JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles | Behavioral/Cognitive

Uncovering latent cognitive, metacognitive, and neural bases of problem-solving deficits in children with developmental dyscalculia

https://doi.org/10.1523/JNEUROSCI.2200-25.2026

Received: 21 November 2025
Revised: 15 July 2026
Accepted: 18 July 2026

Copyright © 2026 Lasnick et al.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

This Early Release article has been peer reviewed and accepted, but has not been through the composition and copyediting processes.The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Uncovering latent cognitive, metacognitive, and neural bases of problem-solving deficits in children with developmental dyscalculia

Running title: Latent problem-solving mechanisms in dyscalculia

O.H.M. Lasnick[1^*], Yunji Park[1^], Percy Mistry[1], Vinod Menon[1,2,3,4*]

[1]Department of Psychiatry & Behavioral Sciences,

[2]Department of Neurology & Neurological Sciences,

[3]Stanford Neurosciences Institute,

[4]Symbolic Systems Program,

Stanford University, Stanford, California, CA, 94305, US

Address: 401 Quarry Rd., Stanford, CA 94305, US

*Corresponding author(s) email: ohml@stanford.edu; menon@stanford.edu

^These authors contributed equally to this work

Revision Number: R2

Number of pages (excluding abstract): 16

Number of figures: 6

Number of tables: 7

Number of words for abstract: 250

Number of words for introduction: 667

Number of words for discussion: 1500

Conflicts of Interest Statement: The authors declare no competing financial interests.

Acknowledgements: This research was supported by grants from the National Institutes of Health (HD094623, HD059205) and National Science Foundation (DRL-2024856) to V.M and the National Institutes of Health (K99HD118172) to Y.P. We thank Daniel Ansari, Fu Yu Kwok, Eric D. Wilkey, Lien Peters, Ellyn Khiu, Rebecca Bull, and

JNeurosci Accepted Manuscript

Kerry Lee for generously sharing their data. All neuroimaging and behavioral data is available at OpenNeuro (doi: 10.18112/openneuro.ds004791.v1.0.0, https://openneuro.org/datasets/ds004791/versions/1.0.0). The authors declare no usage of generative AI in the writing process.

JNeurosci Accepted Manuscript

Abstract

Developmental dyscalculia (DD) affects 7-10% of children and is characterized by persistent difficulties in arithmetic problem-solving, which depends on the dynamic interplay between cognitive and metacognitive control processes governing strategy execution and selection. Conventional behavioral and neuroimaging approaches do not simultaneously quantify these latent processes and link them to distributed neural mechanisms, leaving the neurocognitive and metacognitive underpinnings of DD insufficiently characterized. To address this gap, we integrated Bayesian computational modeling with whole-brain decoding to uncover the cognitive, metacognitive, and neural bases of problem-solving deficits in DD using a sample of 68 children aged 8-10 years (38 female). Children with DD showed deficits across all problem-solving strategies (counting, retrieval, and decomposition), indicating reduced strategy execution efficiency. They also exhibited significant metacognitive impairments: longer baseline switching times between strategies, reduced sensitivity to problem difficulty, and lower adaptivity in selecting optimal strategies. Whole-brain ElasticNet decoding accurately distinguished DD from TA peers based on distributed activation patterns across frontoparietal cortex, hippocampus, and ventral-visual regions, whereas univariate approaches did not. Linear multivariate brain-behavioral analyses and nonlinear transformer-based predictions converged in showing that these distributed signatures predicted counting efficiency and switching sensitivity. Notably, brain-behavior relations reversed in direction between groups, positive in TA children but negative in DD, most strikingly in the hippocampus. These findings demonstrate that arithmetic difficulties in DD arise from deficits in both strategy execution efficiency and metacognitive control, reflecting an altered mapping between distributed neural activity, and highlight the power of integrating computational modeling with brain-wide decoding to understand learning disabilities.

JNeurosci Accepted Manuscript

Significance Statement

Developmental dyscalculia affects 7-10% of children, yet underlying mechanisms remain poorly understood. We integrated Bayesian computational modeling with whole-brain decoding to demonstrate that children with dyscalculia exhibit deficits in both strategy execution efficiency (counting, retrieval, decomposition) and metacognitive control (strategy switching, adaptive strategy selection). Using brain-wide decoding, we identified distributed neural signatures across frontoparietal, hippocampal, and visual regions that distinguished dyscalculia from typically-developing children, patterns undetectable by conventional approaches. These neural signatures predicted strategy selection and execution. Our findings demonstrate that developmental dyscalculia reflects cognitive and neural dysfunction across multiple processes, highlighting the power of combining computational modeling with multivariate neuroimaging analysis for understanding individual variability in learning disabilities.

JNeurosci Accepted Manuscript

Introduction

Developmental dyscalculia (DD) affects 7-10% of school-age children and is characterized by persistent deficits in arithmetic problem-solving that significantly impair academic achievement and daily functioning (Barbaresi et al., 2005; Geary, 2011). Successful problem-solving depends on the dynamic interplay between cognitive processes, such as executing retrieval or backup procedural strategies, and metacognitive control processes that guide adaptive strategy selection and flexible switching across trials (Siegler, 1996, 2007; Siegler et al., 1996). Yet the cognitive, metacognitive, and neural sources of the problem-solving difficulties in children with DD remain insufficiently understood.

Behavioral studies based on self-reported and observational strategy assessments have documented that children with DD draw on the same strategy repertoire as typically-achieving (TA) peers: counting, fact retrieval, and decomposition (e.g., solving 3+8 by reformulating as [2+8]+1). However, they tend to make more errors in executing these strategies (Russell & Ginsburg, 1984; Geary, 1992; Geary et al., 2012). Beyond strategy execution, children with DD rely more heavily on inefficient backup counting strategies (Ostad, 1997; Geary et al., 2004; Geary et al., 2012) and are less adaptive in strategy selection (Torbeyns et al., 2004; Peters et al., 2013), implicating impairments in the top-down process governing adaptive and flexible strategy selection across trials. Such processes reflect metacognitive control, the regulation of cognition through goal-setting and planning as well as metacognitive monitoring, the evaluation of one's own performance (Nelson & Narens, 1990; Ohtani & Hisasaka, 2018).

These approaches have yielded valuable insights into strategy-level deficits in DD. However, self-reports and aggregate behavioral measures are not designed to quantify the latent cognitive and metacognitive control processes that unfold trial-by-trial, leaving the dynamic mechanisms driving performance differences poorly characterized. Only a handful of computational models have begun to uncover these latent processes, and primarily in TA children (Siegler & Shipley, 1995; Supekar et al., 2021; Mistry et al., 2025).

Neuroimaging studies have reported atypical functional activation in children with DD during arithmetic problem-solving (Ashkenazi et al., 2012; Rosenberg-Lee et al., 2015; Peters & De Smedt, 2017), yet interpretation remains constrained by two gaps. First, few studies have directly modeled children's in-scanner problem-solving strategies, making it unclear whether atypical activation reflects disruptions in strategy execution or selection. Second, most studies rely on traditional univariate analyses that emphasize regional activity while overlooking distributed network coordination. Both limitations are underscored by Kwok et al. (2023), who observed clear behavioral impairments, but no atypical functional activation, in a well-characterized DD sample using conventional methods, highlighting the need for more sensitive and deeper computational approaches.

To address these gaps, we integrated advanced Bayesian computational modeling with multivariate neuroimaging to link specific cognitive and metacognitive control processes to distributed neural activity, uncovering the mechanisms underlying arithmetic deficits in children with DD. We analyzed a publicly available dataset from Kwok et al. (2023), comprising a relatively large sample of 65 children (30 DD, 35 TA) diagnosed through systematic screening by the Singaporean Ministry of Education (see **Materials and Methods**), affording greater statistical power than typical neuroimaging studies of DD.

JNeurosci Accepted Manuscript

Our first goal was to quantify the latent processes governing strategy execution and selection in children with DD using the Bayesian Model of Arithmetic Problem-Solving (BMAPS; **Figure 1**) (Supekar et al., 2021; Mistry et al., 2025). BMAPS estimates trial-by-trial probabilities of using counting, retrieval, and decomposition strategies, along with execution efficiency based on accuracy and reaction time patterns, and yields inter-trial metacognitive control parameters governing strategy selection, including switching time, switching sensitivity, and adaptivity. Our second goal was to identify distributed neural patterns that distinguish children with DD from TA children, using Elastic Net classification, a multivariate machine learning approach well-suited for identifying distributed neural signatures (Zou & Hastie, 2005). Finally, employing a novel AI-based transformer model, we tested whether these distributed neural signatures predict the latent cognitive and metacognitive processes captured by BMAPS parameters, linking brain function to the computational mechanisms underlying strategy execution and selection in DD. Together, this integrated approach offers a new mechanistic account of the sources of numerical problem-solving deficits and individual variability in children with DD.

Materials and Methods

Participants

We used the publicly-available dataset from Kwok et al. (2023) on OpenNeuro (doi: 10.18112/openneuro.ds004791.v1.0.0). The original sample included 68 children aged 8-10 years (M = 8.95 years, SD = 0.34; 38 female). Of these, 30 children were classified as having developmental dyscalculia (DD) based on the following criteria: they were either (a) identified by Singapore's Ministry of Education as having difficulty in math through screening at the beginning of primary school and subsequently referred to an intervention program, or (b) scored in the bottom 10th percentile of the Test of Early Mathematics Ability (TEMA-3). Children who scored above the 25th percentile on the TEMA-3 were classified as typically achieving (TA, N = 38). We excluded three additional participants who were missing all data from either the arithmetic verification task or the matching task, resulting in a sample of 65 children (36 female; 28 with DD and 37 TA) for traditional behavioral and neuroimaging analyses. An additional two participants were missing one run from the arithmetic verification task, and these two were excluded from the BMAPS analyses, as the model could not be run with inconsistent data between participants. This resulted in 63 participants for BMAPS model parameter estimation (35 female, 27 DD, 36 TA). Additional details regarding sample characteristics and group classification are provided in Kwok et al. (2023).

Experimental design and statistical analyses

Arithmetic verification task. In this task, participants were presented with an addition equation comprised of two single-digit addends and a proposed solution that could be either correct or incorrect (e.g., 5+3=7, incorrect trial). They were instructed to determine whether the presented equation was correct or incorrect. Half of the trials presented a correct sum and the remaining half presented an incorrect sum that exceeded the true answer by 1 or 2. One-third of trials were considered "Large" problems (solution greater than 10), one-third were "Small" problems (solution less than or equal to 10), and one-third were "Plus1" problems, where one addend equaled 1 (e.g., 2+1=3). Each run consisted of 36 trials presented in block design, with two runs presented to each participant.

JNeurosci Accepted Manuscript

Matching task. In this task, participants were presented simultaneously with two stimuli side-by-side and instructed to determine whether the two stimuli matched or not. Half of the trials had matching stimuli, and the other half had non-matching stimuli. There were three types of stimuli: numbers, shapes, or faces. In the "Number" condition (one-third of trials), children were presented with a symbolic number and a set of dots and asked to judge whether the quantities the two stimuli represented were the same. Incorrect trial stimuli differed from one another in quantity by two, three, or four (e.g., a set of six dots could be presented with the digit "3"). In the "Shape" condition (one-third of trials), they were presented with two regular geometric shapes (e.g., a circle and a square) and asked to judge whether the shapes were the same. There were nine total shapes. In the "Faces" condition (one-third of trials), they were presented with two front-facing human faces derived from 3D models and asked to judge whether the faces were the same. There were nine total "base" faces. For incorrect trials, one of the base faces was paired with an altered version of itself, with either small or large differences. Each run consisted of 54 trials presented in block design, with two runs presented to each participant.

Behavioral analysis. To compare differences in task performance between the DD and TA groups cross different conditions, we conducted independent samples *t*-tests for both accuracy and RT separately in both tasks (arithmetic, matching) and across all task conditions separately (**Supplementary Figure S1**). All behavioral analyses were conducted using a Bayesian statistical framework as implemented in JASP software v0.19.3 (https://jasp-stats.org/) to quantify the strength of evidence for group differences and accurately assess the presence or absence of effects across task conditions. We expected that the DDs would have lower accuracy and higher RT than the TAs for all arithmetic conditions (especially the Large problems), and for the number-matching condition.

Statistical significance was determined using Bayes factors (BF_{10}) which allow evaluation of the presence or absence of evidence for the alternative hypothesis (Keysers et al., 2020). BF_{10} values exceeding 3 indicate evidence for the test hypothesis, while values between 0.33 and 3 suggest no decisive evidence, and values lower than 0.33 indicate evidence for the null hypothesis. The strength of evidence also increases with BF_{10}: values 3-10 indicate moderate evidence, 10-30 strong evidence; 30-100 very strong evidence, and values above 100 indicate extreme evidence (Faulkenberry et al., 2020). Effect size (δ) was also estimated from the posterior probability distribution as the difference between the group means divided by the standard deviation, analogous to Cohen's *d*.

Bayesian model of arithmetic problem-solving and strategy-selection. To uncover cognitive variability underlying arithmetic problem solving in children with DD, we employed a recently developed Bayesian model for arithmetic problem-solving (Mistry et al., 2025) (BMAPS; **Figure 1**) to model in-scanner arithmetic task performance. This model infers the likelihood of using counting, retrieval, and *decomposition* (i.e., breaking down addends into smaller components to simplify problem-solving, for example solving 3+8 by splitting 3 into 1 and 2 to create a base-10 unit from 2+8) strategies on a trial-by-trial basis (**Supplementary Tables S1-S2**) by using the joint distribution of accuracy and RT, while accounting for the variability in item difficulty across trials.

BMAPS estimates individual-level cognitive and metacognitive control parameters related to strategy execution, such as strategy-specific *drift rate* (reflecting efficiency) and strategy selection parameters. The model estimates several key metacognitive control parameters: First, *baseline* (*strategy*) *switching time* refers to the average time required to switch strategies.

Neuroscience Accepted Manuscript

Second, *switching sensitivity* indicates the extent to which switching time is influenced by item difficulty. Third, *retrieval propensity* represents the tendency to favor retrieval over other strategies. Lastly, *adaptivity* reflects the optimality of an individual's retrieval strategy selection, calculated from the correlation between their retrieval efficiency and actual strategy choices across trials. A full list of parameters is provided in **Table 1**.

