANN model was only able to increasingly predict arithmetic-related activity with age in a cluster of the left frontal cortex, located in the L. PreCS. In other words, we did not find an age-related increase in ANN prediction accuracy in most of the brain, including in regions that are classically associated with mental arithmetic such as the bilateral IPS, left AG, and hippocampus.[38,39]

A lack of similarity between ANN predictions and IPS activity may be explained by the fact that IPS-related activity in mental arithmetic is often interpreted as reflecting explicit manipulation of quantities rather than retrieval of associations.[40] We note that ANN prediction was observed in the right IPS in the 11-year-old group (Figure 2C), although this finding was isolated to a single age group and did not emerge in the between-group comparison of the L. PreCS. The interpretation of this observation remains unclear, and we refrain from drawing strong developmental conclusions from it. However, the lack of similarity with the left AG and the hippocampus is surprising, given the hypothesized role of these regions in memory-based retrieval of associations.[38,41–43] In fact, we even found a decrease in prediction accuracy with age using the ANN model in the right hippocampus (Figure S1).

We can see two possible explanations for why the age-related increase in ANN prediction may be restricted to the L. PreCS. First, it is possible that this region may play a more critical role in memory retrieval for overlearned problems than regions such as the hippocampus,[38] which may support encoding of associations between operands and answers in the early stages of arithmetic learning and more stable representation of mental arithmetic in later stages of development.[42] Indeed, in contrast to the hippocampus, activations in the left frontal cortex (Figures 2 and 3) have frequently been observed in arithmetic task fMRI. Coordinate-based meta-analyses indicate that activation of the L. PreCS/PreCG and neighboring L. IFG is frequently observed in tasks involving addition problem solving.[39,44] Activation in the L. IFG in arithmetic tasks also overlapped with sentence comprehension[45,46] and semantic memory retrieval tasks,[47] which may suggest that this region contributes to the retrieval of arithmetic facts from memory. Overall, these observations are in line with developmental neuroimaging studies reporting brain-behavior correlations between arithmetic skills and structural properties of left frontal areas,[48,49] as well as anatomical connectivity between the L. PreCS/IFG and IPS.[50] Note that our study identifies a cluster located in the L. PreCS rather than the L. IFG per se. However, because volume-based analysis may ignore information on cortical folding patterns and is less sensitive to local spatial patterns,[51,52] we adopt here a surface-based analysis.[53,54] Therefore, clusters in the dorsal and posterior parts of the IFG identified in previous volume-based analyses may be relatively close to the L. PreCS identified here.

There is a second possibility to explain why the ANN-based model accounts for arithmetic-related activity only to a limited extent. Under a purely associative account of arithmetic learning, one might expect ANN-derived features to broadly predict arithmetic-related activity in adults, including in regions such as the AG and hippocampus that have been specifically associated with memory-based retrieval of arithmetic facts.[38,41] The fact that the age-related increase in ANN prediction is restricted to the L. PreCS (and is even reversed in the right hippocampus, see Figure S1), suggests that associative retrieval may only partially account for arithmetic processing in adults. Indeed, a series of recent studies[6,55–60] suggest that counting procedures that may be used by young children when learning to add numbers may be accelerated up to the point that they become automatic and unconscious.[57] This may explain a range of behavioral and neuroimaging findings, including the presence of a problem size effect with small addition problems,[57] the difference between single-digit addition and single-digit multiplication,[56,61,62] and the recent findings that patterns of brain activity associated with the problem size effect change quantitatively but not qualitatively over development.[60] Critically, automatized counting is fundamentally different from the type of statistical learning at the core of an ANN, because it involves the rapid execution of a sequential procedure rather than the retrieval of a stored association. It has recently been proposed that automatized counting may coexist with retrieval of associations from memory.[63] Under such a dual-mechanism account, the restricted scope of ANN predictions in the present study would be expected. That is, ANN-derived features would capture the associative component of arithmetic processing, while the procedural component (reflecting automatized counting) would not.

Interestingly, among the four models implemented in this study, the difficulty model was the only one to show greater similarity with brain activity patterns in younger than older participants (in the primary motor cortex) (Figure S8). We speculate that activity in this region may reflect some finger-related processing supporting counting. For example, previous studies reported activity of the primary motor and premotor cortices during counting tasks.[64–66] Indeed, the difficulty model showed significant prediction in the dorsal part of the central sulcus (Figure S8), which is associated with finger movements rather than with the mouth and tongue.[67,68] The high predictive accuracy of this model for the younger groups, as well as a tendency of negative correlation between the ANN and difficulty model (Figure 4), is consistent with the fact that younger participants may frequently adopt silent counting strategies.[7–9,42,59] This pattern suggests that the difficulty model captures effortful, procedure-based strategies such as explicit counting, which are more prevalent in early stages of development and distinct from the associative retrieval processes partly captured by the ANN model. Note that under the automatized counting account,[63] the developmental trajectory would involve a transition from these overt, effortful counting procedures (reflected in the prediction of the difficulty model in the primary motor cortex in younger participants) to fast, automatized counting procedures that would no longer engage primary motor cortex but may still rely on sequential processing rather than memory retrieval.

To shed some light on the nature of the relation between the ANN model and brain activity, we performed several additional analyses comparing the ANN model to other models in which arithmetic problems were categorized in different ways. There was a positive correlation between prediction performance of the ANN model and prediction performance of a model in which addition problems were each represented separately, suggesting that the ANN model may represent (at least in part) each addition problem discretely (Figure 4). As we observed a highly positive correlation across the entire cortex, this effect is likely due to the shared features underlying the ANN and that "discrete model," rather than to the specific brain activity pattern of the L. PreCS. PCA results further support these findings. Increasing weight variance (Figure 5), as well as their two-dimensional visualization, indicated that older participants have more discrete representations of each addition problem in the brain. Moreover, the ANN-based reconstruction analysis indicates that such discrete representations can be obtained using the ANN model, even using continuous latent features of addition problems. These results suggest that the ANN-based model may better account for arithmetic-related activity in older participants to the extent that their neural representations are more discretely organized. The increased dispersion observed in PCA space may reflect increasingly differentiated representations of arithmetic problems. However, the interpretation of such sparsity in low-dimensional space remains an open question. While this pattern may be consistent with the strengthening of associations between operands and answers, it does not uniquely imply a specific cognitive mechanism or a region-specific arithmetic representation, and may arise from multiple representational changes. Importantly, these results do not imply that brain activity related to arithmetic can be fully accounted for by simple discrete models alone. Rather, the ANN model may provide a useful framework to characterize representational structure in a continuous feature space, complementing simpler models based on discrete problem representations.

