Quantitative modelling demonstrates format-invariant representations of mathematical problems in the brain

Tomoya Nakai | Shinji Nishimoto

Abstract
Mathematical problems can be described in either symbolic form or natural language. Previous studies have reported that activation overlaps exist for these two types of mathematical problems, but it is unclear whether they are based on similar brain representations. Furthermore, quantitative modelling of mathematical problem solving has yet to be attempted. In the present study, subjects underwent 3 h of functional magnetic resonance experiments involving math word and math expression problems, and a read word condition without any calculations was used as a control. To evaluate the brain representations of mathematical problems quantitatively, we constructed voxel-wise encoding models. Both intra- and cross-format encoding modelling significantly predicted brain activity predominantly in the left intraparietal sulcus (IPS), even after subtraction of the control condition. Representational similarity analysis and principal component analysis revealed that mathematical problems with different formats had similar cortical organization in the IPS. These findings support the idea that mathematical problems are represented in the brain in a format-invariant manner.

KEYWORDS
encoding, fMRI, IPS, mathematics, RSA

1 | INTRODUCTION

Mathematical problems can be presented in different formats. In symbolic form (math expression [ME]), mathematical problems are represented as a combination of digits and operators (e.g., ‘3 × 2 = 6’); in linguistic form (math word [MW]), the same mathematical problems are described in natural language (e.g., ‘There are 3 eggs in each box. How many eggs are in 2 boxes?’). Although
most mathematical problems are represented in symbolic ME format, the MW format is still ‘one of the most common materials in school curricula for teaching students to transfer mathematical knowledge into real-world contexts’ (Ng et al., 2021). Studies with a large sample of schoolchildren have found that performance in MW problems is related to later performance in more advanced algebra (Fuchs et al., 2012; Powell & Fuchs, 2014), suggesting that a shared cognitive basis of mathematical problem solving may exist across different formats.

In several neuroimaging studies, the format-independency of mathematical problem solving has been examined using univariate methods (Chang et al., 2019; Newman et al., 2011; Sohn et al., 2004; Zhou et al., 2018). Sohn et al. reported that MW problems induce activity in the left prefrontal cortex, whereas ME problems induce activity in the bilateral parietal cortex (Sohn et al., 2004). Newman et al. found evidence for shared activation of MW and ME problems in the bilateral intraparietal sulcus (IPS) and left frontal cortex (Newman et al., 2011). Zhou et al. tested MW problems with different complexities and found activations that included the left inferior frontal gyrus (IFG) and dorsomedial prefrontal cortex (Zhou et al., 2018). Chang et al. found that, compared with sentence comprehension, MW problems induced higher activation in the IPS and dorsolateral prefrontal cortex (Chang et al., 2019). Collectively, these studies suggest the involvement of the IPS and prefrontal cortex in both MW and ME problem solving. However, it remains unclear whether the two types of mathematical problems share brain representations. Furthermore, quantitative modelling of mathematical problem solving has yet to be attempted.

To address these issues, we used a voxel-wise encoding modelling approach (Naselaris et al., 2011). Encoding models predict brain activity in a quantitative manner based on a combination of features extracted from the presented stimuli. Researchers have adopted this approach to examine brain responses to visual (Çukur, Huth, et al., 2013; Çukur, Nishimoto, et al., 2013; Huth et al., 2012; Kay, Naselaris, et al., 2008; Nishimoto et al., 2011), auditory (De Angelis et al., 2018; Nakai, Koide-Majima, & Nishimoto, 2021), semantic (Huth et al., 2016; Kiremitçi et al., 2021; Nakai, Yamaguchi, & Nishimoto, 2021; Nishida et al., 2021; Shahdloo et al., 2022) and emotional (Horikawa et al., 2020; Koide-Majima et al., 2020) stimuli and to assess many complex cognitive functions (Nakai & Nishimoto, 2020). In addition, the approach allows for quantitative evaluation of the generalizability of models constructed under a certain format to data in another format (Nakai, Yamaguchi, & Nishimoto, 2021). Moreover, encoding models enable visualization of multivoxel patterns with different categorical information and evaluation of their similarity across different formats.

In the current study, subjects underwent 3-h functional magnetic resonance imaging (fMRI) experiments (Figure 1a) and each attempted to solve a series of mathematical problems in the MW and ME formats. As a control for non-math features such as visual attention, motor control and linguistic processing, we also included a read word (RW) condition, under which the subjects read the same sentences as under the MW condition but did not conduct calculations (see Methods for details). We prepared the control condition only for the MW condition as this condition would be more demanding than a control corresponding to the ME condition, thus suitable for eliminating a possible influence by features of noninterest. Using sparse operator features, we modelled brain activity specific to each mathematical problem (Figure 1b). By applying the cross-modal modelling technique developed in our previous study (Nakai, Yamaguchi, & Nishimoto, 2021), we defined a format invariance (FI) index, which quantifies how much information can be captured in a format-invariant manner in each brain region. In addition, representational similarity analysis (RSA) and principal component analysis (PCA) revealed representational relationships among different mathematical problems. Thus, through quantitative evaluations, this study provides new evidence of brain representations in math problem solving.