In BMAPS, trial-level responses (choice – accurate or inaccurate, and reaction time) are modeled as a drift diffusion model (DDM). The response of an individual *i* on trial *t* is denoted as x_{it} where $x_{it} = 1$ if the response was correct and $x_{it} = -1$ if the response was incorrect. The response time is similarly denoted as r_{it}. The choice coded response time is represented as $y_{it} = x_{it} r_{it}$. Here, y_{it} is treated as a sample from random variable Y_{it} which is distributed according to a Wiener distribution (Vandekerckhove et al., 2011; Wabersich & Vandekerckhove, 2014). The Wiener distribution describes the probability density function for Y_{it}, or in other words, the joint probability density function for the random variables X_{it} and R_{it}. The Wiener distribution is a four-parameter distribution that describes the first passage time of the diffusion process, that is, the distribution of times taken for the diffusion process to first hit the positive ($x = 1$) and negative ($x = -1$) decision boundaries. This is a canonical approach for modeling the joint distribution of choice and response time in two-choice processes (Ratcliff et al., 2016; Wagenmakers, 2009).

$$y_{i,t} \sim \text{Wiener}\left(\alpha_{i,t}, \tau_{i,t}, \beta_{i,t}, \delta_{i,t}\right) \quad (1)$$

Behavior is modeled as a *mixture model* of these latent strategies across trials. A latent value codes which strategy produces a response on any single trial: *S(t)*. The latent mixture drift diffusion model implemented can then be represented as:

$$y_{i,t} \sim \text{Wiener}\left(\alpha_{i,S(t)}, \tau_{i,S(t)}, \beta_{i,S(t)}, \delta_{i,S(t)}\right) \quad (2)$$

The selection of strategies is assumed to be dependent on a sequential process, with the first step modeling whether there is persistence of memory retrieval. This probability is modeled as a traditional 1-parameter logistic item response theory (1PL-IRT) model. Thus, for step 1, the probability of an individual *i* selecting a memory retrieval strategy for an item *k* ($s_{ik} = M$), depends on the individuals' proclivity towards memory retrieval μ_{iM} as well as how amenable the item is to be retrieved from memory η_{kM} (this is measured at a group level), and follows the form of 1PL-IRT model. The η_{kM} parameter can be interpreted as the ease of retrieval of item *k*. Importantly, this breaks down the propensity of retrieval into orthogonal individual and item level effects.

$$p(s_{ik} = M) = \frac{1}{1 + e^{-(\mu_{iM} + \eta_{kM})}}$$

In the second step, assuming that the result of the first step is to switch away from retrieval ($s_{ik} \neq M$), a nominal response model (Bock, 1972) is considered. This step models the probability of selecting an alternate strategy *S* by simultaneously taking into consideration the remaining 2 alternate strategies (counting and decomposition), where the probability of counting depends on item level effects (η_{kC}):

$$p(s_{ik} = C \mid s_{ik} \neq M) = \frac{1}{1 + e^{-(\eta_{kC})}}$$

JNeurosci Accepted Manuscript

This sequential process implies some switching point at which an individual might give up on the retrieval strategy and switch to an alternate strategy, in other words, the point at which the second step is initiated. Our model characterizes this switching time (ξ_{ik}) as the point at which a decision is taken on whether to switch away from memory by individual *i* for item *k.* It is possible that the actual retrieval takes longer, but enough confidence has been gained by then to take a decision on whether to switch. This switching time will vary by individual, but it may further vary depending on the nature of the item. We propose that people may have some form of access to the ease of retrieval of an item η_{kM} and their switching time may thus vary depending on a baseline switching time γ_i , switching sensitivity ω_i , and ease of item retrieval η_{kM}.

$$\xi_{ik} = \gamma_i + \omega_i \eta_{kM}$$

The memory retrieval process is characterized as a drift diffusion process, which posits that evidence accumulates over time resulting in a decision when a decision threshold is reached. The efficiency of evidence accumulation is characterized by a memory retrieval drift rate parameter (δ^M), with higher values characterizing faster and more accurate retrieval. The decision threshold parameter (α) captures the degree of confidence required to conclusively decide that a solution is retrieved from memory, with higher values characterizing slower but more accurate retrievals. The memory retrieval drift rate for an individual *i* for an item *k* is proposed as a combination of individual (δ_i^M) and item level effects:

$$\delta_{ik}^M = \frac{\delta_i^M}{1 + e^{-\eta_{kM}}}$$

Thus, for items which are considered very easy for retrieval (high positive values of η_{kM}), $\delta_{ik}^M \rightarrow \delta_i^M$, whereas for items considered very difficult for retrieval (high negative values of η_{kM}), $\delta_{ik}^M \rightarrow 0$.

The counting process is also characterized as a drift diffusion process, with a combination of individual and item level effects. The counting strategy used in this model is assumed to be *min* counting (counting up from the larger addend). For an item *k*, with the first addend presented being n_{1k} and the second being n_{2k}, the drift rate for counting is modeled as follows:

$$\delta_{ik}^{C\,[min]} = \frac{\delta_i^C}{\min(n_{1k}, n_{2k})}$$

The decomposition process can be characterized as a set of n_D steps taken to break down the addition process. For instance, 8+7 may be broken down as 7 = 2+5; 8+2 = 10; 10+5 = 15, which involves $n_D = 3$ steps.

$$\delta_{ik}^D = \frac{\delta_i^D}{n_D}$$

Each of these steps is at least as difficult as a counting step, so we apply the constraint, $\delta_i^D \leq \delta_i^C$

Additional details of the model are available in the Supplementary Material of Mistry et al. (2025).

To test whether children with DD exhibit deficits in the cognitive and metacognitive parameters of interest compared to their TA peers, we conducted independent-samples *t*-tests within a

JNeurosci Accepted Manuscript

Bayesian statistical framework for each parameter. Additional bivariate correlation analyses between BMAPS parameters and in-scanner arithmetic performance (accuracy and reaction times) were conducted for descriptive purposes only, to help readers contextualize the BMAPS parameters in relation to more intuitive performance measures. These correlations were not used for inferential conclusions.

As a control analysis to validate our BMAPS parameters, we examined problem size effects in strategy use. We first calculated the percentage of trials on which each strategy was used, separately for Small (sum $\leq$ 10), Medium (sum = 11-12), and Large (sum = 13-14) problems, based on sum values of the addition problems – consistent with the distinction between simple (small sum) and complex (medium and large sum) problems used by Geary et al. (2004) – and performed a chi-squared test across problem size and strategy bins.

fMRI data acquisition and preprocessing. Images from Kwok et al. (2023) were acquired on a Siemens 3T MAGNETOM Prisma scanner using a 32-channel head coil (Siemens, Erlangen, Germany). A T1-weighted high-resolution structural image was acquired using the MPRAGE sequence (TR = 2300ms, TE = 2.29ms, FOV = 204mm, 0.9mm isotropic voxels, and slice thickness of 0.94mm). A T2*-weighted EPI functional image was acquired using the following parameters: TR = 1000ms, TE = 28ms, FOV = 208mm, flip angle = 40°, 2.5mm isotropic voxels, and slice thickness of 2.5mm.

Images were preprocessed and analyzed using SPM12. All structural images were skull-stripped and segmented for co-registration. The preprocessing pipeline for functional data consisted of realignment, slice-timing correction, co-registration to each participant's individual structural T1-weighted image and normalization to a standardized 2mm Montreal Neurological Institute (MNI152) template in SPM and smoothing using a 6mm full-width half-maximum Gaussian kernel to decrease spatial noise. Translational motion (x, y, z) and rotational motion (pitch, roll, yaw) were calculated in millimeters based on the SPM12 realignment procedure. Mean framewise displacement of the participants did not exceed 0.6mm in any single run.

fMRI data analysis: General linear model. Following preprocessing, a general linear model (GLM) implemented in SPM12 was used to compute individual-level task-related brain responses. Brain responses associated with correct trials for each task and condition of interest (Arithmetic task: Large, Small, and Plus1 conditions; Matching task: Number and Shape conditions) were modeled using boxcar functions aligned with the corresponding trial duration. These functions were convolved with a canonical hemodynamic response function and a temporal derivative to account for voxel-wise variations in hemodynamic response latency. Additionally, both translational and rotational head motion parameters were included as covariates of no interest. We used correct trials only to maintain compatibility with Kwok et al. (2023) and because they better capture neural activity specific to arithmetic problem-solving and exclude processes unrelated to the task such as inattention, fatigue, or guessing that may be reflected in incorrect trials. Voxel-wise contrast maps were generated for each participant for each task. For the arithmetic task, the primary contrasts of interest were *Large* + *Small* vs. *Plus1,* as in Kwok et al. (2023), and *Large* vs. *Plus1*, based on Matejko and Ansari (2019), both of which showed strong effects at the neural level. Although *Small vs. Plus 1* was not our primary contrast of interest, voxel-wise contrast maps were generated to assess its significance. For the matching task, the contrast of interest was *Number* vs. *Shape*, which allowed us to assess nonsymbolic and symbolic quantity processing, distinct from the visual processing of shapes.

Neuroscience Accepted Manuscript

Next, we performed a whole-brain one-sample *t*-test to assess population-level within-subjects effects for each contrast of interest. A separate GLM was applied for the arithmetic and matching tasks. All statistical maps were masked with a grey matter mask, and significant clusters were thresholded with voxel-wise FDR-correction at $p < 0.01$ following Liu et al. (2025). Two binary masks derived from the identified significant voxels were selected for the subsequent EN decoding procedure. Masks were generated from the *Large* > *Plus1* contrast from the arithmetic verification task and the *Number* > *Shape* contrast from the matching task. The *Number* > *Shape* mask was included because this contrast captures number-specific visual and semantic processing relevant to early arithmetic problem-solving stages and encompasses a broader set of numerical cognition-related regions than the *Large* > *Plus1* mask alone. This approach restricted the voxel pool used for classification while preserving features relevant to arithmetic processing. The union of these masks was taken and used for subsequent decoding analysis.

Elastic Net decoding analysis. To test whether functional response patterns across distributed brain regions can distinguish children with DD from TA controls, we implemented an elastic net (EN) classifier using the python package glmnet_python (Balakumar, 2016) with leave-one-out cross-validation (LOOCV). This classifier integrates feature elimination from the least absolute shrinkage and selection operator (LASSO) method and feature coefficient reduction from ridge regression, resulting in shrinking the value of coefficients for model features that are relatively unimportant (Russell & Ginsburg, 1984; Geary, 1990; Geary et al., 2012).

Prior to decoding, we identified the task contrast and feature mask to be used. Among the arithmetic contrasts examined (Large + Small > Plus1, Large > Plus1, and Small > Plus1), Large > Plus1 produced the strongest and most widespread activation across frontoparietal and temporal regions (**Supplementary Table S3**) and corresponded to the condition with the most robust behavioral group difference; Small > Plus1 yielded no significant activation. Consistent with Kwok et al. (2023), none of these contrasts revealed univariate group differences. We therefore selected Large > Plus1 for decoding, as it maximized arithmetic-related signal while avoiding the dilution associated with the weaker Small > Plus1 contrast. For feature reduction, we constructed a union mask from the Large > Plus1 and Number > Shape (matching task) contrasts, comprising 24,669 voxels (17.54% of the whole brain; **Supplementary Figure S2**). All voxels within the union mask served as input features for the EN classifier, which was trained to classify the DD vs. TA groups using multivoxel data from the *Large* > *Plus1* contrast values.

The significance of classification accuracy was determined through permutation tests with 5000 iterations. The relative contribution of each voxel to classification (i.e., feature importance) was quantified using the coefficients estimated by the EN model. We first identified the peak coordinates of these brain feature clusters and created spherical regions of interest (ROIs) with a 6mm radius centered on a peak coordinate. To further examine the directionality of group differences between the DD and TA groups, we calculated Cohen's *d* values within the identified ROIs. Brain regions were labeled by the Brainnetome atlas (Fan et al., 2016) and Julich atlas (Amunts et al., 2020). The full analysis pipeline is outlined in **Figure 2**.

Control analysis. Although analyzing correct trials only better captures neural signals of interest, the lower number of correct trials in the DD group compared to the TA group may have introduced a bias in the estimated group differences in neural response. To verify our results are not driven by this potential bias, we re-ran the second-level GLM and subsequent EN decoding analyses using contrasts derived from all trials. We then examined whether the brain feature

Neuroscience Accepted Manuscript

weights identified by the original correct-trials-only EN model remain robust and consistent in the all-trials EN model. To this end, we performed a bivariate correlation of voxelwise coefficients estimated from each model within the brain mask defined by the original correct-trials-only EN model. Convergence between two models would confirm that the group differences identified by the correct-trials-only EN model are not driven by the lower trial count in the DD group.

Multivariate brain-behavior association analyses. To investigate whether functional brain features that distinguish children with DD from their TA peers are associated with latent cognitive mechanisms underlying arithmetic problem-solving, we conducted a multivariate canonical correlation analysis (CCA). This approach identifies linear combinations of brain measures (*Large* > *Plus1* functional contrasts in the 38 ROIs identified by EN decoding, excluding 2 brainstem ROIs) and behavioral measures (BMAPS parameters of interest) that maximize the brain-behavior correlations. CCA yields a set of canonical modes, each consisting of a pair of linear combinations for the behavioral and brain datasets that maximize their correlation, along with loadings that indicate the contribution of each individual behavioral and brain measures to these modes. This analysis enables us to capture the shared variance between functional activity and latent cognitive processes involved in arithmetic problem-solving, and to determine which ROIs/BMAPS parameters drive this relationship.

Transformer-based brain-behavior predictions. To further investigate which functional brain features predict specific latent cognitive mechanisms underlying arithmetic problem-solving, we employed Tabular Prior-data Fitted Network (TabPFN), a transformer-based foundation model designed for regression and classification on small- to medium-sized tabular datasets (Hollmann et al., 2025). Like large language models, TabPFN employs in-context learning, but rather than being trained on text, it is pre-trained on synthetic tabular datasets generated from causal models and Bayesian neural networks. This allows to learn a broad prior over data-generating processes and autonomously acquire effective strategies for common real-world data challenges – such as missing values, outliers, and uninformative features – without requiring the hand-engineered preprocessing steps, required by classical regression or machine learning approaches. Critically, rather than returning a single point estimate for each observation as in classical regression, TabPFN approximates the posterior predictive distribution (PPD) for each new dataset, implementing prediction under Bayesian principles and providing uncertainty quantification alongside its predictions. This offers more statistically principled inference for small- to medium-sized datasets. TabPFN was chosen over traditional machine learning approaches as it can simultaneously adapt to both parametric and nonparametric structure without any hyperparameter tuning, and has been shown to be competitive with, and in some cases outperform, Least Absolute Shrinkage and Selection Operator (LASSO) regression even when parametric assumptions are correctly specified (Zhang et al., 2025).