Indeed, a model in which problems were categorized as a function of their size also showed a positive (though reduced) correlation with the ANN model (Figure 4), suggesting that the ANN model may contain some information about problem-size, which is well known to influence brain activity.[60] However, this correspondence was substantially weaker than that observed for the discrete model, indicating that problem size alone is insufficient to account for the neural representations captured by the ANN model. Nevertheless, although we did not observe a monotonic effect of age for the problem-size model, a significant group main effect was found in the right parietal cortex (Figure S7E). This suggests that problem-size-related processing may vary across developmental stages, particularly in parietal regions associated with quantity processing, even if it does not show a systematic increase or decrease with age.

The hierarchical prediction pattern in the left frontal and visual cortices (Figure 6) is in line with previous studies in visual neuroscience combining ANN and human data,[24–26] where deeper ANN layers showed an increase in prediction performance in higher-order visual cortex. Such patterns have also been observed in the auditory,[27,69] language,[28,32] and arithmetic processing in adults.[29] In arithmetic processing, input information is thought to ascend the brain's hierarchy from lower-level sensory areas to higher-level association cortex,[70,71] in line with the observed decrease in prediction in areas of the visual cortex and the increase in prediction in the left frontal cortex. This might indicate that deeper ANN layers hold higher-order information than simple visual features. More generally, sensory perception such as vision and audition and abstract cognitive functions such as language and mathematics may not follow exactly the same hierarchical pathways in the brain. However, the current study shows that there may be common principles in how ANN-derived features account for brain activity across these domains.

Limitations of the study

We note five limitations of the current study. First, our encoding models did not directly predict brain activity but rather beta estimates after estimating first-level GLM for each participant. Although this approach deviates from conventional encoding models using naturalistic stimuli (e.g., movie viewing task),[72,73] it enables us to minimize the impact of performance (e.g., response times) variations from trial to trial and to compare model performance across age groups using the same training and testing sample size. Nevertheless, because this method involves fitting the data in two stages, it might downplay information about the dynamic processing of arithmetic in the first stage. In addition, we used a standard GLM to estimate condition-level responses rather than single-trial approaches such as GLMsingle,[74] which may provide improved noise modeling. Future work could benefit from incorporating them.

Second, our task design contained only addition problems. Therefore, it remains unclear whether our results may apply to other arithmetic operations. Nonetheless, in a previous study, we have shown that an ANN encoding model may predict brain activity of addition, subtraction, multiplication, and division problems, as well as their combinations (e.g., "(3 + 2) × 4").[29]

Third, the ANN used in the current study (MathBERT) was trained with a very large dataset of mathematical problems, which contains not only simple arithmetic problems but also more complex algebra.[17] This arguably departs from the way children acquire simple arithmetic skills. Future studies using more realistic models based on actual learning materials (i.e., math textbooks) may allow for more accurate models that could potentially predict brain representations of arithmetic in children.

Fourth, the present study primarily focused on a single ANN architecture (MathBERT), with an additional convolutional neural network (ResNet-18) included as a control for low-level visual features, and did not systematically compare performance across a broader range of computational models. Although our control analysis using ResNet-18 suggests that the observed effects are not driven by visual properties of the stimuli, future work incorporating a wider variety of architectures, such as alternative language models or biologically inspired models, will be important to more precisely identify the computational properties that best account for arithmetic-related brain activity.

Finally, the ANN model used in the present study was not explicitly trained on the current task, as we extracted features from a pre-trained model. Although such models are trained to capture input-output relationships in large corpora, their internal representations may reflect statistical regularities that are broadly consistent with associative representations. Therefore, our use of ANN-derived features should be understood as testing whether brain representations exhibit similar patterns, rather than assuming a specific underlying mechanism. Also, the present results do not allow us to determine whether the ANN model performs calculation-like operations or relies on associative representations. Because the ANN-derived features were extracted from a pre-trained model, they may capture multiple aspects of arithmetic problems, including both structural and relational information. As a result, the observed correspondence between ANN predictions and brain activity does not imply a specific computational mechanism, but rather suggests that ANN representations share some representational properties with neural responses during arithmetic processing.

In summary, this study demonstrates that neural activity associated with addition problem solving can be increasingly captured by ANN-based models with learning and development, particularly in the L. PreCS. This pattern is likely to arise from increasingly discrete organization of addition problems in the brain, especially for smaller problems. On the one hand, this generally supports associative models of arithmetic learning (e.g., Ashcraft, 1992[3]). On the other hand, the limited correspondence between arithmetic-related activity and predictions from the ANN outside the precentral sulcus suggests that associative retrieval may only partially account for arithmetic processing in adults, leaving open the possibility that non-associative mechanisms such as automatized counting procedures[57,63] also contribute. This, in turn, highlights the need for integrative computational models that combine ANN-based associative representations with automatized counting mechanisms. These findings provide insights into the neural and computational foundations of mathematical cognition, suggesting that ANN-based models may characterize the associative component of arithmetic learning while highlighting the need for complementary computational approaches to capture the full range of mechanisms underlying arithmetic development.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Tomoya Nakai (tnakai@g.ecc.u-tokyo.ac.jp).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The analysis code is available from Zenodo (https://zenodo.org/records/15395718).
- The dataset that supports the findings of the current study will be available from the corresponding author upon reasonable request.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Franck Lamberton, Danielle Ibarrola, Thomas Bastelica, Jessica Léone, Justine Epinat, and Andrea Díaz-Barriga Yáñez for their help with data collection. We are also grateful to the children and parents who participated. This study was partially supported by JSPS KAKENHI (JP24H02172 and JP26H00539), JST FOREST Program (JPMJFR231V), and JST Moonshot R&D (JPMJMS259G) for T.N., and Agence Nationale de la Recherche (ANR-18-CE93-0001 and ANR-23-CE28-0002) for J.P. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

T.N.: conceptualization, formal analysis, visualization, writing – original draft preparation, and funding acquisition. J.P.: supervision, writing – review and editing, and funding acquisition.