2 | MATERIALS AND METHODS

2.1 | Subjects

Eight healthy college students (aged 20–23 years, three females, all with normal vision), denoted as ID01–ID08, participated in this study. The number of subjects is in the same range as previous encoding model studies (Horikawa et al., 2020; Huth et al., 2016; Kay, Naselaris, et al., 2008; Nakai & Nishimoto, 2020; Nishimoto et al., 2011). The small number of subjects is compensated by a large number of samples for each subject (i.e., 3 h of tests). Subjects are all right-handed (laterality quotient = 80–100) as assessed using the Edinburgh inventory (Oldfield, 1971). Written informed consent was obtained from all subjects prior to their participation in the study, and the experiment was approved by the ethics and safety committee of the National Institute of Information and Communications Technology in Osaka, Japan.
2.2 Stimuli and testing procedure

Subjects performed arithmetic problems in two different formats, namely MW and ME conditions (see Figure S1 for the behavioural results). We selected MW problems with a single operation of addition (Add), subtraction (Sub), multiplication (Mul) and division (Div) from the IL dataset (Roy & Roth, 2015) and MW problems with two operations, including addition and subtraction (AddSub), addition and multiplication (AddMul), addition and division (AddDiv), subtraction and multiplication (SubMul) and subtraction and division (SubDiv), from the CC dataset (Roy et al., 2015). Each condition consisted of 35 instances. Both of the abovementioned datasets have been widely used in previous machine learning studies on MW problems (Mandal & Naskar, 2019). For our MW conditions, the original English sentences were translated into Japanese by the first author (T.N.). Corresponding arithmetic expression problems were created based on the MW conditions; thus, these conditions have the same numerical properties (Table 1).

In the MW condition, an arithmetic word problem (e.g., ‘There are 3 eggs in each box. How many eggs are in 2 boxes?’) was presented for 6 or 10 s (6 s for the single-operator problems [Add, Sub, Mul and Div] and 10 s for the double-operator problems [AddSub, AddMul, AddDiv, SubMul and SubDiv]). Subjects had access to a two-button response pad in their left hand, and they were instructed to perform a calculation based on the presented problem and press the left button when they had calculated an answer. After 1–2 s during which a fixation cross stimulus was displayed, a probe digit stimulus was presented for 2 s (e.g., ‘6’). Subjects were asked to press the left button if the presented digit matched their answer and the right button if it did not match their answer. The next trial started after 1–2 s, during which a time fixation cross stimulus was shown. The number of letters used in each problem was as follows: 61.3 ± 13.3 (Add), 57.0 ± 6.6 (Sub), 66.9 ± 12.0 (Div), 89.4 ± 13.4 (AddSub), 85.6 ± 13.1 (AddMul), 91.2 ± 16.3 (AddDiv), 83.8 ± 7.2 (SubMul) and 91.4 ± 13.6 (SubDiv).

In the ME condition, an arithmetic expression problem (e.g., ‘3 × 2 = ?’) was presented for 4 or 6 s (4 s for the single-operator problems [Add, Sub, Mul and Div], and 6 s for the double-operator problems [AddSub, AddMul, AddDiv, SubMul and SubDiv]). As in the MW
condition, the subjects were instructed to perform a calculation based on the presented problem and to press the left button when they had calculated an answer. The fixation cross stimulus was then presented for 1–2 s, after which a probe digit stimulus was presented for 2 s (e.g., ‘6’). Subjects were again asked to press the left or right button if the presented digit matched or did not match their answer, respectively. The next trial began after the fixation cross stimuli was again displayed for 1–2 s. The number of letters used in each problem was as follows: 5.5 ± 0.5 (Add), 6.5 ± 0.5 (Sub), 5.0 ± 0.2 (Mul), 5.9 ± 0.4 (Div), 9.3 ± 0.9 (AddSub), 8.9 ± 0.3 (AddMul), 10.1 ± 0.7 (AddDiv), 9.6 ± 0.5 (SubMul) and 10.9 ± 0.5 (SubDiv). The numerical problem size in each problem was as follows: 26.3 ± 22.9 (Add), 43.1 ± 16.1 (Sub), 11.0 ± 5.5 (Mul), 13.2 ± 8.4 (Div), 23.2 ± 11.1 (AddSub), 16.2 ± 6.1 (AddMul), 11.4 ± 6.7 (AddDiv), 14.0 ± 6.2 (SubMul) and 22.8 ± 9.5 (SubDiv).