Brain activity from each ROI was used as the feature vector input to TabPFN to predict individual BMAPS parameters that showed significant group differences between the DD and TA groups (see **Results**). Parameters of particular interest were the drift rates for each strategy, baseline switching time, switching sensitivity, and adaptivity. Model performance for each BMAPS parameter was assessed using R^2 and Pearson's r, computed by correlating leave-one-out cross-validation (LOOCV)-predicted values with observed BMAPS values. To confirm cross-model consistency, we additionally ran LASSO and ridge regression models and compared R^2 and Pearson's r across all three approaches.

Neuroscience Accepted Manuscript

After confirming model consistency, feature importance for each ROI was computed using Shapley Additive Explanations (SHAP) (Lundberg & Lee, 2017; Hollmann et al., 2025), a model-agnostic framework for interpreting predictions. For a given feature, SHAP values quantify its contribution to shifting the model's prediction away from the baseline, positively or negatively, for each individual datapoint. This yields both an average SHAP value per feature, reflecting its overall importance across the ROIs, and a distribution of individual SHAP values, one per participant.

Code availability

All code used for this project will be shared via the Stanford Cognitive & Systems Neuroscience Laboratory GitHub webpage (https://github.com/scsnl).

Results

Poor arithmetic and number matching performance in children with DD

We first examined whether children with DD exhibit differences in performance across conditions in the arithmetic verification task. Bayesian independent samples *t*-tests revealed that children with DD showed lower accuracy across conditions (BF_{10} = 126.53, δ = -.90 [95% CIs = (-1.43, -.39)]) compared to TA children, reproducing the findings of Kwok et al. (2023) (**Table 2**). No group differences were observed in total reaction times (RT; BF_{10} = .88, δ = .38 [95% CIs = (.04, .86)]). Similar analysis conducted for each condition revealed group differences in the Large and Small conditions, but not in the Plus1 condition (BF_{10} = .53, δ = -.28 [95% CIs = (-.75, .17)]). In the Large condition, the DD group exhibited markedly lower accuracy than the TA group, yielding the largest observed group difference (BF_{10} = 705.81, δ = -1.05 [95% CIs = (-1.58, -.51)]). No group difference was observed in RT (BF_{10} = .27, δ = -.07 [95% CIs = (-.52, .38)]). In the Small condition, the DD group also exhibited lower accuracy (BF_{10} = 35.04, δ = -.79 [95% CIs = (-1.31, -.29)]) and slower RT (BF_{10} = 10.52, δ = .67 [95% CIs = (.19, 1.18)]) compared to the TA group.

Similar group differences were also observed in the matching task. Children with DD showed lower accuracy (BF_{10} = 6.30, δ = -.62 [95% CIs = (-1.12, -.15)]) compared to TA children (**Table 2**), but no differences in RT (BF_{10} = 0.34, δ = .24 [95% CIs = (.01, .66)]). Condition-specific analyses revealed a strong group difference in the Faces condition, with no effects in the Shapes (BF_{10} = 1.99, δ = -.49 [95% CIs = (-.98, -.08)]) and Number condition (BF_{10} = .56, δ = -.32 [95% CIs = (-.78, -.03)]). In the Faces condition, the DD group exhibited lower accuracy (BF_{10} = 10.34, δ = -.67 [95% CIs = (-1.18, -.19)]) than the TA group. None of the three conditions showed group differences in the RT (BF_{10} < .61). All group differences across conditions for both the arithmetic and matching tasks are visualized in **Supplementary Figure S1**.

These results suggest that group differences between the DD and TA groups, previously observed in Kwok et al. (2022), varied across task conditions.

Lower efficiency in arithmetic strategy execution and selection in children with DD

J Neurosci Accepted Manuscript

Our first goal was to examine whether children with DD differ from their TA peers in arithmetic strategy selection and execution processes. The Bayesian model of arithmetic problem-solving (BMAPS) (Mistry et al., 2025) estimated, for each trial, the probabilities of each strategy use (i.e., counting, retrieval, and decomposition), as well as cognitive parameters indexing strategy execution efficiency (e.g., strategy-specific drift rates) and metacognitive parameters related to strategy selection (e.g., baseline switching time and switching sensitivity).

To validate the BMAPS model in estimating strategy use, we performed a chi-squared test on the frequency of trials in which each strategy was used across small (sum $\leq$ 10), medium (sum = 11-12), and large (sum = 13-14) problem size bins. The results revealed a clear problem size effect: decomposition strategy use increases as the problem size increases (from 0.7% to 12.6% to 33.7% across small, medium, and large problems), whereas retrieval strategy use decreases as the problem size increases (from 55.5% to 4.4% and 7.1%). Results of the chi-squared test confirm that strategy dominance is dependent on problem size, $\chi^2(4) = 1630.51$, $p < .001$ (see **Supplementary Table S2**).

Across trials, the averaged probability of each strategy was 9.6% (SD = 15.6%) for decomposition, 46.8% (SD = 40.4%) for retrieval, and 43.6% (SD = 35.6%) for counting. The proportion of trials in which each strategy had the highest probability of being selected was 7.6% for decomposition, 38.8% for retrieval, and 53.6% for counting. Trial-averaged strategy probabilities did not differ between TAs and DDs (*ps* > .694; **Supplementary Table S1**), suggesting that both groups relied on similar strategies overall, with decomposition being the least frequently used.

For the efficiency measures, independent samples *t*-tests revealed that children with DD exhibited lower efficiency for all three strategies—as indicated by reduced drift rates (retrieval: $BF_{10} = 19.37$; counting: $BF_{10} = 348.10$; decomposition: $BF_{10} = 8.20 \times 10^{11}$; **Table 3**; **Figure 3**).

Similar to the results with drift rates, children with DD showed deficits in metacognitive processes related to strategy selection, including longer baseline switching time ($BF_{10} = 307.25$), lower switching sensitivity ($BF_{10} = 7.04 \times 10^6$), and reduced adaptivity ($BF_{10} = 4.29$), compared to TA children.

These results indicate that although children with DD and TA children employed comparable strategies, both strategy execution efficiencies and metacognitive control involved in strategy selection may be impaired in children with DD.

Whole-brain multivariate decoding distinguishes children with DD from typically-achieving peers

Our second goal was to determine whether distributed functional brain patterns during arithmetic problem-solving jointly distinguish children with DD from their TA peers. We employed a whole-brain multivariate decoding approach using an elastic net (EN) classifier with cross-validation. The classifier assigned feature-importance values to all voxels within a mask derived from the arithmetic verification task and matching tasks, indexing the relative contribution of each brain region to classifying the DD and TA groups (see details in **Methods**).

Decoding focused on the Large vs. Plus1 contrast from the arithmetic verification task, which yielded the strongest and most widespread arithmetic-related activation and showed the most

Neurosci Accepted Manuscript

robust behavioral group difference of all conditions; the rationale for selecting this contrast, and for the 24,669-voxel union mask used for feature reduction, is detailed in Methods (**Supplementary Tables S3–S4, Supplementary Figure S2**). Consistent with Kwok et al. (2023), no univariate group differences were detected for any contrast.

The EN model successfully distinguished the DD group from the TA group, achieving a classification accuracy of 69.23%, (p = .028 by permutation testing). **Table 4** reports model performance across EN mixing parameter α values). Results for alternative *Large* > *Plus1* and *Number* > *Shape* mask models are provided in **Supplementary Table S5**.

To identify regions contributing most to classification, we extracted voxels with peak feature importance from all significant clusters and defined 6mm spherical ROIs centered on these peaks. The resulting 40 ROIs spanned distributed cortical and subcortical regions, including the bilateral inferior and superior parietal lobules (IPL/SPL) including the intraparietal sulcus (IPS), superior and middle frontal gyrus (SFG/MFG), precentral gyrus (PreCG), lingual gyrus (LG), and fusiform gyrus (FG), as well as the right hippocampus and thalamus (**Figure 4**).

We then quantified the direction and magnitude of group differences in each ROI using Cohen's *d* (**Table 5**). The TA group showed greater activation in 26 of the 40 ROIs, while the DD group showed greater activation in the remaining 14. Among regions with the highest EN decoding coefficients, those with greater activation in the TA group included the right hippocampus (HIPP); right SFG, MFG, inferior frontal gyrus (IFG); right LG, right FG; and left SPL/IPL including the IPS. Regions with the highest peak coefficients that had higher activation in DDs included the bilateral PreCG and right MFG.

Together, these results show that children with DD exhibit functional activation patterns that are reliably distinguishable from those of their TA peers across distributed brain regions, even in the absence of detectable univariate group differences.

Control analysis: decoding is robust to trial-count differences between groups

Because only correct trials were included in the primary decoding analysis, and children with DD answered fewer trials correctly than TA children, the lower trial count in the DD group could in principle have biased the estimated group differences. To evaluate this possibility, we re-ran the EN decoding analysis using Large > Plus1 contrast maps generated from all trials, regardless of accuracy, thereby equating the two groups more closely in the number of trials contributing to each contrast map. We then correlated the voxelwise EN coefficients from this all-trials model with those from the original correct-trials-only model, restricted to the brain mask defined by the original model.

The two sets of coefficients were strongly and significantly correlated (r = .68, p < .001; **Supplementary Figure S3**), indicating that the spatial pattern of regional contributions to classification was largely preserved when all trials were included. This convergence confirms that the group differences identified by our primary decoding model are robust to differences in trial count between groups and are unlikely to be artifacts of the lower number of correct trials in the DD group.

JNeurosci Accepted Manuscript

Distributed brain activity patterns predict latent cognitive and metacognitive processes: Canonical correlation analysis

Having shown that distributed activation patterns distinguish the DD and TA groups, we next asked whether and how these patterns relate to the latent cognitive and metacognitive processes underlying arithmetic problem-solving. To capture system-level brain-behavior associations across the DD and TA groups, we conducted a canonical correlation analysis (CCA) relating the full set of EN-derived ROIs (brain data) to five BMAPS parameters (behavioral data): counting and retrieval drift rates, switching sensitivity, baseline switching time, and adaptivity. Decomposition drift rate was excluded because the probability of the decomposition strategy was rarely selected (<10% of trials in both groups) and carried high error rates (~50%), making it unlikely that activation would reliably track its use.

This analysis revealed a single significant canonical mode with a brain-behavior correlation of r = .956 (R^2 = 0.243), which distinguished the DD and TA groups along both the behavioral and brain dimensions (**Figure 5A**). Among behavioral parameters, counting drift rate and switching sensitivity loaded most strongly, whereas baseline switching time loaded most strongly in the negative direction (**Figure 5B**; **Table 6**). Among brain regions, the right HIPP, left SPL/IPS, right LG, and right cuneus showed the highest positive loadings, and the left SFG the highest negative loadings (**Figure 5C**; **Table 6**; **Supplementary Figure S4**).

Together, these results provide system-level evidence that the distributed neural signature distinguishing DD from TA children is jointly organized with the latent cognitive and metacognitive parameters governing arithmetic problem-solving, establishing a brain-behavior axis that we next decompose at the level of individual processes.

Distributed brain activity patterns predict latent cognitive and metacognitive processes: Tabular Prior-data Fitted Network prediction analysis

To identify which neural regions support each latent process and whether their contributions differ between groups, we used a transformer-based foundation model approach, the Tabular Prior-data Fitted Network (TabPFN; Hollmann et al., 2025), to predict BMAPS parameters from ROI activation, and quantified each region's contribution using normalized SHAP values. We then compared mean SHAP values between TA and DD children for each ROI (**Table 7**). Significant LOOCV results emerged for counting drift rate (R^2 = 0.053, Pearson's r = 0.271, p = .032) decomposition drift rate (R^2 = 0.205, r = 0.467, p < .001), and switching sensitivity (R^2 = 0.189, r = 0.492, p < .001). To evaluate cross-method consistency, we additionally ran LASSO and ridge regression models. Results from both methods were consistent with TabPFN, yielding significant prediction to counting drift rates, decomposition drift rate, and switching sensitivity, and nonsignificant predictions for the remaining parameters (see details in **Supplementary Table S6)**. These results confirm the robustness of our predictions.

Distributed brain activity patterns predict latent cognitive and metacognitive processes: Tabular Prior-data Fitted Network-based feature extraction

Across all three parameters, group comparisons revealed a striking and consistent reversal in the directionality of regional contributions: SHAP values were positive or near-zero in TA

eNeuro Accepted Manuscript

children but negative in DD children, such that contributions were systematically higher in TA than DD children (**Table 7**; **Figure 6**; **Supplementary Figure S5**). Thus, the same distributed regions that contributed positively to predicting these behavioral parameters in TA children contributed negatively in children with DD.

This directional reversal was most pervasive for switching sensitivity, our index of metacognitive control, where 36 of 38 ROIs (not including two brainstem clusters) differed significantly between groups (most at $p < .001$) with 22 ROIs showing group reversals, spanning frontal, parietal, hippocampal, visual, and subcortical regions. This near brain-wide group difference indicates that metacognitive control over strategy selection is supported by broadly positive distributed contributions in TA children but broadly negative contributions in DD. Group differences were also widespread for decomposition drift rate (26 of 38 ROIs) and more focal for counting drift rate (12 of 38 ROIs): both showed significant differences across distributed hippocampal, frontal, parietal, and dorsomedial frontal regions, while the latter was limited to hippocampus, lingual gyrus, superior frontal gyrus, and visual cortex. In every case, the significant differences followed the same TA-positive, DD-negative pattern.

The right hippocampus showed the most pronounced group reversal across all three parameters. Its contribution to counting drift rate reversed from positive in TA children (mean SHAP = 0.126) to strongly negative in DD children (−0.220; $p = .008$), with parallel reversals for decomposition drift rate (0.088 vs. −0.193; $p = .001$) and switching sensitivity (0.051 vs. −0.031; $p < .001$). Comparable counting-related reversals were observed in the right visual cortex/lingual gyrus (clusters 17 and 32: 0.078 vs. -0.128, $p < .001$; and 0.034 vs. −0.020; $p < .001$) and left superior frontal gyrus (clusters 22, 27, and 33: 0.133 vs. -0.118, $p = .005$; 0.041 vs. −0.025, $p = .002$; and 0.024 vs. -0.018, $p = .004$) (**Table 7**).

Together, these findings indicate that the distributed neural architecture supporting arithmetic strategy execution and metacognitive control is engaged in opposing directions in children with DD relative to their TA peers, with hippocampal disengagement during procedural computation as the most prominent group-distinguishing feature.