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

- METHOD DETAILS
  - Experimental procedure
  - fMRI data acquisition
  - fMRI data preprocessing
  - Trial-based signal estimation
  - Feature extraction
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Vertex-wise encoding model fitting
  - Principal component analysis
  - Weight reconstruction
  - Prediction accuracy in other anatomical ROIs
  - Effect of tie problems and sum-to-10 problems
  - Noise ceiling and signal-to-noise ratio
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2026.117180.

Received: November 27, 2025
Revised: April 14, 2026
Accepted: July 28, 2026

REFERENCES

1. Mandelbaum, E., and Millière, R. (2025). Associationist theories of thought. In The Stanford Encyclopedia of Philosophy (Fall 2025 Edition), E.N. Zalta and U. Nodelman, eds. (Metaphysics Research Lab, Stanford University).
2. Anderson, J.R., and Bower, G.H. (2014). Human Associative Memory (Psychology Press). https://doi.org/10.4324/9781315802886.
3. Ashcraft, M.H. (1992). Cognitive arithmetic: a review of data and theory. Cognition *44*, 75–106. https://doi.org/10.1016/0010-0277(92)90051-i.
4. Lemaire, P., and Siegler, R.S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. J. Exp. Psychol. Gen. *124*, 83–97. https://doi.org/10.1037/0096-3445.124.1.83.
5. Geary, D.C., Hoard, M.K., and Nugent, L. (2012). Independent contributions of the central executive, intelligence, and in-class attentive behavior to developmental change in the strategies used to solve addition problems. J. Exp. Child Psychol. *113*, 49–65. https://doi.org/10.1016/j.jecp.2012.03.003.
6. Thevenot, C., Barrouillet, P., Castel, C., and Uittenhove, K. (2016). Ten-year-old children strategies in mental addition: A counting model account. Cognition *146*, 48–57. https://doi.org/10.1016/j.cognition.2015.09.003.
7. Wu, S.S., Meyer, M.L., Maeda, U., Salimpoor, V., Tomiyama, S., Geary, D.C., and Menon, V. (2008). Standardized Assessment of Strategy Use and Working Memory in Early Mental Arithmetic Performance. Dev. Neuropsychol. *33*, 365–393. https://doi.org/10.1080/87565640801982445.
8. Groen, G.J., and Parkman, J.M. (1972). A chronometric analysis of simple addition. Psychol. Rev. *79*, 329–343. https://doi.org/10.1037/H0032950.
9. Ashcraft, M.H. (1982). The development of mental arithmetic: A chronometric approach. Dev. Rev. *2*, 213–236. https://doi.org/10.1016/0273-2297(82)90012-0.
10. Logan, G.D. (1988). Toward an instance theory of automatization. Psychol. Rev. *95*, 492–527. https://doi.org/10.1037/0033-295X.95.4.492.
11. Chen, Y., and Campbell, J.I.D. (2018). "Compacted" procedures for adults' simple addition: A review and critique of the evidence. Psychon. Bull. Rev. *25*, 739–753. https://doi.org/10.3758/s13423-017-1328-2.
12. Bishop, C.M. (2006). Pattern Recognition and Machine Learning (Springer).
13. Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). Deep Learning (MIT press). https://doi.org/10.4258/hir.2016.22.4.351.
14. Matsumoto, D., and Nakai, T. (2023). Syntactic theory of mathematical expressions. Cogn. Psychol. *146*, 101606. https://doi.org/10.1016/j.cogpsych.2023.101606.
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems 30 (NIPS 2017). https://doi.org/10.48550/arXiv.1706.03762.
16. Schlag, I., Smolensky, P., Fernandez, R., Jojic, N., Schmidhuber, J., and Gao, J. (2019). Enhancing the Transformer with explicit relational encoding for math problem solving. Preprint at arXiv. https://doi.org/10.48550/arXiv.1910.06611.
17. Peng, S., Yuan, K., Gao, L., and Tang, Z. (2021). MathBERT: A pre-trained model for mathematical formula understanding. Preprint at arXiv. https://doi.org/10.48550/arXiv.2105.00377.
18. Azerbayev, Z., Schoelkopf, H., Paster, K., Dos Santos, M., McAleer, S., Jiang, A.Q., Deng, J., Biderman, S., and Welleck, S. (2024). Llemma: An open language model for mathematics. International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.2310.10631.
19. Zong, M., and Krishnamachari, B. (2023). Solving math word problems concerning systems of equations with GPT-3. Proc. Conf. AAAI Artif. Intell. *37*, 15972–15979. https://doi.org/10.1609/aaai.v37i13.26896.
20. Yang, A., Zhang, B., Hui, B., Gao, B., Yu, B., Li, C., Liu, D., Tu, J., Zhou, J., Lin, J., et al. (2024). Qwen2.5-math technical report: Toward mathematical expert model via self-improvement. Preprint at arXiv. https://doi.org/10.48550/arXiv.2409.12122.
21. Russin, J., Fernandez, R., Palangi, H., Rosen, E., Jojic, N., Smolensky, P., and Gao, J. (2021). Compositional processing emerges in neural networks solving math problems. CogSci *2021*, 1767–1773. https://doi.org/10.48550/arXiv.2105.08961.
22. Yuan, Z., Yuan, H., Tan, C., Wang, W., and Huang, S. (2023). How well do Large Language Models perform in Arithmetic tasks?. Preprint at arXiv. https://doi.org/10.48550/arXiv.2304.02015.
23. Zhang, H., Da, J., Lee, D., Robinson, V., Wu, C., Song, W., Zhao, T., Raja, P., Zhuang, C., Slack, D., et al. (2024). A careful examination of large language model performance on grade school arithmetic. Advances in Neural Information Processing Systems 37 (NeurIPS 2024). https://doi.org/10.48550/arXiv.2405.00332.
24. Güçlü, U., and van Gerven, M.A.J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. J. Neurosci. *35*, 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015.
25. Horikawa, T., and Kamitani, Y. (2017). Generic decoding of seen and imagined objects using hierarchical visual features. Nat. Commun. *8*, 15037. https://doi.org/10.1038/ncomms15037.
26. Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., and Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Sci. Rep. *6*, 27755. https://doi.org/10.1038/srep27755.
27. Kell, A.J.E., Yamins, D.L.K., Shook, E.N., Norman-Haignere, S.V., and McDermott, J.H. (2018). A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. Neuron *98*, 630–644.e16. https://doi.org/10.1016/j.neuron.2018.03.044.
28. Caucheteux, C., Gramfort, A., and King, J.-R. (2023). Evidence of a predictive coding hierarchy in the human brain listening to speech. Nat. Hum. Behav. *7*, 430–441. https://doi.org/10.1038/s41562-022-01516-2.
29. Nakai, T., and Nishimoto, S. (2023). Artificial neural network modelling of the neural population code underlying mathematical operations. Neuroimage *270*, 119980. https://doi.org/10.1016/j.neuroimage.2023.119980.
30. Naselaris, T., Kay, K.N., Nishimoto, S., and Gallant, J.L. (2011). Encoding and decoding in fMRI. Neuroimage *56*, 400–410. https://doi.org/10.1016/j.neuroimage.2010.07.073.