In the RW condition, an arithmetic word problem was presented for 6 or 10 s as described above for the MW condition. The set of sentences presented was the same as that used in the MW condition (e.g., ‘There are 3 eggs in each box. How many eggs are in 2 boxes?’). However, subjects were instructed to carefully read the sentence and press the left button when they had memorized all non-numerical nouns. After a fixation cross stimulus was shown for 1–2 s, two words were presented for 2 s (e.g., ‘eggs’ and ‘boxes’). Subjects were asked to press the left button if both nouns were included in the original sentence and the right button if one of the two nouns were not included in the sentence. Two nouns were presented in the answer-matching phase to ensure that the subjects memorized all the nouns that appeared in the sentence. The next trial started after a fixation cross stimulus was displayed for 1–2 s.

Stimuli were presented on a projector screen inside the scanner (21.0 × 15.8° visual angle at 30 Hz). During scanning, subjects wore MR-compatible ear tips. Presentation software (Neurobehavioral Systems, Albany, CA, USA) was used to control the stimulus presentation and collection of behavioural data. To measure the button responses (BRs), optic response pads with two buttons were used (HHSC-2 × 2, Current Designs, Philadelphia, PA, USA).

The experiment was performed for 3 days, with six runs performed each day. In total, 18 runs were conducted, of which six training and two test runs were in the MW condition, and another six training and two test runs were in the ME condition. Two test runs were conducted in the RW condition. The presentation order was as follows: Day 1, MW, ME, MW, ME, MW and RW (57.5 min); Day 2, MW, ME, MW, ME, MW and ME (54.5 min); Day 3, RW, ME, MW, ME, MW and ME (54.5 min). A T1 anatomical image (6 min) was acquired.

### Table 1: List of example stimulus for each operation type.

| Problems  | MW condition | ME condition |
|-----------|--------------|--------------|
| **Add**   | ‘Carol collects 2 peanuts. Carol’s father gives Carol 5 more. How many peanuts does Carol have?’ | ‘2 + 5 = ?’ |
| **Sub**   | ‘Ernest starts with 17 crayons. Jennifer takes 6 away. How many crayons does Ernest end with?’ | ‘17 – 6 = ?’ |
| **Mul**   | ‘There are 3 eggs in each box. How many eggs are in 2 boxes?’ | ‘3 × 2 = ?’ |
| **Div**   | ‘The school is planning a field trip. There are 14 students and 2 seats on each school bus. How many buses are needed to take the trip?’ | ‘14/2 = ?’ |
| **AddSub** | ‘Faye and her mom were picking carrots from their garden. Faye picked 23 and her mother picked 5. If only 12 of the carrots were good, how many bad carrots did they have?’ | ‘(23 + 5) – 12 = ?’ |
| **AddMul** | ‘At the town carnival Billy rode the ferris wheel 7 times and the bumper cars 3 times. If each ride cost 5 tickets, how many tickets did he use?’ | ‘(7 + 3) × 5 = ?’ |
| **AddDiv** | ‘A toy store had 4 giant stuffed bears in stock when they got another shipment with 10 bears in it. The put the bears onto shelves with 7 on each shelf. How many shelves did they use?’ | ‘(4 + 10) / 7 = ?’ |
| **SubMul** | ‘A new building needed 14 windows. The builder had already installed 5 of them. If it takes 4 hours to install each window, how long will it take him to install the rest?’ | ‘(14 – 5) × 4 = ?’ |
| **SubDiv** | ‘A pet store had 18 puppies. In one day they sold 3 of them and put the rest into cages with 5 in each cage. How many cages did they use?’ | ‘(18 – 3) / 5 = ?’ |

Annotations: Add, addition; AddDiv, addition and division; AddMul, addition and multiplication; AddSub, addition and subtraction; Div, division; ME, math expression; Mul, multiplication; MW, math word; Sub, subtraction; SubDiv, subtraction and division; SubMul, subtraction and multiplication.
at the end of the first day. The presentation order was identical for all subjects. The data from each pair of test runs were averaged to increase the signal-to-noise ratio. Each run contained 45 trials consisting of five trials for each of the nine problems, and the presentation order within each run was randomized. The duration of a single run was 595 s for the MW condition and 455 s for the ME condition. At the beginning of each run, 10 s of dummy scans were acquired, during which the fixation cross was displayed, and these dummy scans were later omitted from the final analysis to reduce noise. We also obtained 10 s of scans at the end of each run, during which the fixation cross was displayed; however, these scans were included in the analyses.