Discussion

We investigated cognitive, metacognitive, and neural mechanisms underlying arithmetic problem-solving deficits in children with developmental dyscalculia (DD), integrating Bayesian computational modeling of behavior with multivariate brain-wide decoding of fMRI data from Kwok et al. (2023). Three key insights emerged: First, children with DD exhibit deficits in both strategy execution efficiency, indexed by reduced drift rates for counting, retrieval, and decomposition, and metacognitive control, indexed by impaired switching sensitivity, longer switching times, and reduced adaptivity. Second, whole-brain Elastic Net classification identified distributed neural patterns across frontoparietal, hippocampal, and visual regions that distinguished DD from typically-achieving (TA) children. Third, canonical correlation analysis (CCA) and transformer-based predictions converged in showing that these distributed neural signatures predicted latent cognitive and metacognitive parameters, with the intraparietal sulcus (IPS), hippocampus, and frontoparietal cortical regions as critical nodes. Together, our findings provide the first comprehensive, mechanistic account linking latent cognitive processing deficits to their neural substrates during arithmetic problem-solving in DD.

Neurosci Accepted Manuscript

Our application of the Bayesian model of arithmetic problem-solving (BMAPS) (Mistry et al., 2025; Supekar et al., 2021; **Figure 1**) revealed pervasive deficits in DD extending beyond accuracy or strategy preferences. Both groups relied primarily on counting and retrieval strategies, with minimal use of decomposition, consistent with typical development at ages 8-10 for the problem set used in our study (Geary et al., 2004). Critically, no group differences emerged in the proportion of chosen strategies, confirming that children with DD employ the same strategic repertoire as their TA peers. However, they executed all three strategies less efficiently (**Figure 3A**), with lower drift rates for counting (d = 1.098), retrieval (d = 0.834), and decomposition (d = 2.629), indicating less accurate evidence accumulation. These results extend prior reports showing reduced accuracy in children with mathematical learning disabilities (Russell & Ginsburg, 1984; Geary, 1990; Geary et al., 2012) by quantifying trial-by-trial execution efficiency through rigorous computational modeling.

Children with DD exhibited pronounced impairments in metacognitive control (**Figure 3B**). Longer baseline switching time (d = -1.088) reflect general delays in switching from the initially chosen strategy, while reduced switching sensitivity (d = 1.829) revealed that their decisions to switch strategies were insensitive to problem difficulty. Unlike TA children, they did not modulate switching according to whether problems were easy or difficult. Reduced adaptivity (d = 0.666) further indicated that when they did switch strategies, their choices were often suboptimal. This pattern suggests that children with DD may lack both awareness that their current strategy is failing and the flexibility to shift adaptively to more effective approaches, leading to persistent use of suboptimal strategies. Because metacognition encompasses not only control but also monitoring processes such as error detection and performance evaluation (Nelson & Narens, 1990), and because effective switching presupposes accurate monitoring, future studies should adopt experimental and analytic designs that dissociate monitoring from control to clarify how the two interact in DD.

Our whole-brain decoding with ElasticNet revealed distributed brain activation patterns jointly distinguishing the groups (**Figure 4**, **Tables 4-5**), a result particularly notable given that univariate analyses of these data yielded null results (Kwok et al., 2023), underscoring the importance of multivariate approaches for characterizing learning disabilities. The identified regions encompassed the frontoparietal network commonly implicated in arithmetic processing (Rykhlevskaia et al., 2009; Peters & De Smedt, 2017; Arsalidou et al., 2018; Popescu et al., 2019); the salience network necessary for maintaining attentional focus and network switching (Ünal et al., 2025); the lingual and fusiform gyrus, which represent complex visual symbols including numbers (Cohen et al., 2002; Yeo et al., 2017; Chen et al., 2019); and the hippocampus, central for associative memory formation and recall (Qin et al., 2014; Cho et al., 2012; Giovanello et al., 2004). These findings converge on recent meta-analysis documenting atypical brain activation across similar regions in mathematical learning disabilities, including DD (Tablante et al., 2023).

This distributed signature indicates that DD reflects dysfunction in multiple brain systems, consistent with systems neuroscience models of mathematical cognition (Iuculano et al., 2018; Menon & Chang, 2022; Menon et al., 2024). Rather than implicating a single region, classification patterns encompassed “domain-specific” regions anchored on the intraparietal sulcus (IPS) (Butterworth et al., 2011; Price et al., 2007) and “domain-general” networks, including frontal regions associated with executive functions, working memory, and metacognition (Menon, 2016; Bellon et al., 2020). These findings suggest that arithmetic

J Neurosci Accepted Manuscript

problem-solving deficits in DD cannot be reduced to a single impaired representation but instead arise from disruption across distributed systems supporting numerical magnitude processing, visual-symbolic recognition, memory encoding, and cognitive control, consistent with multi-component accounts of DD (Fias et al., 2013; Menon, 2016; Menon & Chang, 2022; Liu et al., 2025). This pattern is consistent with evidence that mathematical processing draws on domain-general cognitive control systems shared with other skills, suggesting that DD involves dysfunction extending beyond number-specific representations (Ünal et al. 2025).

Brain-behavior analyses established direct mechanistic links between distributed neural signatures and specific latent deficits using two complementary approaches: CCA to characterize system-level brain-behavior covariation, and a transformer-based Tabular Prior-data Fitted Network (TabPFN; Hollmann et al., 2025) to identify which regions support each latent cognitive process and whether their contributions differ between groups. CCA revealed robust integration between distributed neural patterns and cognitive-metacognitive profile distinguishing the groups (R^2 = 0.243), with switching sensitivity and counting efficiency showing strong positive behavioral loadings, and parietal magnitude-processing regions, the hippocampus, and ventral-visual areas showing strong positive neural loadings (**Figure 5, Table 6**). Complementing these findings, TabPFN revealed that the relationship between distributed neural activity and latent processes was not simply weaker in DD, but directionally reversed: across switching sensitivity and counting and decomposition drift rates, regional contributions were positive or near-zero in TA children yet negative in children with DD (**Figure 6**, **Table 7**).

This reversal helps explain why conventional univariate analyses revealed no group differences (Kwok et al., 2023): the groups differed not in how strongly individual regions activated, but in how their distributed activity mapped onto behavior. The reversal was most pronounced in the hippocampus, the only region to differ significantly across all three parameters. The hippocampal contribution to counting and decomposition efficiency is particularly noteworthy, as hippocampal engagement is typically associated with direct memory retrieval rather than procedural strategies (Cho et al., 2011; Qin et al., 2014; Supekar et al., 2021). In children with DD, these contributions were strongly negative, suggesting that disengagement of the memory system may facilitate procedural strategy execution when retrieval and procedural processes would otherwise compete. This may point to disruptions in the typical coordination between memory and procedural systems, such that hippocampal recruitment competes with, rather than scaffolds, counting and decomposition, a mechanism that may be particularly characteristic of atypical numerical development.

A parallel reversal characterized the neural basis of metacognitive control. For switching sensitivity, contributions reversed in sign between groups across nearly the entire network, encompassing frontal, parietal, and salience regions implicated in performance monitoring and adaptive control (Bellon et al., 2020; Fleur et al., 2021; Ünal et al., 2025). In TA children, coordinated engagement of these regions supported switching sensitivity to problem difficulty, whereas in DD the same regions contributed negatively, pointing to a system-wide failure to flexibly mobilize the networks governing adaptive strategy selection. Frontal executive and parietal magnitude-processing regions routinely engaged during arithmetic problem-solving (Polspoel et al., 2017; Declercq et al., 2022) thus appear to be recruited in qualitatively different ways in DD, rather than simply being underactivated. Collectively, our results indicate that DD is characterized less by localized hypo- or hyperactivation than by an altered, often inverted,

Neuroscience Accepted Manuscript

mapping between distributed neural activity and the latent processes governing strategy execution and selection.

These differential associations may help inform the targeting of future intervention studies, pointing to strategy execution efficiency and adaptive strategy selection as candidate mechanisms. However, whether interventions addressing these mechanisms induce neural change in DD remains unknown, given the scarcity of neuroimaging-based intervention research in arithmetic problem-solving; establishing such links is an important next step.

An important open question is whether domain-general frontal cortex involvement is specific to DD or reflects signatures common across neurodevelopmental disorders. Mathematical and reading processing recruit shared cognitive control networks, with children showing broader engagement than adults (Ünal et al., 2025) and similar domain-general dysfunction is reported in dyslexia (Ashkenazi et al., 2013; Martin et al., 2015), suggesting frontal cortex recruitment may reflect executive demands arising whenever children struggle to deploy task-appropriate strategies. Critically, however, differential IPS involvement points to a “domain-specific” component unlikely to generalize, as IPS dysfunction is consistently linked to numerical magnitude processing (Butterworth et al., 2011; Price et al., 2007). Stratified clinical comparison groups will be needed to disentangle domain-general from DD-specific contributions.

In conclusion, our study provides the first comprehensive computational quantification of latent cognitive and metacognitive control processes governing strategy execution and selection in children with DD. By integrating Bayesian cognitive modeling with brain-wide decoding, we uncovered latent processes invisible to conventional approaches and linked neural patterns directly to cognitive mechanisms. More broadly, this integrative framework demonstrates the power of combining computational modeling with multivariate neuroimaging to understand individual variability in DD.

JNeurosci Accepted Manuscript

References

Amunts K, Mohlberg H, Bludau S, Zilles K (2020) Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. Science 369:988–992.

Arsalidou M, Pawliw-Levac M, Sadeghi M, Pascual-Leone J (2018) Brain areas associated with numbers and calculations in children: Meta-analyses of fMRI studies. Developmental Cognitive Neuroscience 30:239–250.

Ashkenazi S, Black JM, Abrams DA, Hoeft F, Menon V (2013) Neurobiological Underpinnings of Math and Reading Learning Disabilities. J Learn Disabil 46:549–569.

Ashkenazi S, Rosenberg-Lee M, Tenison C, Menon V (2012) Weak task-related modulation and stimulus representations during arithmetic problem solving in children with developmental dyscalculia. Developmental Cognitive Neuroscience 2:S152–S166.

Baird B, Smallwood J, Gorgolewski KJ, Margulies DS (2013) Medial and Lateral Networks in Anterior Prefrontal Cortex Support Metacognitive Ability for Memory and Perception. J Neurosci 33:16657–16665.

Balakumar BJ (2016) glmnet_python. Department of Statistics, Stanford University, Stanford, CA

Barbaresi WJ, Katusic SK, Colligan RC, Weaver AL, Jacobsen SJ (2005) Math learning disorder: incidence in a population-based birth cohort, 1976-82, Rochester, Minn. Ambul Pediatr 5:281–289.

Bellon E, Fias W, Ansari D, De Smedt B (2020) The neural basis of metacognitive monitoring during arithmetic in the developing brain. Human Brain Mapping 41:4562–4573.

Bock DR (1972) Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika 37:29–51.

Butterworth B, Varma S, Laurillard D (2011) Dyscalculia: From Brain to Education. Science 332:1049–1053.

Cai W, Griffiths K, Korgaonkar MS, Williams LM, Menon V (2021) Inhibition-related modulation of salience and fronto-parietal networks predicts cognitive control ability and inattention symptoms in children with ADHD. Mol Psychiatry 26:4016–4025.

Carpenter PA, Just MA, Shell P (1990) What One Intelligence Test Measures: A Theoretical Account of the Processing in the Raven Progressive Matrices Test. Psychological Review 97:404–431.

Chang H, Mistry PK, Zhang Y, Schwartz F, Menon V (2026) Latent Neurocognitive Mechanisms Underlying Quantity Discrimination in Children with and without Mathematical Learning Disabilities. J Neurosci 46:e2385242025.

Chang H, Rosenberg-Lee M, Qin S, Menon V (2019) Faster learners transfer their knowledge better: Behavioral, mnemonic, and neural mechanisms of individual differences in children's learning. Developmental Cognitive Neuroscience 40.

Chen L, Wassermann D, Abrams DA, Kochalka J, Gallardo-Diez G, Menon V (2019) The visual word form area (VWFA) is part of both language and attention circuitry. Nat Commun 10:5601.

J Neurosci Accepted Manuscript

Cho S, Metcalfe AWS, Young CB, Ryali S, Geary DC, Menon V (2012) Hippocampal-prefrontal engagement and dynamic causal interactions in the maturation of children's fact retrieval. J Cogn Neurosci 24:1849–1866.

Cho S, Ryali S, Geary DC, Menon V (2011) How does a child solve 7 + 8? Decoding brain activity patterns associated with counting and retrieval strategies. Dev Sci 14:989–1001.

Chua EF, Ahmed R (2016) Electrical stimulation of the dorsolateral prefrontal cortex improves memory monitoring. Neuropsychologia 85:74–79.

Cohen L, Lehéricy S, Chochon F, Lemer C, Rivaud S, Dehaene S (2002) Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. Brain 125:1054–1069.

Cooney JB, Ladd SF (1992) The influence of verbal protocol methods on children's mental computation. Learning and Individual Differences 4:237–257.

Declercq M, Bellon E, Sahan MI, Fias W, De Smedt B (2022) Arithmetic learning in children: An fMRI training study. Neuropsychologia 169:108183.

Fan L, Li H, Zhuo J, Zhang Y, Wang J, Chen L, Yang Z, Chu C, Xie S, Laird AR, Fox PT, Eickhoff SB, Yu C, Jiang T (2016) The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. Cereb Cortex 26:3508–3526.

Faulkenberry T, Ly A, Wagenmakers E-J (2020) Bayesian Inference in Numerical Cognition: A Tutorial Using JASP. Journal of Numerical Cognition.

Fazio BB (1994) The counting abilities of children with specific language impairment: a comparison of oral and gestural tasks. J Speech Hear Res 37:358–368.

Fias W, Menon V, Szucs D (2013) Multiple components of developmental dyscalculia. Trends in Neuroscience and Education 2:43–47.

Fleur DS, Bredeweg B, van den Bos W (2021) Metacognition: ideas and insights from neuro- and educational sciences. npj Sci Learn 6:13.

Geary DC (1990) A componential analysis of an early learning deficit in mathematics. Journal of Experimental Child Psychology 49:363–383.

Geary DC (2004) Mathematics and Learning Disabilities. J Learn Disabil 37:4–15.

Geary DC (2011) Consequences, Characteristics, and Causes of Mathematical Learning Disabilities and Persistent Low Achievement in Mathematics. Journal of Developmental & Behavioral Pediatrics 32:250–263.