31. Khosla, M., Ngo, G.H., Jamison, K., Kuceyeski, A., and Sabuncu, M.R. (2021). Cortical response to naturalistic stimuli is largely predictable with deep neural networks. Sci. Adv. *7*, eabe7547. https://doi.org/10.1126/sciadv.abe7547.

32. Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S.A., Feder, A., Emanuel, D., Cohen, A., et al. (2022). Shared computational principles for language processing in humans and deep language models. Nat. Neurosci. *25*, 369–380. https://doi.org/10.1038/s41593-022-01026-4.

33. Zbrodoff, N.J. (1995). Why is 9 + 7 harder than 2 + 3? Strength and interference as explanations of the problem-size effect. Mem. Cognit. *23*, 689–700. https://doi.org/10.3758/bf03200922.

34. Hamann, M.S., and Ashcraft, M.H. (1986). Textbook presentations of the basic addition facts. Cogn. Instr. *3*, 173–202. https://doi.org/10.1207/s1532690xci0303_2.

35. Schoppe, O., Harper, N.S., Willmore, B.D.B., King, A.J., and Schnupp, J.W.H. (2016). Measuring the Performance of Neural Models. Front. Comput. Neurosci. *10*, 10. https://doi.org/10.3389/fncom.2016.00010.

36. Ras, G., Xie, N., Van Gerven, M., and Doran, D. (2022). Explainable deep learning: A field guide for the uninitiated. J. Artif. Intell. Res. *73*, 329–397. https://doi.org/10.1613/jair.1.13200.

37. van der Velden, B.H.M., Kuijf, H.J., Gilhuijs, K.G.A., and Viergever, M.A. (2022). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Med. Image Anal. *79*, 102470. https://doi.org/10.1016/j.media.2022.102470.

38. Sokolowski, H.M., Hawes, Z., and Ansari, D. (2023). The neural correlates of retrieval and procedural strategies in mental arithmetic: A functional neuroimaging meta-analysis. Hum. Brain Mapp. *44*, 229–244. https://doi.org/10.1002/hbm.26082.

39. Istomina, A., and Arsalidou, M. (2024). Add, subtract and multiply: Meta-analyses of brain correlates of arithmetic operations in children and adults. Dev. Cogn. Neurosci. *69*, 101419. https://doi.org/10.1016/j.dcn.2024.101419.

40. Zamarian, L., Ischebeck, A., and Delazer, M. (2009). Neuroscience of learning arithmetic–evidence from brain imaging studies. Neurosci. Biobehav. Rev. *33*, 909–925. https://doi.org/10.1016/j.neubiorev.2009.03.005.

41. Grabner, R.H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., and Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. Neuropsychologia *47*, 604–608. https://doi.org/10.1016/j.neuropsychologia.2008.10.013.

42. Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D.C., and Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. Nat. Neurosci. *17*, 1263–1269. https://doi.org/10.1038/nn.3788.

43. Klein, E., Willmes, K., Bieck, S.M., Bloechle, J., and Moeller, K. (2019). White matter neuro-plasticity in mental arithmetic: Changes in hippocampal connectivity following arithmetic drill training. Cortex *114*, 115–123. https://doi.org/10.1016/j.cortex.2018.05.017.

44. Arsalidou, M., and Taylor, M.J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. Neuroimage *54*, 2382–2393. https://doi.org/10.1016/j.neuroimage.2010.10.009.

45. Nakai, T., and Okanoya, K. (2018). Neural Evidence of Cross-domain Structural Interaction between Language and Arithmetic. Sci. Rep. *8*, 12873. https://doi.org/10.1038/s41598-018-31279-8.

46. Nakai, T., and Okanoya, K. (2020). Cortical collateralization induced by language and arithmetic in non-right-handers. Cortex *124*, 154–166. https://doi.org/10.1016/j.cortex.2019.11.009.

47. Heidekum, A.E., Grabner, R.H., De Smedt, B., De Visscher, A., and Vogel, S.E. (2019). Interference during the retrieval of arithmetic and lexico-semantic knowledge modulates similar brain regions: Evidence from functional magnetic resonance imaging (fMRI). Cortex *120*, 375–393. https://doi.org/10.1016/j.cortex.2019.06.007.

48. Rotzer, S., Kucian, K., Martin, E., von Aster, M., Klaver, P., and Loenneker, T. (2008). Optimized voxel-based morphometry in children with developmental dyscalculia. Neuroimage *39*, 417–422. https://doi.org/10.1016/j.neuroimage.2007.08.045.

49. Viesel-Nordmeyer, N., and Prado, J. (2023). Arithmetic skills are associated with left fronto-temporal gray matter volume in 536 children and adolescents. npj Sci. Learn. *8*, 56. https://doi.org/10.1038/s41539-023-00201-x.

50. Tsang, J.M., Dougherty, R.F., Deutsch, G.K., Wandell, B.A., and Ben-Shachar, M. (2009). Frontoparietal white matter diffusion properties predict mental arithmetic skills in children. Proc. Natl. Acad. Sci. USA *106*, 22546–22551. https://doi.org/10.1073/pnas.0906094106.