2.3 | MRI data acquisition

The experiment was conducted using a 3.0 T scanner (MAGNETOM Prisma; Siemens, Erlangen, Germany) with a 64-channel head coil. We scanned 72 interleaved axial slices that were 2-mm thick without a gap, parallel to the anterior and posterior commissure line, using a T2*-weighted gradient echo multiband echo-planar imaging sequence (repetition time [TR] = 1000 ms; echo time [TE] = 30 ms; flip angle [FA] = 62°; field of view [FOV] = 192 × 192 mm²; resolution = 2 × 2 mm²; multiband factor = 6). We obtained 605 volumes for the MW and RW conditions and 465 volumes for the ME condition, with each set following 10 dummy images. For anatomical reference, high-resolution T1-weighted images of the whole brain were also acquired from all subjects with a magnetization-prepared rapid acquisition gradient echo sequence (TR = 2530 ms; TE = 3.26 ms; FA = 9°; FOV = 256 × 256 mm²; voxel size = 1 × 1 × 1 mm³).

2.4 | fMRI data preprocessing

Motion correction in each run was performed using the statistical parametric mapping toolbox (SPM12; Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). All volumes were aligned to the first EPI image for each subject. Low-frequency drift was removed using a median filter with a 120-s window. Slice timing correction was performed against the first slice of each scan. The response for each voxel was then normalized by subtracting the mean response and scaling it to the unit variance. We used FreeSurfer (https://surfer.nmr.mgh.harvard.edu/) to identify the cortical surfaces from the anatomical data and to register these to the voxels of the functional data. For each subject, the voxels identified in the cerebral cortex (59,499–75,980 voxels per subject) were used in the analysis.

2.5 | Operator features

The operator features were composed of one-hot vectors, which were assigned values of 1 or 0 for each time bin during the stimulus presentation of arithmetic problems, indicating whether one of the nine tested problems (Add, Sub, Mul, Div, AddSub, AddMul, AddDiv, SubMul and SubDiv) was performed in that period; thus, nine operator features were used.

2.6 | Motion energy features

We employed a Motion energy (MoE) model used in previous studies (Koide-Majima et al., 2020; Nakai & Nishimoto, 2020; Nishimoto et al., 2011) that can be found in a public repository (https://github.com/gallantlab/motion_energy_matlab). First, movie frames and pictures were spatially downsampled to 96 × 96 pixels. The RGB pixel values were then converted into the Commission International de l’Eclairage LAB colour space, and the colour information was discarded. The luminance (L*) pattern was passed through a bank of three-dimensional spatiotemporal Gabor wavelet filters; the outputs of the two filters with orthogonal phases (quadrature pairs) were squared and summed to yield local MoE. Subsequently, MoE was compressed with a log-transformation and temporally downsampled to 0.5 Hz. Filters were tuned to six spatial frequencies (0, 1.5, 3.0, 6.0, 12.0 and 24.0 cycles per image) and three temporal frequencies (0, 4.0 and 8.0 Hz) without directional parameters. Filters were positioned on a square grid that covered the screen. The adjacent filters were separated by 3.5 standard deviations of their spatial Gaussian envelopes. To reduce the computational load, the original MoE features, which had 1395 dimensions, were reduced to 300 dimensions using principal component analysis.

2.7 | Button response feature

The button response (BR) feature was constructed based on the number of button responses per second. One BR feature was used.
2.8 | Letter feature

The letter feature was constructed based on the number of letters that appeared in each stimulus. One letter feature was used.

2.9 | Reaction time feature

For each time bin during the presentation of a mathematical problem, the reaction time (RT) of that trial was assigned as a cognitive load of the trial. One RT feature was used.

2.10 | Accuracy feature

For each time bin during the presentation of a mathematical problem, the accuracy of the operator type of the given problem (for the given subject) was assigned as another cognitive load index. One accuracy feature was used.

2.11 | Encoding model fitting

In the encoding model, the cortical activity in each voxel was fitted with a finite impulse response model that captured the slow hemodynamic response and its coupling with neural activity (Kay, David, et al., 2008; Nishimoto et al., 2011). The feature matrix $F_E \in \mathbb{R}^{T \times 6N}$ was modelled by concatenating sets of $[T \times N]$ feature matrices with six temporal delays of 2–7 s ($T =$ number of samples; $N =$ number of features). The cortical response $R_E \in \mathbb{R}^{T \times V}$ was then modelled by multiplying the feature matrix $F_E$ by the weight matrix $W_E \in \mathbb{R}^{6N \times V}$ ($V =$ number of voxels):

$$\hat{R}_E = F_E W_E$$

We used an L2-regularized linear regression with the training dataset to obtain the weight matrix $W_E$. The training dataset consisted of 3630 (3630 s) and 2790 (2790 s) samples for the MW and ME conditions, respectively, whereas, a training dataset was not provided for the RW condition. The optimal regularization parameter was assessed using 10-fold cross-validation where the 11 different regularization parameters ranged from 1 to $2^{10}$.