Geary DC, Bow-Thomas CC, Yao Y (1992) Counting knowledge and skill in cognitive addition: a comparison of normal and mathematically disabled children. J Exp Child Psychol 54:372–391.

Geary DC, Hoard MK, Byrd-Craven J, Catherine DeSoto M (2004) Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. Journal of Experimental Child Psychology 88:121–151.

Geary DC, Hoard MK, Nugent L, Bailey DH (2012) Mathematical Cognition Deficits in Children With Learning Disabilities and Persistent Low Achievement: A Five-Year Prospective

JNeurosci Accepted Manuscript

Study. J Educ Psychol 104:206–223.

Giovanello KS, Schnyer DM, Verfaellie M (2004) A critical role for the anterior hippocampus in relational memory: evidence from an fMRI study comparing associative and item recognition. Hippocampus 14:5–8.

Guo L (2022) The effects of self-monitoring on strategy use and academic performance: A meta-analysis. International Journal of Educational Research 112:101939.

Hasson S, Ashkenazi S (2026) Heterogeneity in mathematics: Investigating cognitive profiles and reading comorbidities among children. Psychological Research 90:47.

Haxby JV, Connolly AC, Guntupalli JS (2014) Decoding neural representational spaces using multivariate pattern analysis. Annu Rev Neurosci 37:435–456.

Hollmann N, Müller S, Purucker L, Krishnakumar A, Körfer M, Hoo SB, Schirrmeister RT, Hutter F (2025) Accurate predictions on small data with a tabular foundation model. Nature 637:319–326.

Iuculano T, Padmanabhan A, Menon V (2018) Chapter 15 - Systems Neuroscience of Mathematical Cognition and Learning: Basic Organization and Neural Sources of Heterogeneity in Typical and Atypical Development. In: Heterogeneity of Function in Numerical Cognition (Henik A, Fias W, eds), pp 287–336. Academic Press.

Keysers C, Gazzola V, Wagenmakers E-J (2020) Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. Nat Neurosci 23:788–799.

Kirk EP, Ashcraft MH (2001) Telling stories: the perils and promise of using verbal reports to study math strategies. J Exp Psychol Learn Mem Cogn 27:157–175.

Kubota M, Hadley LV, Schaeffner S, Könen T, Meaney J-A, Morey CC, Auyeung B, Moriguchi Y, Karbach J, Chevalier N (2023) The effect of metacognitive executive function training on children's executive function, proactive control, and academic skills. Dev Psychol 59:2002–2020.

Kwok FY, Wilkey ED, Peters L, Khiu E, Bull R, Lee K, Ansari D (2023) Developmental dyscalculia is not associated with atypical brain activation: A univariate fMRI study of arithmetic, magnitude processing, and visuospatial working memory. Hum Brain Mapp 44:6308–6325.

Liu R, Chang H, El-Said D, Wassermann D, Zhang Y, Menon V (2025) Brain-wide decoding of numbers and letters: Converging evidence from multivariate fMRI analysis and probabilistic meta-analysis. Cortex 189:256–274.

Lundberg S, Lee S-I (2017) A Unified Approach to Interpreting Model Predictions. Available at: http://arxiv.org/abs/1705.07874 [Accessed April 28, 2026].

Martin A, Schurz M, Kronbichler M, Richlan F (2015) Reading in the brain of children and adults: a meta-analysis of 40 functional magnetic resonance imaging studies. Hum Brain Mapp 36:1963–1981.

Matejko AA, Ansari D (2019) The neural association between arithmetic and basic numerical processing depends on arithmetic problem size and not chronological age. Dev Cogn Neurosci 37:100653.

JNeurosci Accepted Manuscript

Menon V (2016) Working memory in children's math learning and its disruption in dyscalculia. Current Opinion in Behavioral Sciences 10:125–132.

Menon V, Chang H (2022) Theories of Dyscalculia. In: The Cambridge Handbook of Dyslexia and Dyscalculia (Skeide MA, ed), pp 25–44 Cambridge Handbooks in Psychology. Cambridge: Cambridge University Press.

Menon V, Padmanabhan A, Schwartz F (2024) Cognitive Neuroscience of Dyscalculia and Math Learning Disabilities. In: Oxford Handbook of Developmental Cognitive Neuroscience (Cohen Kadosh K, ed), pp 729–764. Oxford University Press.

Mistry PK, Chang H, El-Said D, Menon V (2025) Unraveling latent cognitive, metacognitive, strategic, and affective processes underlying children's problem-solving using Bayesian cognitive modeling. :2025.01.29.635409 Available at: https://www.biorxiv.org/content/10.1101/2025.01.29.635409v1 [Accessed April 29, 2025].

Nelson T, Narens L (1990) Metamemory: A Theoretical Framework and New Findings. In: Psychology of Learning and Motivation, pp 125–173. Academic Press. Available at: https://www.sciencedirect.com/science/chapter/bookseries/abs/pii/S0079742108600535 [Accessed April 2, 2026].

Ohtani K, Hisasaka T (2018) Beyond intelligence: a meta-analytic review of the relationship among metacognition, intelligence, and academic performance. Metacognition Learning 13:179–212.

Ostad SA (1997) Developmental differences in addition strategies: a comparison of mathematically disabled and mathematically normal children. British Journal of Educational Psychology 67:345–357.

Peters L, De Smedt B (2017) Arithmetic in the developing brain: A review of brain imaging studies. Dev Cogn Neurosci 30:265–279.

Peters G, De Smedt B, Torbeyns J, Verschaffel L, Ghesquière P (2014) Subtraction by addition in children with mathematical learning disabilities. Learning and Instruction 30:1–8.

Polspoel B, Peters L, Vandermosten M, De Smedt B (2017) Strategy over operation: neural activation in subtraction and multiplication during fact retrieval and procedural strategy use in children: Brain Activity in Children's Arithmetic Strategies. Hum Brain Mapp 38:4657–4670.

Popescu T, Sader E, Schaer M, Thomas A, Terhune DB, Dowker A, Mars RB, Cohen Kadosh R (2019) The brain-structural correlates of mathematical expertise. Cortex 114:140–150.

Price GR, Holloway I, Räsänen P, Vesterinen M, Ansari D (2007) Impaired parietal magnitude processing in developmental dyscalculia. Curr Biol 17:R1042-1043.

Qin S, Cho S, Chen T, Rosenberg-Lee M, Geary DC, Menon V (2014) Hippocampal-neocortical functional reorganization underlies children's cognitive development. Nat Neurosci 17:1263–1269.

Ratcliff R, Smith PL, Brown SD, McKoon G (2016) Diffusion Decision Model: Current Issues and History. Trends in Cognitive Sciences 20:260–281.

Rosenberg-Lee M, Ashkenazi S, Chen T, Young CB, Geary DC, Menon V (2015) Brain hyper-

JNeurosci Accepted Manuscript

connectivity and operation-specific deficits during arithmetic problem solving in children with developmental dyscalculia. Dev Sci 18:351–372.

Russell RL, Ginsburg HP (1984) Cognitive analysis of children's mathematics difficulties. Cognition and Instruction 1:217–244.

Rykhlevskaia E, Uddin LQ, Kondos L, Menon V (2009) Neuroanatomical correlates of developmental dyscalculia: combined evidence from morphometry and tractography. Front Hum Neurosci 3:51.

Siegler RS (1988) Individual Differences in Strategy Choices: Good Students, Not-So-Good Students, and Perfectionists. Child Development 59:833–851.

Siegler RS (1996) Emerging Minds: The Process of Change in Children's Thinking. Oxford University Press. Available at: https://academic.oup.com/book/40965 [Accessed June 30, 2025].

Siegler RS (2007) Cognitive variability. Developmental Science 10:104–109.

Siegler RS, Adolph KE, Lemaire P (1996) Strategy Choices Across the Life Span. In: Implicit Memory and Metacognition. Psychology Press.

Siegler RS, Lemaire P (1997) Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. Journal of Experimental Psychology: General 126:71–92.

Siegler RS, Shipley C (1995) Variation, selection, and cognitive change. In: Developing cognitive competence: New approaches to process modeling, pp 31–76. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Smith-Chant BL, LeFevre J-A (2003) Doing as they are told and telling it like it is: self-reports in mental arithmetic. Mem Cognit 31:516–528.

Sokolowski HM, Fias W, Mousa A, Ansari D (2017) Common and distinct brain regions in both parietal and frontal cortex support symbolic and nonsymbolic number processing in humans: A functional neuroimaging meta-analysis. Neuroimage 146:376–394.

Sokolowski HM, Hawes Z, 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.

Starling-Alves I, Peters L, Wilkey ED (2025) Beyond the sum of their parts: A multi-dimensional approach to dyscalculia-dyslexia comorbidity integrating studies of the brain, behavior, and genetics. Dev Cogn Neurosci 72:101510.

Supekar K, Chang H, Mistry PK, Iuculano T, Menon V (2021) Neurocognitive modeling of latent memory processes reveals reorganization of hippocampal-cortical circuits underlying learning and efficient strategies. Commun Biol 4:405.

Swanson HL, Hoskyn M, Lee C (1999) Interventions for students with learning disabilities: A meta-analysis of treatment outcomes. New York, NY, US: The Guilford Press.

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

JNeurosci Accepted Manuscript

Torbeyns J, Verschaffel L, Ghesquiere P (2004) Strategy Development in Childrenwith Mathematical Disabilities: Insights from the Choice/No-Choice Method and the Chronological-Age/ Ability-Level—Match Design. J Learn Disabil 37:119–131.

Ünal ZE, Park Y, Simsek E, Menon V, Geary DC (2025) Neurodevelopmental commonalities in cognitive control networks for mathematics and reading in meta-analysis of 3308 participants. Nat Commun 16:8398.

Vandekerckhove J, Tuerlinckx F, Lee MD (2011) Hierarchical diffusion models for two-choice response times. Psychological Methods 16:44–62.

Wabersich D, Vandekerckhove J (2014) Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). Behav Res 46:15–28.

Wagenmakers E-J (2009) Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. European Journal of Cognitive Psychology 21:641–671.

Wilson AJ, Andrewes SG, Struthers H, Rowe VM, Bogdanovic R, Waldie KE (2015) Dyscalculia and dyslexia in adults: Cognitive bases of comorbidity. Learning and Individual Differences 37:118–132.

Yeo DJ, Wilkey ED, Price GR (2017) The search for the number form area: A functional neuroimaging meta-analysis. Neuroscience & Biobehavioral Reviews 78:145–160.

Zhang Q, Tan YS, Tian Q, Li P (2025) TabPFN: One Model to Rule Them All? Available at: http://arxiv.org/abs/2505.20003 [Accessed May 6, 2026].

Zou H, Hastie T (2005) Regularization and Variable Selection Via the Elastic Net. Journal of the Royal Statistical Society Series B: Statistical Methodology 67:301–320.

JNeurosci Accepted Manuscript

Figure Legend

Figure 1. Schematic of Bayesian Model of Arithmetic Problem-Solving (BMAPS) framework. BMAPS models behavior as emerging from a latent mixture of drift-diffusion models, each representing a distinct cognitive strategy – retrieval, counting, and decomposition – embedded within a hierarchical metacognitive framework. Strategy selection follows two steps: ***Step 1.*** *an initial bias toward retrieval* during which evidence accumulates (green trajectory) toward either a correct (upper) or an incorrect (lower) response. ***Step 2.*** *potential mid-trial switch* when retrieval fails to reach threshold. In such cases, a probabilistic switch may occur from retrieval (**A**) to continued retrieval, (**B**) counting, or (**C**) decomposition, with the new strategy continuing until the decision threshold is reached or the trial ends. BMAPS infers the strategy use on each trial by comparing likelihoods across strategy-specific parameterization, yielding both trial-level and individual-level measures of cognitive and metacognitive processes. ***Cognitive parameters*** include strategy-specific evidence accumulation efficiency – retrieval (δ^R), counting (δ^C), and decomposition (δ^D). ***Metacognitive parameters*** include strategy switching adaptivity (Ψ), baseline switching time (γ), switching sensitivity to problem difficulty (ω), and switching uncertainty and variability (mean entropy Ω and standard entropy Σ). This sequential, within-trial modeling captures dynamic strategy adaptations that traditional self-report methods cannot detect. Figure adapted from Mistry et al. (2025).

Figure 2. Analysis pipeline. ***GLM-based contrast generation.*** For the arithmetic verification task, participants determined whether addition equations were correct (e.g., 8+7=15) across three difficulty levels: Large problems (sum >10), Small problems (sum ≤10), and Plus1 problems (one addend equals 1). For the matching task, children judged whether paired stimuli matched across three conditions: Number (dots vs. numerals), Shape (geometric forms), and Faces. We focused on the Large > Plus1 contrast for arithmetic and Number > Shape contrast for matching, as these produced the strongest within-subjects activation patterns (***Top panel***). ***Mask creation and feature selection.*** Statistical significance testing with voxelwise FDR correction ($p < .05$) and cluster-correction ($p < .01$) identified significant voxels for each contrast, visualized on inflated brain surfaces (red/blue = positive/negative contrasts). Binary masks were created from these significant clusters, then combined via union operation to create a comprehensive mask for brain decoding (***Middle***). ***Whole-brain Elastic Net decoding and ROI identification.*** Voxel-wise activation values (Large > Plus1 contrast) served as input features for Elastic Net classification to distinguish DD from TA children. Model training employed leave-one-out cross-validation (LOOCV). Statistical significance was determined via permutation testing with 5,000 iterations (*Bottom*).