51. Oosterhof, N.N., Wiestler, T., Downing, P.E., and Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. Neuroimage *56*, 593–600. https://doi.org/10.1016/j.neuroimage.2010.04.270.

52. Tucholka, A., Fritsch, V., Poline, J.-B., and Thirion, B. (2012). An empirical comparison of surface-based and volume-based group studies in neuroimaging. Neuroimage *63*, 1443–1453. https://doi.org/10.1016/j.neuroimage.2012.06.019.

53. Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage *9*, 179–194. https://doi.org/10.1006/nimg.1998.0395.

54. Fischl, B., Sereno, M.I., and Dale, A.M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage *9*, 195–207. https://doi.org/10.1006/nimg.1998.0396.

55. Barrouillet, P., and Thevenot, C. (2013). On the problem-size effect in small additions: can we really discard any counting-based account? Cognition *128*, 35–44. https://doi.org/10.1016/j.cognition.2013.02.018.

56. Fayol, M., and Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. Cognition *123*, 392–403. https://doi.org/10.1016/j.cognition.2012.02.008.

57. Uittenhove, K., Thevenot, C., and Barrouillet, P. (2016). Fast automated counting procedures in addition problem solving: When are they used and why are they mistaken for retrieval? Cognition *146*, 289–303. https://doi.org/10.1016/j.cognition.2015.10.008.

58. Poletti, C., Perez, J.-F., Houillon, J.-C., Prado, J., and Thevenot, C. (2021). Priming effects of arithmetic signs in 10- to 15-year-old children. Br. J. Dev. Psychol. *39*, 380–392. https://doi.org/10.1111/bjdp.12363.

59. Poletti, C., Díaz-Barriga Yáñez, A., Prado, J., and Thevenot, C. (2023). The development of simple addition problem solving in children: Reliance on automatized counting or memory retrieval depends on both expertise and problem size. J. Exp. Child Psychol. *234*, 105710. https://doi.org/10.1016/j.jecp.2023.105710.

60. Díaz-Barriga Yáñez, A., Longo, L., Chesnokova, H., Poletti, C., Thevenot, C., and Prado, J. (2023). Neural evidence for procedural automatization during cognitive development: Intraparietal response to changes in very-small addition problem-size increases with age. Dev. Cogn. Neurosci. *64*, 101310. https://doi.org/10.1016/j.dcn.2023.101310.

61. Mathieu, R., Epinat-Duclos, J., Léone, J., Fayol, M., Thevenot, C., and Prado, J. (2018). Hippocampal spatial mechanisms relate to the development of arithmetic symbol processing in children. Dev. Cogn. Neurosci. *30*, 324–332. https://doi.org/10.1016/j.dcn.2017.06.001.

62. Mathieu, R., Epinat-Duclos, J., Sigovan, M., Breton, A., Cheylus, A., Fayol, M., Thevenot, C., and Prado, J. (2018). What's Behind a "+" Sign? Perceiving an Arithmetic Operator Recruits Brain Circuits for Spatial Orienting. Cereb. Cortex *28*, 1673–1684. https://doi.org/10.1093/cercor/bhx064.

63. Prado, J., and Thevenot, C. (2026). Counting on automatization: The automatized counting theory of mental arithmetic. Dev. Rev. *79*, 101252. https://doi.org/10.1016/j.dr.2026.101252.

64. Hinton, S.C., Harrington, D.L., Binder, J.R., Durgerian, S., and Rao, S.M. (2004). Neural systems supporting timing and chronometric counting: an

FMRI study. Brain Res. Cogn. Brain Res. *21*, 183–192. https://doi.org/10.1016/j.cogbrainres.2004.04.009.

65. Piazza, M., Mechelli, A., Price, C.J., and Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: an fMRI study. Brain Res. *1106*, 177–188. https://doi.org/10.1016/j.brainres.2006.05.104.
66. Zago, L., Petit, L., Mellet, E., Joliot, M., Mazoyer, B., and Tzourio-Mazoyer, N. (2010). Neural correlates of counting large numerosity. ZDM *42*, 569–577. https://doi.org/10.1007/s11858-010-0254-9.
67. Penfield, W., and Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain *60*, 389–443. https://doi.org/10.1093/BRAIN/60.4.389.
68. Gordon, E.M., Chauvin, R.J., Van, A.N., Rajesh, A., Nielsen, A., Newbold, D.J., Lynch, C.J., Seider, N.A., Krimmel, S.R., Scheidter, K.M., et al. (2023). A somato-cognitive action network alternates with effector regions in motor cortex. Nature *617*, 351–359. https://doi.org/10.1038/s41586-023-05964-2.
69. Tuckute, G., Feather, J., Boebinger, D., and McDermott, J.H. (2023). Many but not all deep neural network audio models capture brain responses and exhibit correspondence between model stages and brain regions. PLoS Biol. *21*, e3002366. https://doi.org/10.1371/journal.pbio.3002366.
70. Dehaene, S., Molko, N., Cohen, L., and Wilson, A.J. (2004). Arithmetic and the brain. Curr. Opin. Neurobiol. *14*, 218–224. https://doi.org/10.1016/j.conb.2004.03.008.
71. Vogel, S.E., and De Smedt, B. (2021). Developmental brain dynamics of numerical and arithmetic abilities. npj Sci. Learn. *6*, 22. https://doi.org/10.1038/s41539-021-00099-3.
72. Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J.L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. *21*, 1641–1646. https://doi.org/10.1016/j.cub.2011.08.031.
73. Huth, A.G., Nishimoto, S., Vu, A.T., and Gallant, J.L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron *76*, 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014.
74. Prince, J.S., Charest, I., Kurzawski, J.W., Pyles, J.A., Tarr, M.J., and Kay, K.N. (2022). Improving the accuracy of single-trial fMRI response estimates using GLMsingle. eLife *11*, e77599. https://doi.org/10.7554/eLife.77599.
75. Prado, J., Mutreja, R., and Booth, J.R. (2014). Developmental dissociation in the neural responses to simple multiplication and subtraction problems. Dev. Sci. *17*, 537–552. https://doi.org/10.1111/desc.12140.
76. Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., et al. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods *16*, 111–116. https://doi.org/10.1038/s41592-018-0235-4.
77. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., and Gee, J.C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage *54*, 2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025.
78. Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging *20*, 45–57. https://doi.org/10.1109/42.906424.
79. Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage *17*, 825–841. https://doi.org/10.1016/s1053-8119(02)91132-8.
80. Cox, R.W., and Hyde, J.S. (1997). Software tools for analysis and visualization of fMRI data. NMR Biomed. *10*, 171–178. https://doi.org/10.1002/(sici)1099-1492(199706/08)10:4/5%253C171::aid-nbm453%253E3.0.co;2-l.
81. Greve, D.N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. Neuroimage *48*, 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060.
82. Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., and Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. *8*, 14. https://doi.org/10.3389/fninf.2014.00014.
83. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
84. Gao, J.S., Huth, A.G., Lescroart, M.D., and Gallant, J.L. (2015). Pycortex: an interactive surface visualizer for fMRI. Front. Neuroinform. *9*, 23. https://doi.org/10.3389/fninf.2015.00023.
85. Aiken, L.R., and Williams, E.N. (1973). Response times in adding and multiplying single-digit numbers. Percept. Mot. Skills *37*, 3–13. https://doi.org/10.2466/pms.1973.37.1.3.
86. LeFevre, J.A., Shanahan, T., and DeStefano, D. (2004). The tie effect in simple arithmetic: An access-based account. Mem. Cognit. *32*, 1019–1031. https://doi.org/10.3758/BF03196878.
87. Sahani, M., and Linden, J. (2002). How Linear are Auditory Cortical Responses? In Advances in Neural Information Processing Systems 15 (NIPS 2002), pp. 125–132.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
fMRIPrep version 21.0.2	https://fmriprep.org	RRID:SCR_016216
Python version 3.8.8	http://www.python.org/	RRID:SCR_008394
Scikit-learn version 0.24.1	http://scikit-learn.org/	RRID:SCR_002577
Nilearn version 0.8.1	http://nilearn.github.io	RRID:SCR_001362
MathBERT	https://huggingface.co/tbs17/MathBERT	Hugging Face model: tbs17/MathBERT
Pycortex version 1.2.1	https://github.com/gallantlab/pycortex	GitHub repository: gallantlab/pycortex
Analysis code in this study	https://zenodo.org/records/15395718	https://doi.org/10.5281/zenodo.15395718