The test dataset consisted of 605 samples (605 s) for the MW and RW conditions and 465 samples (465 s) for the ME condition (each repeated twice). Two repetitions of the test dataset were averaged to increase the signal-to-noise ratio. Statistical significance (one-sided) was computed by comparing estimated correlations to the null distribution of correlations between two independent Gaussian random vectors with the same length as the test dataset. The statistical threshold was set at $P < 0.05$ and corrected for multiple comparisons using the FDR procedure (Benjamini & Hochberg, 1995). For data visualization on the cortical maps, we used pycortex (Gao et al., 2015) and fsbrain (Schäfer & Ecker, 2020).

2.12 | Encoding model fitting excluding regressors of noninterest

To evaluate the possible effect of sensorimotor features and general cognitive load on the model predictions, we performed an additional encoding model fitting by excluding the features of noninterest. To this end, we concatenated the MoE (visual), BR (motor), RT (cognitive load), accuracy (cognitive load) and letter (orthographic) features. The concatenated features were used as a feature matrix for the encoding modelling. The original training dataset was divided into five subtraining runs and a single subtest run. Encoding models were trained using the subtraining dataset, and the prediction accuracy was calculated using the remaining subtest dataset. L2-regularized linear regressions were applied in the same manner as that described in the ‘Encoding model fitting’ subsection in the Methods. This procedure was repeated for all runs in the original training dataset. Cortical voxels significantly predicted in at least five out of six repetitions were regarded as non-math voxels and excluded from the analysis of target encoding models.

2.13 | Quantification of format invariance

To quantify how the encoding models explained brain activity in each voxel regardless of the presentation format, we defined FI values. In our previous study, we used a similar measurement to quantify how linguistic information is processed in a modality-invariant manner (Nakai, Yamaguchi, & Nishimoto, 2021). Note that the variance partitioning analysis (de Heer et al., 2017) cannot be used to analyse the format-invariance in the current experiment because this analysis compares explained variances of two (or more) different types of features extracted from the same dataset. In contrast, this experiment aims at comparing the same operator features extracted from two datasets. To quantify FI based on prediction accuracy, we used the geometric mean of prediction accuracy rather than the weight correlation. This
was justified because models with similar weight values have similar predictive performance. FI consisted of two components: $D_W$ and $D_E$. $D_W$ was defined as the degree of predictability for the MW test dataset regardless of the training format:

$$D_W = \sqrt{R_{WW} \cdot R_{EW}}$$

where $R_{WW}$ and $R_{EW}$ are the intra-format prediction accuracy for the MW model and the cross-format prediction accuracy for the ME model when applied to the test dataset for the MW condition, respectively. Similarly, $D_E$ was defined as the degree of predictability calculated for the ME test dataset regardless of the training format:

$$D_E = \sqrt{R_{WE} \cdot R_{EE}}$$

where $R_{EE}$ and $R_{WE}$ are the intra-format prediction accuracy using an encoding model trained with the ME condition and the cross-format prediction accuracy trained with the MW condition when applied to the test dataset for the ME condition, respectively. For all voxels with negative prediction accuracies, the prediction accuracy was reset to 0 to avoid obtaining imaginary values. FI was then calculated for each voxel as a geometric mean between $D_W$ and $D_E$ as follows:

$$FI = \sqrt{D_W \cdot D_E}$$

FI values ranged from 0 to 1, and a high FI indicates that the target features are represented in a format-invariant manner, whereas an FI of zero indicates that the target voxel does not have a shared representation for MW and ME conditions. To calculate the null distribution of FI values, a set of four correlation coefficients (produced by Gaussian random vector pairs) were used instead of $R_{WW}$, $R_{EE}$, $R_{EW}$ and $R_{WE}$. The statistical threshold was set at $P < 0.05$ and corrected for multiple comparisons using the FDR procedure (Benjamini & Hochberg, 1995).

### 2.14 Format specificity

To quantify how the encoding models explained brain activity that was specific for a certain format, we defined format specificity, which was calculated in each voxel for each format ($FS_W$ for the MW condition and $FS_E$ for the ME condition) as the difference between the intra-format and cross-format prediction accuracies:

$$FS_W = R_{WW} - R_{EW}$$

$$FS_E = R_{EE} - R_{WE}$$

FS value ranges from $-1$ to $1$. A high FS value indicates that the target features are represented specifically according to the target format, where a negative FS indicates that the target voxel does not have a format-specific representation. Significance and FDR corrections for multiple comparisons were calculated as described for the FI values.