Figure 3. Children with DD exhibit deficits in both strategy execution efficiency and metacognitive control. (A) ***Cognitive parameters.*** Children with DD demonstrated substantially lower drift rates compared to TA children across all three strategies: **Retrieval (δ^R**; Cohen's $d = 0.834$, $BF_{10} = 19.37$), **Counting (δ^C**; $d = 1.098$, $BF_{10} = 348.10$), and **Decomposition (δ^D; $d = 2.629$, $BF_{10} = 8.20 \times 10^{11}$)**. Lower drift rates indicate slower and less accurate accumulation of evidence toward correct responses. **(B)** ***Metacognitive parameters.*** Children with DD required more time to switch between strategies (**Baseline switching time, γ;** $d = -1.088$, $BF_{10} = 307.25$), indicating reduced strategic flexibility. They were also less sensitive to problem difficulty when adjusting strategies (**Switching sensitivity, ω;** $d = 1.829$, $BF_{10} = 7.04 \times 10^6$). Across trials, children with DD selected less optimal strategies (**Adaptivity, Ψ;** $d = 0.666$, $BF_{10} = 4.29$). Together, these cognitive and metacognitive impairments indicate that

Neurosci Accepted Manuscript

children with DD have difficulties executing strategies efficiently and adaptively and flexibly switching between them. Yellow boxes = DD group; Blue boxes = TA group. Box boundaries show the 25th-75th percentiles; center line = median. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$ based on Bayes Factors. *d* indicates Cohen's *d.*

Figure 4. Distributed neural signature of developmental dyscalculia (DD) identified through whole-brain Elastic Net decoding encompasses frontoparietal, hippocampal, and visual regions. Whole-brain Elastic Net classification with cross-validation identified brain regions whose distributed activation patterns jointly distinguished children with DD from TA peers with ($p = .028$, permutation testing). Brain regions are mapped to the Brainnetome atlas (Fan et al., 2016). **Key regions included** the bilateral precentral gyrus, right middle frontal gyrus, superior occipital gyrus, right hippocampus, right superior/middle/inferior frontal gyrus, bilateral superior and inferior parietal lobules including intraparietal sulcus, fusiform gyrus, and areas of the visual cortex including the lingual gyrus. Numbered labels for each cluster from **Table 5** are shown in brackets.

Abbreviations (ROIs): **HIPP** = hippocampus, **PreCG** = precentral gyrus, **MFG** = middle frontal gyrus, **AI** = anterior insula, **LG** = lingual gyrus, **FG** = fusiform gyrus, **SFG** = superior frontal gyrus, **OG** = occipital gyrus, **SPL** = superior parietal lobule, **IPS** = intraparietal sulcus, **dmPFC** = dorsomedial prefrontal cortex, **ACC** = anterior cingulate cortex, **IFG** = inferior frontal gyrus, **sup.** = superior.

Figure 5. Canonical correlation analysis revealed system-level brain-behavior associations across the DD and TA groups. (A) First canonical mode maximizing the brain-behavior correlation. Scatter plot shows the relationship between z-scores of the first canonical variate's brain measures (x-axis) and behavioral measures (y-axis). Brain data were derived from 38 cortical-subcortical ROIs and behavioral data were derived from 5 BMAPS parameters. The first canonical mode achieved a strong brain-behavior correlation of 0.956 ($R^2 = 0.243$, $p < .001$), effectively distinguishing the DD group (yellow points/ellipse) from the TA group (blue points/ellipse) across both neural and behavioral dimensions. Each point represents individual participant. Ellipses indicate 90% confidence regions. This system-level covariation demonstrates that distributed neural patterns and latent cognitive/metacognitive processes jointly characterize DD status. **(B) Behavioral loadings. *Positive loadings*** were found in **switching sensitivity** (ω; 0.77), **counting drift rate** (δ^C; 0.69), retrieval drift rate (δ^R; 0.58), and adaptivity (Ψ; 0.19). ***Negative loadings*** were found in **baseline switching time** (γ; -0.50). **(C) Brain loadings. Strongest *positive loadings*** were observed in right visual regions and the lingual gyrus (LG) (0.80 and 0.67), right anterior cingulate gyrus (ACC; 0.62), right anterior insula (AI; 0.60), right fusiform gyrus (FG; 0.59), right hippocampus (HIPP; 0.56), and two clusters in the left superior parietal lobule/intraparietal sulcus (SPL/IPS; 0.55 and 0.51). ***Negative loadings*** were observed in two clusters in the left superior frontal gyrus (SFG; -0.66 and -0.53). Only loadings with absolute values exceeding 0.50 are displayed; complete loadings can be found in **Table 6**. Loadings for all canonical modes are illustrated in **Supplementary Figure S4**. *Abbreviations*: DD = developmental dyscalculia; TA = typically-achieving.

Figure 6. Distributed brain activity patterns predict metacognitive switching sensitivity and procedural strategy efficiency, with the directionality of regional contributions differing across DD and TA groups. TabPFN feature importance was estimated using Shapley Additive Explanations (SHAP); values indicate the magnitude and direction of each ROI's contribution across participants for each BMAPS parameter. SHAP values reveal a major

Neuroscience Accepted Manuscript

divergence between the TA and DD groups in both the magnitude and direction of regional contributions. The most pronounced group reversals were found in the right hippocampus for all three BMAPS parameters (positive in the TA group, negative in the DD group), as well as in the dmPFC/ACC and visual cortex. Group differences were most pervasive for switching sensitivity (36 ROIs; brainstem ROIs not shown), with 22 ROIs showing strong group reversals, especially in the right PreCG, right MFG, and right visual cortex. For strategy efficiency, both drift rates showed more focal group differences, though most were in opposite directions between groups. Counting drift rates were concentrated in the left SFG, bilateral visual cortex, and right SPL/IPS. Decomposition drift rates showed group reversals across more widespread brain regions compared to counting, including the frontal and fusiform gyrus.

Regions labeled with an asterisk (*) show a statistically significant ($p < .01$) divergence for feature importance between TAs and DDs, while label color indicates whether the mean SHAP values for TAs and DDs have the same (e.g., positive SHAP in both groups with group difference in magnitude only) or opposite (e.g., positive SHAP in TAs versus negative SHAP in DDs) sign: red = opposite; blue = same. Bar shading (light versus dark) indicates statistical significance of SHAP values based on within-group one-sample t-tests for that region (light = not significant; dark = significant).

Abbreviations (ROIs): **HIPP** = hippocampus, **PreCG** = precentral gyrus, **MFG** = middle frontal gyrus, **AI** = anterior insula, **LG** = lingual gyrus, **FG** = fusiform gyrus, **SFG** = superior frontal gyrus, **OG** = occipital gyrus, **SPL** = superior parietal lobule, **IPS** = intraparietal sulcus, **dmPFC** = dorsomedial prefrontal cortex, **ACC** = anterior cingulate cortex, **IFG** = inferior frontal gyrus**, AG** = angular gyrus, **LOC** = lateral occipital cortex, **sup.** = superior.

JNeurosci Accepted Manuscript

Table 1. Definition of Bayesian model of arithmetic problem-solving (BMAPS) parameters.

BMAPS parameter	Symbol	Definition
Retrieval drift rate	δ^R	Rate of evidence accumulation for retrieval
Counting drift rate	δ^C	Rate of evidence accumulation for counting
Decomposition drift rate	δ^D	Rate of evidence accumulation for decomposition
Retrieval propensity	μ	Tendency to favor retrieval over other strategies
Decision threshold	α	Amount of evidence required for a given decision
Baseline switching time[1]	γ	Time to switch strategies (baseline, across trials)
Switching sensitivity	ω	Modulates strategy switching time for a given trial based on ease of item retrieval
Counting boost	κ	Increase in efficiency for counting vs. memory retrieval
Non-decision time[2]	τ	Portion of RT not dedicated to decision-making
L-R bias	β	Response bias for binary decision (0-1, 0.5 = no bias)
Adaptivity	Ψ	Optimality of strategy selection (across trials)
Mean entropy	Ω	Uncertainty in strategy choice (derived from latent probability of all strategies)
Standard entropy	Σ	Variability of mean entropy across trials

[1]Trial-level switching time is defined as the time at which an individual decides to switch away from the initial retrieval strategy; a longer switching time is expected with either a high switching sensitivity or if that trial's solution is easier to retrieve from memory (or if individual baseline switching time is higher).

JNeurosci Accepted Manuscript

[2]Includes time to encode the presented stimulus prior to beginning execution of the initial strategy; the remaining RT is assumed to reflect the actual strategy selection and decision-making processes.

JNeurosci Accepted Manuscript

Table 2. Mean accuracy and reaction times on arithmetic and matching tasks, and differences between the DD and TA groups.

Condition	Accuracy (SD)		BF_{10}	Reaction times (SD)		BF_{10}
	TA	**DD**	**Group differences**	**TA**	**DD**	**Group differences**
Arithmetic verification task						
Total	87.3 (.09)	76.4 (.13)	126.53***	2410 (430)	2591 (410)	.88
Large	79.2 (.14)	61.5 (.18)	705.81***	2818 (574)	2766 (685)	.27
Small	89.1 (.11)	77.4 (.16)	35.04**	2363 (505)	2745 (507)	10.52**
Plus-1	93.8 (.09)	90.5 (.11)	.53	2049 (403)	2264 (535)	1.06
Matching task						
Total	83.0 (.06)	77.9 (.08)	6.30*	1404 (132)	1431 (127)	.34
Numbers	77.3 (.10)	74.0 (.09)	.56	1450 (160)	1495 (146)	.38
Shapes	93.0 (.06)	89.0 (.07)	1.99	1241 (161)	1299 (168)	.61
Faces	79.0 (.09)	71.2 (.12)	10.52**	1520 (151)	1511 (194)	.28

Mean reaction times and their standard deviations are rounded to the nearest millisecond.

*BF > 3, **BF > 10, ***BF > 10

JNeurosci Accepted Manuscript

1091

JNeurosci Accepted Manuscript

Table 3. Group differences in BMAPS parameters between the TA and DD groups, and their correlations with in-scanner arithmetic performance.

	Independent *t*-tests (Cohen's d)	**Pearson's r**	
BMAPS parameter	TA > DD	Arithmetic acc.	Arithmetic RT
Retrieval drift rate	0.834**	0.844***	-0.261
Counting drift rate	1.098***	0.864***	-0.376**
Decomposition drift rate	2.629***	0.816***	-0.392**
Baseline switch time	-1.088***	-0.389**	0.626***
Switching sensitivity	1.829***	0.359*	0.059
Counting boost	0.797**	0.480***	-0.194
Adaptivity	0.666*	0.213	-0.289
Retrieval propensity	0.149	-0.230	-0.313*
Decision threshold	-0.484	-0.049	0.636***
Non-decision time	0.656*	0.463***	-0.117
L-R bias	0.237	0.243	-0.164
Mean entropy	0.253	-0.341*	-0.426**
Standard entropy	-0.088	0.266	0.294

*BF_{10} > 3, **BF_{10} > 10, ***BF_{10} > 100

Reported correlations between BMAPS parameters and accuracy/RT are presented for descriptive purposes only.

JNeurosci Accepted Manuscript

Table 4. Elastic Net (EN) neural decoding performance depending on alpha value.

α	Accuracy	Sensitivity	Specificity	TN	FP	FN	TP	p-value
0	0.5692	0.4286	0.6757	25	12	16	12	0.225
0.25	0.6000	0.2143	0.8919	33	4	22	6	0.163
0.5	0.6154	0.3929	0.7838	29	8	17	11	0.131
0.75	0.6154	0.2857	0.8649	32	5	20	8	0.112
0.9	0.6462	0.3571	0.8649	32	5	18	10	0.044*
1	0.6923	0.4643	0.8649	32	5	15	13	0.028*

Larger α values incur a greater penalty for having many nonzero features. **TN** = true negative (TA correctly categorized as TA); **FP** = false positive, **FN** = false negative; **TP** = true positive (DD correctly categorized as DD). **Sensitivity** = TP / (TP + FN). **Specificity** = TN / (TN + FP). P-values are based on permutation testing of model accuracy with 5000 iterations.

*$p < .05$

JNeurosci Accepted Manuscript

Table 5. Brain regions identified by Elastic Net (EN) neural decoding with different features between the DD and TA group in the arithmetic verification task.

Centroid #	Region	x	y	z	Peak coefficient	Cohen's d
DD > TA						
2	R PreCG	26	-10	50	1.8220	-0.556
3	R MFG	26	4	48	1.7315	-0.505
9	R sup. OG	26	-78	40	0.5866	-0.337
11	L PreCG	-24	-12	52	0.5191	-0.348
13	L occipital FG	-40	-66	-10	0.4576	-0.020
19	L dmPFC / ACC	-2	24	34	0.3208	-0.313
20	L sup. OG	-20	-68	36	0.3126	-0.215
21	L IFG	-46	16	0	0.3085	-0.067
22	L SFG	-14	16	66	0.2867	-0.442
23	Medial SFG	0	2	72	0.2839	-0.177
24	L frontal pole	-40	52	14	0.2505	-0.064
27	L SFG	-10	12	66	0.2326	-0.442
33	L SFG	-10	14	64	0.0712	-0.313
36	L frontal pole	-36	48	12	0.0379	-0.113
TA > DD						
1	R HIPP	32	-32	0	2.3527	0.714
4	R AI	28	28	2	1.4670	0.653
5	R MFG	40	20	30	0.9132	0.637
6	L Visual Cortex (V1) / LG	-14	-86	2	0.8723	0.768

JNeurosci Accepted Manuscript

7	L FG	-36	-64	-12	0.6797	0.002
8	R SFG	6	24	60	0.6692	0.334
10	L SPL/IPS	-28	-56	56	0.5764	0.317
12	R occipital FG	30	-80	-6	0.4622	0.836
14	R MFG	42	12	42	0.4464	0.670
15	R brainstem	10	-22	-12	0.4218	0.517
16	L occipital FG	-26	-84	-18	0.3785	0.723
17	R Visual Cortex (V1/V2) / LG	12	-86	-2	0.3639	0.889
18	L SPL/IPS	-28	-56	52	0.3239	0.459
25	L IFG	-46	14	16	0.2376	0.627
26	R IPS/AG	46	-48	46	0.2374	0.770
28	R thalamus	16	-32	2	0.2159	0.665
29	R SPL/IPS	36	-46	48	0.1878	0.453
30	R Visual Cortex (V1/V2) / LG	20	-86	-2	0.1537	0.808
31	R ACC	8	4	28	0.1359	0.551
32	R V1/LG	4	-80	6	0.1109	0.531
34	R MFG	52	36	28	0.0577	0.653
35	R brainstem	10	-22	-8	0.0398	0.412
37	R MFG	48	32	38	0.0321	0.734
38	R LOC/IPS	40	-60	40	0.0320	0.723
39	R LOC/IPS	38	-60	38	0.0159	0.723
40	R thalamus	22	-30	4	0.0049	0.631

JNeurosci Accepted Manuscript

Mean peak coefficient is .465 (SD = .527). Cohen's d indicates the effect size associated with a *t*-test comparing functional activity between the TAs and DDs for the Large > Plus1 contrast. *Abbreviations (ROIs)*: **HIPP** = hippocampus, **PreCG** = precentral gyrus, **MFG** = middle frontal gyrus, **AI** = anterior insula, **LG** = lingual gyrus, **FG** = fusiform gyrus, **SFG** = superior frontal gyrus, **OG** = occipital gyrus, **SPL** = superior parietal lobule, **IPS** = intraparietal sulcus, **dmPFC** = dorsomedial prefrontal cortex, **ACC** = anterior cingulate cortex, **IFG** = inferior frontal gyrus**, AG** = angular gyrus, **LOC** = lateral occipital cortex, **sup.** = superior.