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

One hundred and forty-nine children, adolescents, and adults between the ages of 8 and 30 were recruited for the study (see also previous work[60]). The sample size was determined based on the single-digit multiplication and subtraction problem-solving in children.[75] Participants were contacted through advertisements on social media. They were divided into four age groups: 8- to 9-year-olds (hereafter 8-yo, $n = 41$), 11- to 12-year-olds (hereafter 11-yo, $n = 36$), 14- to 15-year-olds (hereafter 14-yo, $n = 30$), and adults over the age of 18 ($n = 42$). Ten participants were excluded from the final analyses because they did not complete the fMRI session (8-yo, $n = 4$; 11-yo, $n = 3$; 14-yo, $n = 3$). Fourteen participants were also excluded after fMRI preprocessing, either because of excessive head motion in the scanner (see criteria below) or poor image quality (8-yo, $n = 8$; 11-yo, $n = 2$; 14-yo, $n = 1$; adults, $n = 3$). Finally, in order to match the number of participants in the four age groups, we randomly selected 26 participants (the minimum group size across the four groups) from each group. The final sample ($n = 104$) consisted of 26 participants in each of the four groups.

All participants were native French speakers. Participants below the age of 18 gave their assent to participate in the study, while their parents gave written informed consent. Adult participants gave written informed consent. The study was approved by a French ethics committee (Comité de Protection des Personnes Est 4, IDRCB 2019-A01918–49). Participants (or parents if participants were below 18) were paid 40 euros per session for their participation.

METHOD DETAILS

Experimental procedure

The study consisted of two separate testing sessions: one behavioral session (which included psychometric and behavioral testing, as well as mock MRI scanning for children and adolescents) and one fMRI session.

In the first testing session, children and adolescents were familiarized with the fMRI environment in a mock scanner. They listened to a recording of the noises associated with all fMRI sequences. A motion tracker system (3D Guidance track STAR, Ascension Technology Corporation) was used to measure head movements and provide online feedback to the participants. Finally, participants practiced 25 trials of the silent arithmetic task in that mock scanner.

In the second testing session, participants performed a mental addition task in the MRI scanner. The task consisted of all 45 single-digit addition problems with a sum less than or equal to 10. Problems were presented in both commutative orders (e.g., 2 + 3 and 3 + 2). Each addition problem was presented five times over five consecutive runs to maximize signal-to-noise ratio in the fMRI scanner, for a total of 225 trials. In each trial, an addition problem (which remained on the screen until a response was detected) was preceded by a red square for a fixed duration of 1,000 ms and followed by a variable period of fixation (i.e., a white square) from 2,000 to 4,000 ms. Participants were instructed to not say the answer to each problem out loud but rather to press a button with their right hand as soon as they came up with the answer in their head. To ensure that participants performed mental calculations, some problems were immediately followed by an exclamation mark on the screen for 1,000 ms. In such cases, participants had to quickly say the answer out loud before the disappearance of the exclamation mark. Answers were recorded manually by the experimenter. There were four arithmetic problems of such kind for every run of 45 trials. These trials with vocal answers were excluded from the fMRI analyses. For further information on the behavioral data, see our previous work.[60]

Trial order within each run and run order was randomized across the participants. Trials with vocal response were pseudo-randomized in the trial list, such that these trials were not located at the beginning of a run and were not consecutive. The task was programmed and presented with PsychoPy3 (v2020.2.5). Visual stimuli were projected on an in-scanner display by a projector in the room adjacent to the scanner. Head movement was minimized by cushions placed around the participant's head.