### 2.15 RSA

For each of the target anatomical regions of interest (ROIs), operator similarity was calculated using the Pearson correlation distance between weight vectors of different operators extracted from the weight matrix of operator feature encoding models. We selected the voxels that showed significant FI values and averaged six time delays for each operator. An RSM was calculated based on all combinations of nine operator types, using concatenated feature vectors across all subjects. The upper triangular part of the RSM was rearranged into a single vector for both the MW and ME conditions, and Spearman’s correlation coefficients for correlations between the two rearranged vectors were calculated.

### 2.16 PCA

For each subject, we performed PCA on the weight matrix of the operator feature model concatenated across the eight subjects. We selected the voxels that showed significant FI values and averaged six time delays for each operator. We applied PCA to the $[9 \times V]$ weight matrices ($V =$ number of voxels) of the MW and ME encoding models. To show the structure of the representational space of mathematical problems, nine operators were mapped onto the two-dimensional space using the loadings of the first and second PCs, that is, PC1 and PC2, as the x-axis and y-axis, respectively. The upper triangular part of the RSM was rearranged into a single vector for both the MW and ME conditions, and Spearman’s correlation coefficients for correlations between the two rearranged vectors were calculated.
permutated the categorical labels of the MW and ME models 5000 times before applying PCA. \(P\) values of the actual correlation coefficients were evaluated according to the number of random correlation coefficients larger than the actual value across all permutations.

3 | RESULTS

3.1 | Encoding models predict brain activity in an intra- and a cross-format manner for each subject

To confirm that the encoding models successfully captured brain activity during the math problem solving, we performed a series of intra-format encoding modelling predictions and examined the prediction accuracy using a test dataset with the same format as that of the training dataset (Figure 1b, green). First, we trained an encoding model using the MW training dataset and predicted brain activity using the MW test dataset. Using operator features, the encoding model significantly predicted the activity in large brain networks, including the bilateral inferior frontal, parietal and occipital cortices (\(P < 0.05\), FDR corrected; \(40.1\% \pm 8.2\%\) of voxels were significant; Figure 2a blue; Table S1). Similarly, we trained encoding models using the ME training dataset and predicted brain activity using the ME test dataset. This encoding model significantly predicted a smaller part of the region, including the bilateral inferior frontal, bilateral parietal and bilateral occipital cortices, compared with that predicted by the MW intra-format encoding model (\(21.5\% \pm 9.4\%\) of voxels were significant; Figure 2b blue; Tables S1).

To determine whether the encoding models were generalizable to mathematical problems in other formats, we performed cross-format encoding model analyses. Specifically, we applied an encoding model trained with the ME training dataset to the MW test dataset; prediction accuracy was significant in the bilateral inferior frontal, bilateral parietal and bilateral occipital cortices (\(21.2\% \pm 8.2\%\) of voxels were significant; coloured in orange or white in Figure 2a; Table S1). Similarly, we applied the encoding model trained with the MW training dataset to the ME test dataset; prediction accuracy was also significant in the bilateral inferior frontal, bilateral parietal and bilateral occipital cortices (\(12.2\% \pm 4.8\%\) of voxels were significant; coloured in orange or white in Figure 2b; Table S1). In summary, our encoding models successfully predicted brain activity induced in large brain regions, including the parietal and frontal cortices, across different formats of mathematical problems.

3.2 | Format-invariant and format-specific processing of mathematical problems

To quantify how much information a certain model captures for each format of mathematical problem, we calculated FI indices by combining the intra-format and cross-
format prediction accuracies. FI was adopted based on modality invariance indices, which have been used previously to quantify modality-invariant processing of semantic information across text and speech modalities (Nakai, Yamaguchi, & Nishimoto, 2021). We found clusters of significant FI values distributed across the bilateral frontal, parietal and occipital cortices (19.7% ± 9.9% of voxels were significant across the whole cortex; Figure S2). To quantitatively evaluate the consistency of our results across test subjects, we calculated the ratio of significant voxels in each anatomical ROI of each subject and averaged across subjects. We focused on ROIs where >20% of voxels had a significant FI value. Suprathreshold FI values were found in large brain regions, including the bilateral prefrontal, parietal and occipital cortices (Figure 3a).

The original encoding models predict not only arithmetic-related regions but also regions related to sensorimotor and semantic information. To exclude regions of noninterest from our analyses, we applied encoding models trained in MW and ME conditions to the control RW condition. The average prediction accuracies of these two models were regarded as baseline accuracies. The baseline prediction accuracy was subtracted from the original intra- and cross-format prediction accuracies, and FI values were calculated. Consequently, we observed significant FI values predominantly in the left IPS (6.2% ± 3.8% of voxels were significant across the cortex; Figure 3b), indicating that format-invariant representations of mathematical problems are robust across different problem structures.