JNeurosci Accepted Manuscript

Table 6. Canonical loadings for brain/behavioral variables onto first canonical mode (CCA analysis).

Behavioral loadings		Brain loadings	
Positive loadings			
Switching sensitivity	.77	*R Visual Cortex (V1/V2) / LG (17)*	.80
Counting drift rate	.69	*R Visual Cortex (V1/V2) / LG (30)*	.67
Retrieval drift rate	.58	*R ACC (31)*	.62
Adaptivity	.19	*R AI (4)*	.60
		R occipital FG (12)	.59
		R HIPP (1)	.56
		L SPL/IPS (18)	.55
		L Visual Cortex (V1) / LG (6)	.52
		L SPL/IPS (10)	.51
Negative loadings			
Baseline switching time	-.50	*L SFG (22)*	-.66
		L SFG (27)	-.53

Cluster numbers are given in parentheses besides the name of each ROI. To save space, only ROIs with loadings greater than |±.50| are shown.

Abbreviations (ROIs): **HIPP** = hippocampus, **AI** = anterior insula, **LG** = lingual gyrus, **FG** = fusiform gyrus, **SFG** = superior frontal gyrus, **SPL** = superior parietal lobule, **IPS** = intraparietal sulcus, **ACC** = anterior cingulate cortex.

JNeurosci Accepted Manuscript

Table 7. Group differences in mean normalized SHAP values between TAs and DDs for each ROI in the TabPFN models for switching sensitivity (ω) and counting/decomposition drift rates (δ^C, δ^D).

	Region	SHAP						p-value		
		ω		δ^C		δ^D		ω	δ^C	δ^D
		TA	**DD**	**TA**	**DD**	**TA**	**DD**			
1	R HIPP	0.051	-0.031	0.126	-0.220	0.088	-0.193	<0.001**	0.008*	0.001*
2	R PreCG	0.144	-0.047	0.020	0	0.044	-0.015	<0.001**	0.098	0.001*
3	R MFG	0.071	-0.026	0.006	0.001	0.026	-0.016	<0.001**	0.581	0.005*
4	R AI	-0.084	-0.072	0.008	-0.003	0.013	-0.012	0.010	0.332	0.001*
5	R MFG	0.024	-0.019	0.033	0.006	0.027	-0.011	<0.001**	0.025	0.001*
6	L V1 & LG	-0.017	-0.021	0.013	-0.061	0.054	-0.066	<0.001**	0.001*	<0.001**
7	L FG	0.02	-0.025	-0.003	-0.015	0.015	-0.012	<0.001**	0.175	0.007*
8	R SFG	-0.008	-0.013	0.019	-0.019	0.019	-0.044	<0.001**	0.027	0.002*
9	R sup. OG	0.047	-0.038	0.009	-0.04	0.041	-0.013	0.005*	0.001*	0.003*
10	L SPL/IPS	-0.047	-0.023	0.009	-0.024	0.011	-0.004	0.005*	0.114	0.036
11	L PreCG	-0.008	-0.011	-0.011	-0.011	0.023	-0.003	<0.001**	0.898	0.002*
12	R occipital FG	-0.016	-0.044	-0.007	0.005	0.017	-0.018	<0.001**	0.533	<0.001**
13	L occipital FG	0.04	-0.024	0.039	0.012	0.022	-0.009	0.006*	0.166	0.054
14	R MFG	0.042	-0.018	0.027	-0.018	0.014	-0.031	0.001*	0.005*	0.005*
16	L occipital FG	0.003	-0.011	0.014	-0.013	0.011	-0.035	<0.001**	0.139	0.019
17	R V1/V2 & LG	0.007	-0.023	0.078	-0.128	0.077	-0.141	<0.001**	<0.001**	<0.001**
18	L SPL/IPS	-0.012	-0.035	0.008	-0.012	-0.001	-0.024	0.003*	0.056	0.008*
19	L dmPFC / ACC	0.034	-0.012	0.051	-0.019	0.034	-0.013	<0.001**	0.001*	<0.001**
20	L sup. OG	0.048	-0.018	0.014	0	0.012	0.004	<0.001**	0.142	0.245

JNeurosci Accepted Manuscript

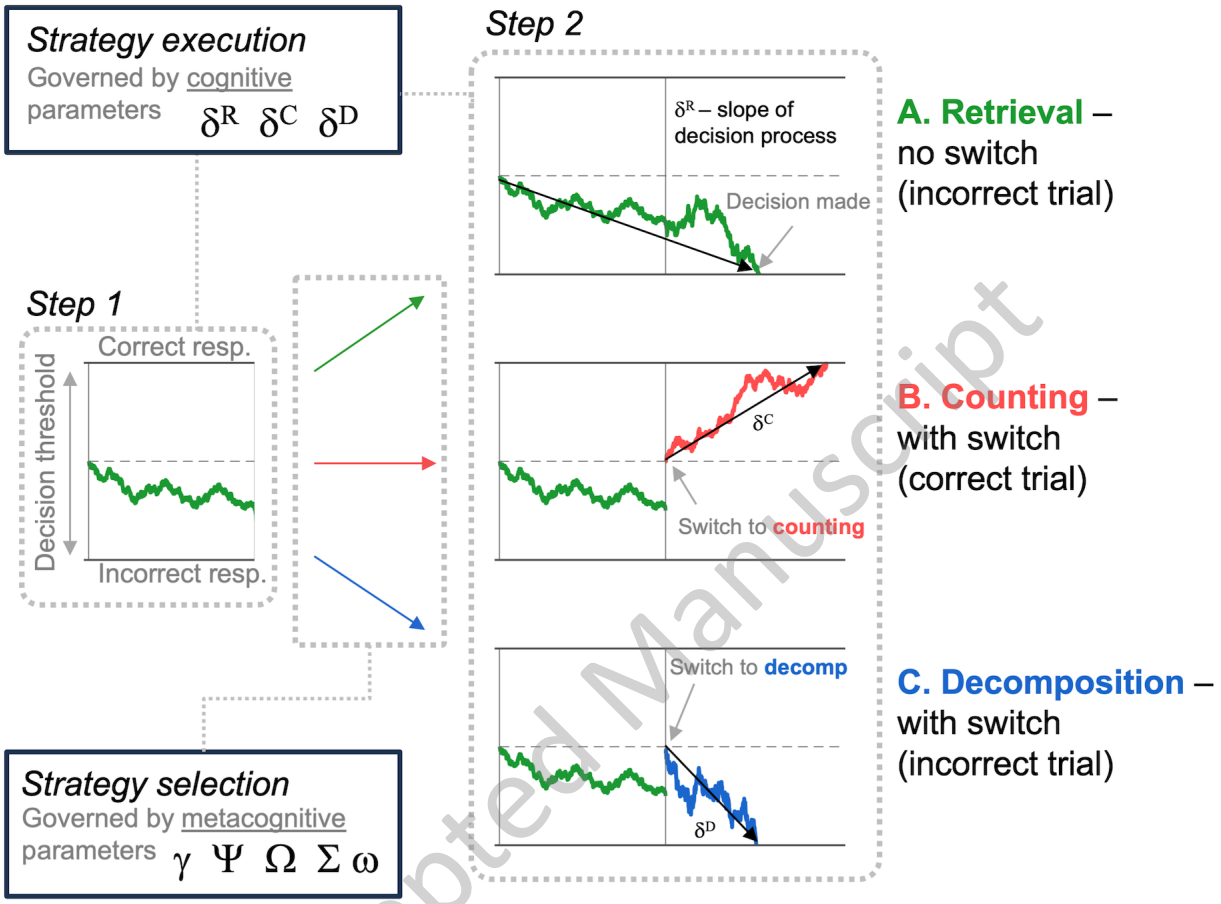
21	L IFG	0.025	-0.049	-0.003	-0.018	0.014	-0.012	0.009*	0.412	0.002*
22	L SFG	-0.017	-0.016	0.133	-0.118	0.115	-0.093	<0.001**	0.005*	0.001*
23	Medial SFG	0.016	-0.009	0.024	-0.009	0.026	-0.020	<0.001**	0.119	0.002*
24	L frontal pole	0.001	-0.005	0.005	0.003	0.011	0.003	<0.001**	0.931	0.242
25	L IFG	0.008	-0.049	0.043	-0.017	0.017	-0.020	<0.001**	0.001*	<0.001**
26	R IPS/AG	-0.015	-0.014	-0.010	-0.01	0.015	0.001	0.138	0.872	0.048
27	L SFG	0.002	-0.010	0.041	-0.025	0.064	-0.030	<0.001**	0.002*	<0.001**
28	R thalamus	-0.017	-0.021	0.027	-0.003	0.010	-0.005	<0.001**	0.020	0.039
29	R SPL/IPS	-0.022	-0.073	0.010	-0.040	0.001	-0.019	0.004*	0.009*	0.003*
30	R V1/V2 & LG	0.036	-0.084	-0.009	-0.025	0.012	-0.032	<0.001**	0.082	<0.001**
31	R ACC	-0.013	-0.015	0.030	-0.029	0.023	-0.025	<0.001**	0.047	<0.001**
32	R V1/LG	0.003	-0.012	0.034	-0.020	0.026	-0.015	<0.001**	<0.001**	<0.001**
33	L SFG	0.005	-0.010	0.024	-0.018	0.032	-0.007	<0.001**	0.004*	<0.001**
34	R MFG	-0.001	-0.019	0.012	0.020	0.004	-0.014	<0.001**	0.334	0.058
36	L frontal pole	-0.011	-0.008	0.034	0.004	0.018	-0.001	<0.001**	0.031	0.034
37	R MFG	-0.016	-0.029	-0.008	-0.003	-0.002	-0.002	<0.001**	0.670	0.985
38	R LOC/IPS	-0.019	-0.010	0.007	0.001	0.008	-0.005	0.001*	0.620	0.070
39	R LOC/IPS	-0.009	-0.016	0.005	0.020	0.009	-0.007	<0.001**	0.392	0.035
40	R thalamus	-0.012	-0.021	0.026	0.005	0.023	-0.010	<0.001**	0.058	<0.001**

Abbreviations (ROIs): **HIPP** = hippocampus, **PreCG** = precentral gyrus, **MFG** = middle frontal gyrus, **AI** = anterior insula, **LG** = lingual gyrus, **FG** = fusiform gyrus, **SFG** = superior frontal gyrus, **OG** = occipital gyrus, **SPL** = superior parietal lobule, **IPS** = intraparietal sulcus, **dmPFC** = dorsomedial prefrontal cortex, **ACC** = anterior cingulate cortex, **IFG** = inferior frontal gyrus**, AG** = angular gyrus, **LOC** = lateral occipital cortex, **sup.** = superior.

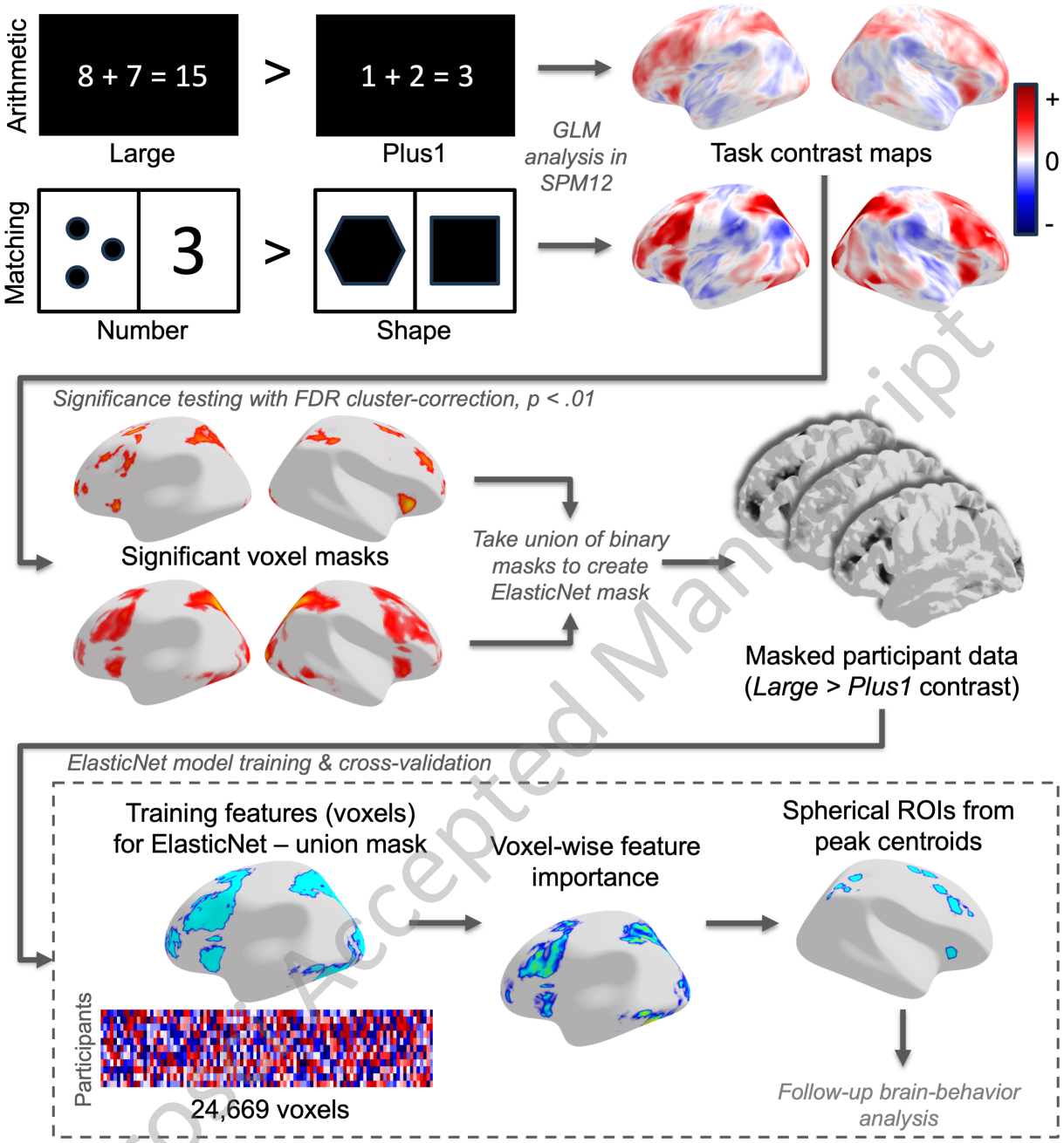
*p < .01, **p < .001. Significance determined using an α level of 0.01. Brainstem clusters not shown.