fMRI data acquisition

Images were collected using a Siemens Prisma 3 T MRI scanner with a 64-channel receiver head-neck coil (Siemens Healthcare, Erlangen, Germany) at the CERMEP Imagerie du vivant in Lyon, France. The blood oxygenation level-dependent (BOLD) signal was measured with a susceptibility-weighted single-shot echo planar imaging sequence. Imaging parameters were as follows: repetition time (TR) = 2,000 ms, echo time (TE) = 24 ms, flip angle = 80°, field of view (FOV) = 220 $\times$ 206 mm^2, resolution = 1.72 $\times$ 1.72 mm^2, slice thickness = 3 mm (0.48 mm gap), number of slices = 32. A high-resolution T1-weighted whole-brain anatomical volume was also collected for each participant. Parameters were as follows: TR = 2,400 ms, TE = 2.81 ms, flip angle = 8°, FOV = 224 $\times$ 256 mm^2, resolution = 1.0 $\times$ 1.0 mm^2, slice thickness = 1.0 mm, number of slices = 192.

fMRI data preprocessing

The preprocessing of both functional and anatomical MRI data was carried out using fMRIPrep 21.0.2.[76] The T1-weighted image underwent a correction for intensity non-uniformity using ANTs 2.3.3 N4BiasFieldCorrection[77] and was skull-stripped using a Nipype implementation of the antsBrainExtraction.sh workflow. Brain tissue segmentation was conducted using fast (FSL 6.0.5.1).[78] Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1).[53] Volume-based spatial normalization to the standard space (MNI152NLin6Asym) was achieved through nonlinear registration with antsRegistration (ANTs 2.3.3).

For the functional scans, a reference volume and its skull-stripped version were initially produced using the custom methodology of fMRIPrep. Prior to spatiotemporal filtering, head-motion parameters, including six rotation and translation parameters, were estimated in relation to the reference scan using mcflirt (FSL 6.0.5.1).[79] All functional scans underwent slice-time correction to the middle slice using 3dTshift from AFNI.[80] The reference scan was subsequently co-registered with the T1w reference using bbregister (FreeSurfer).[81] Three global signals were extracted within the cerebrospinal fluid (CSF), white matter (WM), and whole-brain masks. The blood oxygenation level-dependent time-series were resampled into the standard spaces (MNI152NLin6Asym). Runs in which translations greater than 2 mm or rotations greater than 2 radians were detected in more than 5% of volumes were excluded from the analysis (8-yo, n = 3; 11-yo, n = 2; 14-yo, n = 1 of runs). Mean head motion (after exclusion of high-motion runs, translation: x/y/z in mm; rotation: pitch/roll/yaw in radians) was 0.057/0.057/0.111 mm and 0.002/0.001/0.001 for adults, 0.030/0.052/0.095 mm and 0.002/0.001/0.001 rad for 14-yo, 0.035/0.064/0.132 mm and 0.002/0.001/0.001 rad for 11-yo, and 0.050/0.098/0.265 mm and 0.005/0.002/0.001 rad for 8-yo. Participants having less than three available runs were excluded (see Participants subsection).

The resulting functional data underwent high-pass filtering at 0.01 Hz and low-pass filtering at 0.15 Hz. Confounding factors, including the six head-motion parameters, global signals from the CSF, WM, and whole-brain masks, as well as the derivative of each factor, were eliminated using linear regression implemented in Scikit-learn. The functional data were then standardized to have a zero mean and unit variance for each of the 18,715 cortical vertices in the fsaverage5 space, excluding non-cortical vertices. Spatial smoothing was applied using a Gaussian filter with 8 mm full-width at half-maximum based on the geodesic distance on the cortical surface.

Trial-based signal estimation

First-level individual analysis was performed based on the preprocessed fMRI data using the Nilearn package in Python.[82] Univariate activity (i.e., beta value) associated with the 45 mental addition problems was analyzed using a general linear model, where the fMRI signal for each problem was modeled as an epoch of the same length as the reaction time, starting with the presentation of the problem, convolved with a hemodynamic response function. In addition, we included the following regressors of non-interest: a button response regressor with a duration of 0.5 s starting with each button response, three drift regressors describing low-frequency signals, regressors representing experimental runs, and a constant regressor. Forty-five beta series representing brain activity associated with each arithmetic problem were obtained for each of the 104 participants (i.e., 26 participants for each of the four age groups).

Feature extraction

Latent ANN features were extracted using a pre-trained large language model specifically tailored for understanding mathematical formulas, MathBERT (https://huggingface.co/tbs17/MathBERT).[17] Each addition problem was used as input, and intermediate outputs from the first to the twelfth layers (out of a total of twelve layers) were extracted as features. These resultant series of features were further reduced to 100 dimensions using PCA.

Latent features from ANN models were compared to encoding models that incorporated three features that represented different aspects of the problems: 'discrete' features consisting of 45 one-hot vectors (i.e., 45 dimensions) representing all 45 addition problems as distinct categories, a 'problem-size' feature corresponding to the sum of addition problems, and a 'difficulty' feature corresponding to reaction times of addition problems (averaged across all runs for each participant).

In addition, to control for low-level visual properties of the stimuli, we extracted visual features using a pre-trained convolutional neural network (ResNet-18). The input to the network consisted of images of the arithmetic problems as presented during the experiment. We extracted feature maps from the first residual block and computed the spatial average across each channel, resulting in a 64-dimensional feature vector for each problem.

QUANTIFICATION AND STATISTICAL ANALYSIS

Vertex-wise encoding model fitting

In the encoding model, each vertex's cortical activity (beta series concatenated across participants, denoted as R of dimensions [T $\times$ V]) was modeled by multiplying the feature matrix F of dimensions [T $\times$ N] by the weight matrix W of dimensions [N $\times$ V] (where T = number of samples; N = number of features; V = number of vertices):

$$\hat{R} = FW$$

Beta-series estimates were used to minimize the impact of variability in response times across trials and participants, thereby enabling more comparable model fitting across age groups. Model performance was evaluated using a nested leave-one-out cross-validation across participants. The weight matrix W was obtained using L2-regularized linear regression with the training data of 25 participants, which consisted of 1,125 samples. The optimal regularization parameter was determined using 10-fold cross-validation in the inner loop, where the regularization parameters varied from 1 to 10^6.