We used the RW condition (using the same visual stimuli as the MW condition) as a control for both MW and ME conditions because we assumed that the MW condition would induce larger activations than the ME condition. To confirm this assumption, we quantified FS values for both MW and ME conditions. FS was adopted based on modality specificity indices, which have been used previously to quantify modality-specific processing of semantic information for text or speech modalities (Nakai, Yamaguchi, & Nishimoto, 2021). We found above threshold FS values only for MW condition (Figure 3c). In contrast, no above-threshold FS values were found for the ME condition (Figure 3d), supporting our assumption that the RW condition serves as a control for both MW and ME conditions.

One concern regarding our analyses was whether the encoding model results could be explained solely by the problem complexity factor or also by cognitive load. To examine this further, we first constructed additional encoding models using only single- or double-operators (Figure S3); for both operator types, we found significant FI values in the bilateral IPS (10.4% ± 5.5% and 17.2% ± 8.1% of voxels were significant across the cortex, respectively), indicating that model predictability for the bilateral IPS was not a superficial effect due to differences in task difficulty.

---

**Figure 3** Format invariance (FI) and format specificity (FS).
(a) Group average of FI mapped onto the template brain. FI values were averaged in each anatomical region of interest (ROI) and averaged across eight subjects. Only ROIs with >20% significant voxels are shown. (b) Group average of FI with subtraction of the prediction accuracy in the read word (RW) condition. (c, d) Group average of FS for (c) math word (MW) and (d) math expression (ME), mapped onto the template brain. IPS, intraparietal sulcus; L, left hemisphere.

---

NAKAI and NISHIMOTO
between problem complexities. To further investigate our concern, we performed encoding modelling after excluding cortical voxels that can be predicted by features of noninterest (i.e., visual, motor, orthography and cognitive load; see Section 2 for detail). We calculated intra-format predictions using a leave-one-run-out method within the training dataset. Voxels that were consistently predicted in either the MW or ME condition were excluded from FI analysis; we found significant FI values in the bilateral IPS even after excluding these voxels (15.4% ± 7.2% of voxels were significant; Figure S4), indicating that the FI values in the abovementioned analyses were not explained by the effect of sensorimotor processing or general cognitive load.

3.3 | Similarities in brain representations related to different mathematical problems

To investigate representational differences between different mathematical problems, we performed RSA based on the encoding model weights. In each anatomical ROI, we calculated a representational similarity matrix (RSM) for both MW and ME conditions using Pearson’s correlation distance (Figure 4a). We first confirmed similar correlation patterns across subjects for MW (left IPS, $\rho = 0.436 \pm 0.244$, Wilcoxon signed-rank test, $P < 0.001$; right IPS, $\rho = 0.541 \pm 0.187$, $P < 0.001$) and ME conditions (left IPS, $\rho = 0.491 \pm 0.198$, $P < 0.001$; right IPS, $\rho = 0.310 \pm 0.171$, $P < 0.001$). We thereby compared the resultant RSMs between the two formats and found significant correlations between two RSMs in the bilateral parietal and occipital cortices (Figure 4b). In particular, we found a positive correlation between the representational similarities of MW and ME conditions in the bilateral IPS (Figure 4c,d). The significant correlation existed in the left IPS even after excluding the voxels predicted under the control RW condition (Figure S5), indicating that the similarity of the representational relationships among different operators provide the basis of format invariance.

3.4 | Visualization of brain representations related to different mathematical problems

Both FI analysis and RSA revealed shared brain representations for MW and ME conditions. To visualize the related organizations, we applied PCA to the weight

---

**FIGURE 4** Representational similarity analysis. (a) A representational similarity matrix (RSM) was calculated using correlation distances between weight vectors extracted from the target region of interest (ROI) concatenated across all subjects. Spearman’s correlation analysis, based on the upper triangular parts of the RSM in the math word (MW) and math expression (ME) conditions, was conducted for each target ROI. (b) Correlation coefficients between RSMs in the MW and ME conditions calculated in each ROI, mapped onto the template brain. Only ROIs with a significant correlation coefficient are shown ($P < 0.05$ with Bonferroni correction). (c, d) Scatter plot of the normalized similarities under the MW and ME conditions in the (c) left and (d) right intraparietal sulcus (IPS). Dots represent elements in the RSMs (surrounded by red lines in a).
matrices of encoding models. PCA was also based on the whole cortical voxels. We found that nine tested problems had similar organizations in the MW and ME conditions (Figure 5a,b). The representational map visualized for the first and second PCs (PC1 and PC2) showed that double-operator problems were located in the bottom-right area, whereas single-operator problems were located in the upper-left area. Such organization was common across the MW and ME formats, and the mathematical problems were similarly organized even when we used only the bilateral IPS voxels (Figure S6). To evaluate the similarity between the representational space of MW and ME conditions in a quantitative manner, we calculated Spearman’s correlation coefficients between the PCA loading vectors of two formats. For the top two PCs, the PCA loading vectors were positively correlated between MW and ME conditions (PC1 and PC2 were significant using a permutation test, $P < 0.05$; Figure 5c). This effect was also observed using the bilateral IPS voxels (PC1 and PC2 were significant using a permutation test, $P < 0.05$; Figure 5d and S6). Thus, the nine tested mathematical problems were organized in similar brain representational spaces regardless of their format.