Neurosci Accepted Manuscript

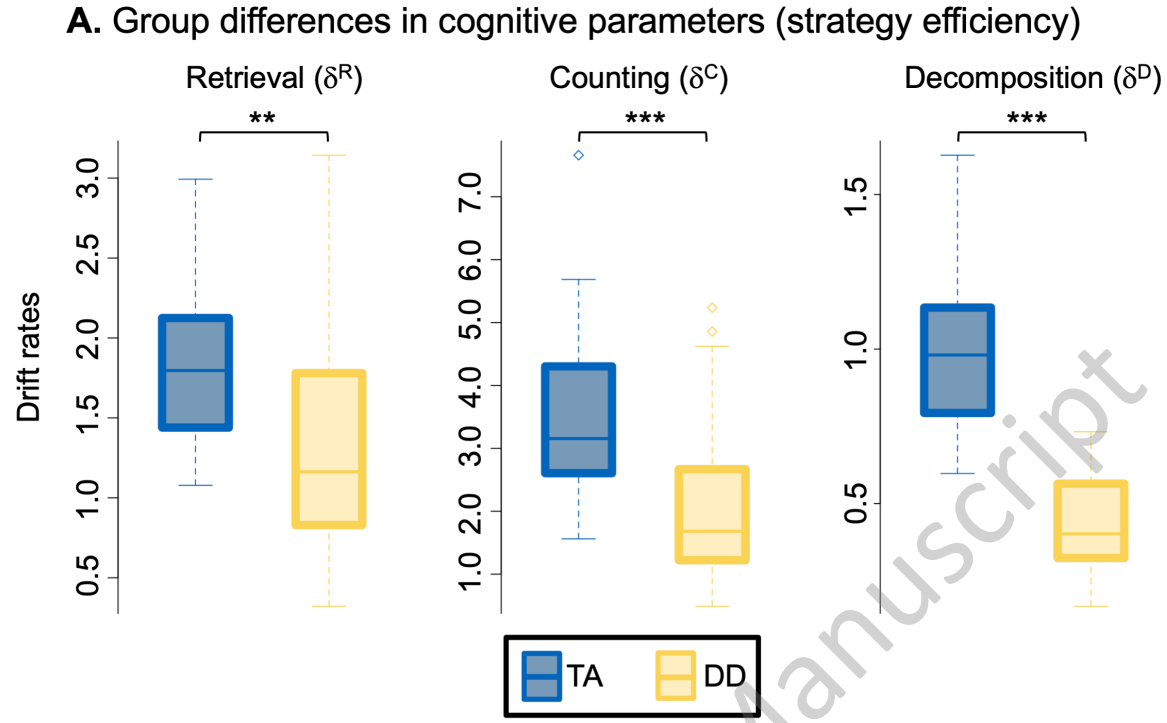
Schematic of BMAPS drift diffusion process for each strategy (3 sample trials)

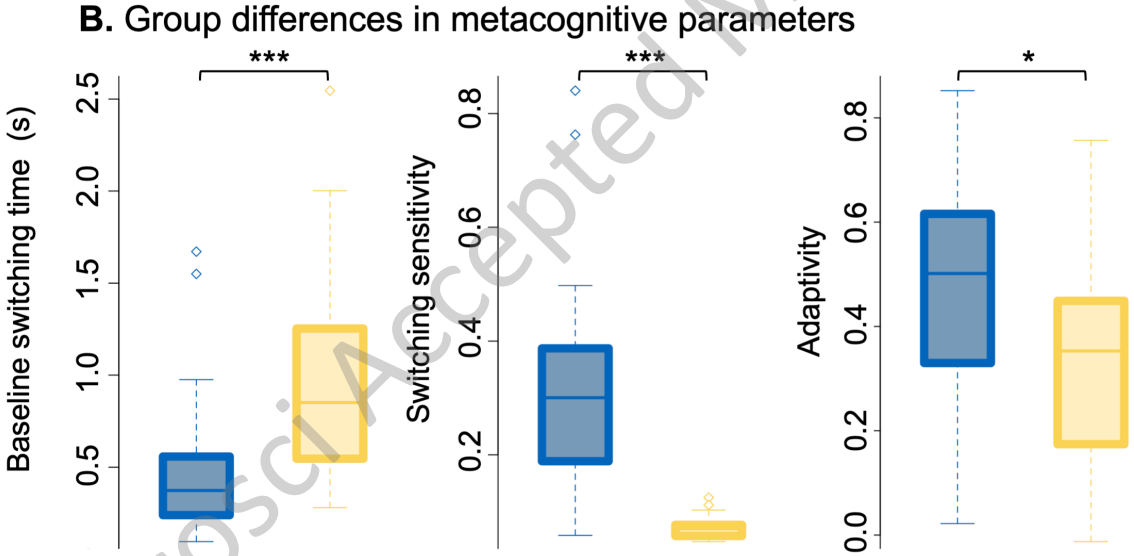


JNeurosci Accepted Manuscript



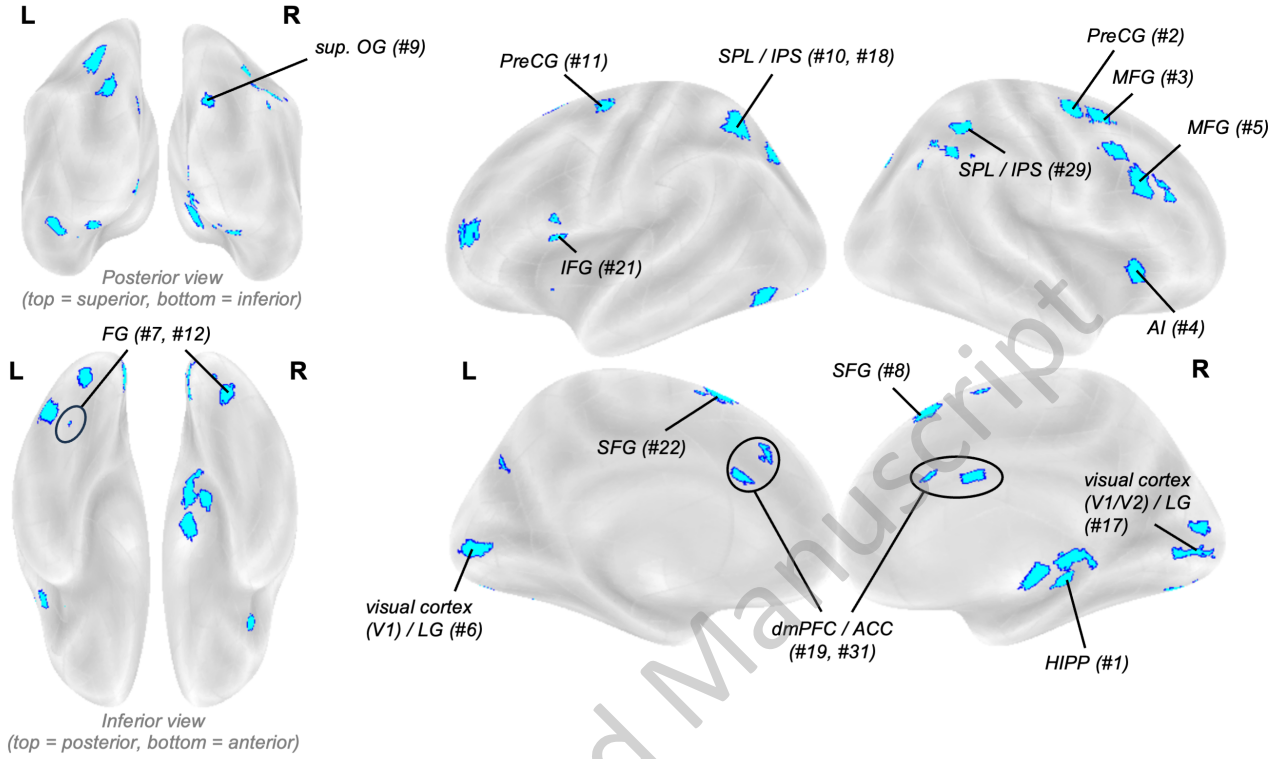
JNeurosci Accepted Manuscript



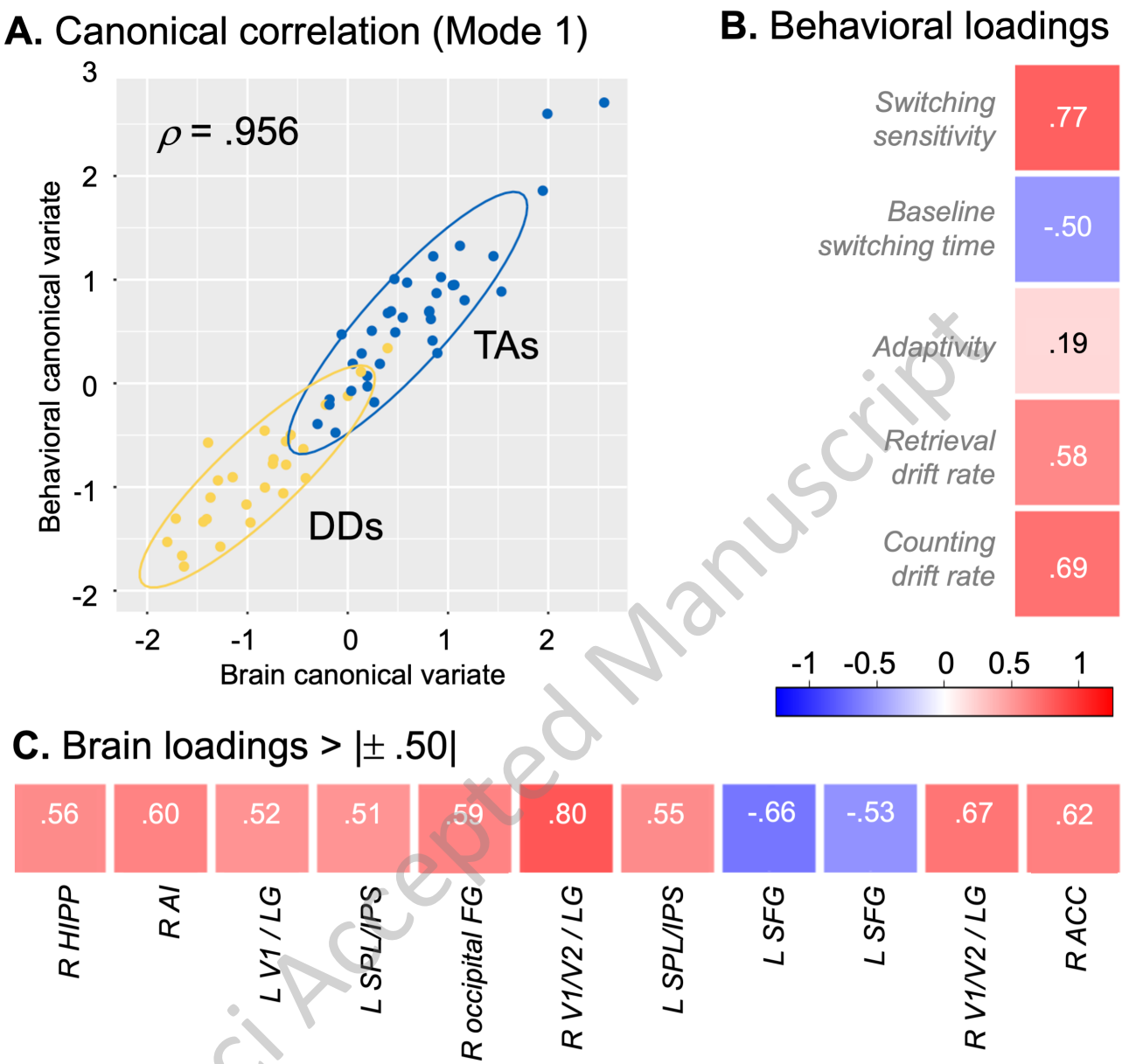


JNeurosci Accepted Manuscript

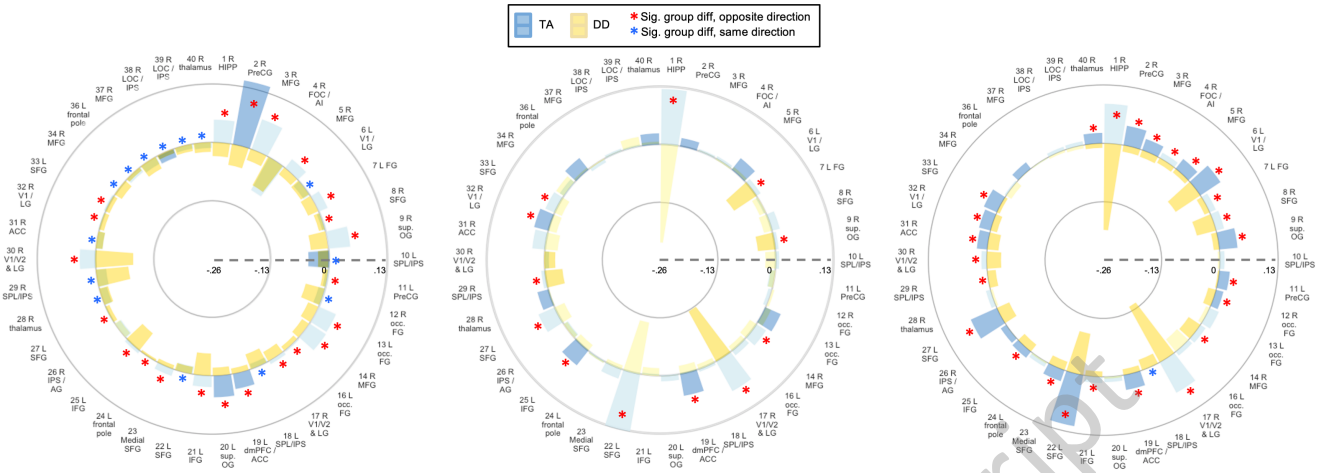
Spherical ROIs around most significant voxels in EN model trained with Union mask (α = 1)



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



Switching sensitivity

Counting drift rate (δ^C)

Decomposition drift rate (δ^D)

JNeurosci Accepted Manuscript