The test data of the left-out participant consisted of 45 samples for each outer loop. The prediction accuracy was evaluated using Pearson's correlation coefficient between the predicted and actual test samples. The peak-level statistical significance was calculated using a one-sided Wilcoxon signed-rank test ($p < 0.001$). A null distribution of cluster size was assessed with 1,000 permutation samples, and cluster-level significance was calculated based on this null distribution ($p < 0.05$), corrected for multiple comparisons using the false-discovery rate (FDR).[83] Pycortex was used to visualize data on cortical maps.[84]

Age-group comparisons were performed using Kruskal–Wallis tests followed by Wilcoxon rank-sum tests, whereas comparisons between small and large problems within each age group were performed using Wilcoxon signed-rank tests. Effect sizes were quantified using η^2 and Cohen's d. Boxplots show the median and interquartile range; whiskers extend to the most extreme data points within 1.5 times the interquartile range, and points beyond the whiskers represent outliers.

Principal component analysis

Brain representations for the 45 addition problems were visualized using PCA. For this analysis and the following weight reconstruction analysis, we calculated a weight matrix using data from all participants (without leaving out one participant) into model training for each age group. To control the effect of varying scales of the weight matrices, we selected the same regularization parameter across all participant groups (i.e., the mode of the four age groups). We applied PCA to the [45*V] weight matrices (V = number of vertices, concatenated across four groups) of the encoding model. To show the structure of the representational space of addition problems, 45 problems were mapped onto the two-dimensional space using the loadings of the first and second PCs as the x axis and y axis, respectively.

Weight reconstruction

To bridge the discrete model and ANN model weights, we reconstructed weight values of 45 addition problems of the discrete model using the ANN model weights. We first extracted latent ANN features of each addition problem (here denoted as *reference feature vectors*) and multiplied these features with the weight vectors obtained from the ANN encoding model. This analysis highlights how well the latent features can mediate between categorical concepts (addition problems in this case) and the brain. As a result, we obtained a 45-dimensional vector (i.e., reconstructed weights) in each vertex, which can be interpreted as a reconstruction of the discrete model weights via ANN latent features. We then performed a PCA using these reconstructed weights.

Prediction accuracy in other anatomical ROIs

We evaluated between-group differences in other anatomical ROIs that have been frequently found implicated in previous studies of mental arithmetic in adults and children.[38,39] These included the left inferior frontal gyrus (IFG), the hippocampus, the angular gyrus (AG), and the intraparietal sulcus (IPS). In the opercular part of the left inferior frontal gyrus (IFG), we found significantly higher prediction accuracy for adult participants than the other groups (Kruskal-Wallis test: $p < 0.001$, $\eta^2 = 0.162$; Wilcoxon rank-sum test: Adult vs. 14-yo, $p < 0.001$, $d = 1.00$; Adult vs. 11-yo, $p < 0.001$, $d = 1.12$; Adult vs. 8-yo, $p = 0.006$, $d = 0.75$; Figure S1A). In contrast, we found a significant main effect of age and larger prediction accuracy for 11-yo than adults in the right hippocampus (Kruskal-Wallis test: $p = 0.035$, $\eta^2 = 0.049$; 11-yo vs. Adult, $p = 0.005$, $d = 0.76$; Figure S1B), but no significant main effect was found in the left hippocampus ($p = 0.071$, $\eta^2 = 0.065$; Figure S1C). In the right AG, we found a significant main effect of age groups, but post-hoc tests did not show a monotonic increase with age (Kruskal-Wallis test: $p = 0.019$, $\eta^2 = 0.081$; Wilcoxon rank-sum test: adults vs. 14-yo, $p = 0.0049$, $d = 0.72$; 11-yo vs. 14-yo, $p = 0.0034$, $d = 0.70$; Figure S1E). No clear between-group difference was observed in the left AG ($p = 0.1436$, $\eta^2 = 0.044$; Figure S1D), left IPS ($p = 0.594$, $\eta^2 = 0.012$; Figure S1F), and right IPS ($p = 0.237$, $\eta^2 = 0.075$; Figure S1G). Overall, these results indicate that the age-related increase in similarity with the ANN is primarily seen in left frontal areas in older participants, though a similarity between the ANN and the right hippocampus may also be seen earlier in development.

Effect of tie problems and sum-to-10 problems

In addition to small and large problems, previous studies have reported faster reaction times for tie problems (e.g., "3 + 3") and sum-to-10 problems (e.g., "3 + 7") compared to other addition problems.[8,85] Indeed, these problems may have a special status in memory.[11,86] We therefore tested whether ties and sum-to-10 problems demonstrate better prediction patterns compared to other problems using the ANN model. We did not find any significant difference for these problems compared to other problems in any groups (Figures S12A–S12D). In the anatomical L. PreCS ROI, no age group showed larger prediction accuracy in ties and sum-to-10 problems than other problems (Wilcoxon signed-rank test: $p > 0.227$ for all age groups; Figure S12E). These results indicate that the ANN model does not show better prediction performance for activity associated with ties and sum-to-10 problems compared to other problems.

Noise ceiling and signal-to-noise ratio

To compare the signal-to-noise level of different age groups, we applied a *noise ceiling* calculation.[35] Due to a need for repeated signals for this calculation, we divided the original fMRI data into two subsets: the first to second runs, and the third to fourth runs (the fifth run was not used in this analysis). A general linear model was fitted for each of these subsets, resulting in two sets of beta series for each subject.

The noise ceiling was calculated using the formula introduced in a previous study.[87] First, the variance of repetitive neural responses (i.e., total power; *TP*) is split into estimators of signal power (*SP*) and noise power (*NP*):

$$TP = \frac{1}{N}\sum_{n} Var(r_n)$$

$$SP = \frac{1}{N-1}\left(N \times Var\left(\frac{1}{N}\sum_{n} r_n\right) - TP\right)$$

$$NP = TP - SP$$

Where N is the total number of repetitions (=2 in the current study), r_n is a time series of neural responses at the nth repetition. Based on these estimators, the noise ceiling value *NC* (maximum prediction accuracy) can be calculated as follows:

$$NC = \sqrt{\frac{SP}{SP + \frac{NP}{N}}}$$

In the present study, we further calculated the signal-to-noise ratio (*SNR*) by dividing the signal power by the noise power:

$$SNR = \frac{SP}{NP}$$

ADDITIONAL RESOURCES

This study did not create any additional resources.