4 | DISCUSSION

To the best of our knowledge, all previous neuroimaging studies on mathematical problem solving have used univariate analysis to assess the format-independency of mathematical problems (Chang et al., 2019; Newman et al., 2011; Sohn et al., 2004; Zhou et al., 2018). Thus, it had not been made clear whether the activation overlap of mathematical problems with different formats reflected shared cortical representations. In contrast to univariate analysis, multivariate analysis (i.e., RSA) can test whether activation patterns are similar across different formats. Notably, training and test datasets had not previously been analysed separately nor had the generalizability of quantitative models to new data been evaluated. In the current study, we demonstrated for the first time that multivariate analyses and encoding model approaches are useful for studying the format-invariance of mathematical problems.

The bilateral IPS is involved in numerical cognition (Cantlon et al., 2006; Cohen Kadosh et al., 2005; Eger et al., 2003; Piazza et al., 2007) and processing complex mathematical problems (Nakai & Okanoya, 2020;...
Nakai & Sakai, 2014). Although format-invariance of quantity information has been questioned by several researchers (Bulthé et al., 2014, 2015; Lyons et al., 2015; Lyons & Beilock, 2018), we provide evidence that complex mathematical problems are processed in this region in a format-invariant manner. In particular, the left-hemisphere dominance of the IPS in the current study might reflect the specific nature of format-invariance in the case of such complex problems. Note that our results do not indicate format-invariance of symbolic and non-symbolic numbers, as both MW and ME problems used in the current study adopted symbolic numbers based on previous studies (Chang et al., 2019; Ng et al., 2021; Sohn et al., 2004; Zhou et al., 2018). Without controlling for the RW condition, we also found FI in the bilateral inferior frontal gyrus (IFG) (Figure 3a). The involvement of the left IFG in mathematical problems has been reported in previous studies (Nakai et al., 2017; Nakai & Okanoya, 2018, 2020; Nakai & Sakai, 2014), and the functional and anatomical connectivity between the IFG and IPS may play an essential role in the development of mathematical ability (Chang et al., 2019; Emerson & Cantlon, 2012; Rosenberg-Lee et al., 2015; Tsang et al., 2009). Thus, our findings on the IFG and IPS are consistent with the results of other mathematical cognition studies.

Our PCA visualization showed that single- and double-operator problems were distinctly clustered, which is consistent with the result in a previous study comparing single- and double-operator problems (Prabhakaran et al., 2001). Researchers have argued that activations in the bilateral IPS during mathematical problem solving are modulated by problem difficulty as measured using RTs (Chang et al., 2019). Although our PCA results seemed to reflect a difference between single- and double-operator problems (Figure 5), the representational space we obtained after controlling for the RW condition also had a similar organization between single- and double-operator problems (Figure S6). In particular, certain double-operator problems were closely located based on the shared single-operator, for example, the AddDiv and SubDiv problems and the AddMul and SubMul problems were paired. Our results for operator-specific representations are consistent with previous reports of double dissociation between different operators (Dehaene & Cohen, 1997) and operator-selective neurons (Kutter et al., 2022), which would contribute to the format-invariant predictability of cortical activations.

This study also revealed format-specific aspects of the MW and ME conditions. The MW condition is based on natural language sentences and might require the subjects more visual attention and cognitive load than the ME condition. The format-specificity analysis visibly reflected such an influence (Figure 3c,d), indicating a more relevant contribution of the bilateral frontal, temporal and occipital regions for the MW condition than for the ME condition. In contrast, it is important that the format-invariance was demonstrated in the IPS despite the seemingly large difference between the two conditions. Our method successfully revealed the latent structure of different operators behind the superficial form in which the mathematical problems were presented.

While the overall representations of math operations are similar across the subjects, we also observed notable individual variability. For instance, subjects ID03 and ID06 exhibited relatively smaller FI values than the other subjects (Figure S2). This might be caused by their lower signal-to-noise ratios (Table S1), as their percentage of significantly predicted voxels was smaller than that of the other subjects even for intra-format analyses. There might be room for improving the FI value definition in that it is affected by intra-format prediction accuracy values.

It is also worth noting that some individuals might use different strategies for solving mathematical problems. Although we observed no clear difference in RTs in the addition and multiplication problems (Figure S1), it is possible to solve multiplication by a series of additions. Moreover, some children (and even a few adults) tend to use a counting strategy rather than a memory-based strategy for solving addition problems (Qin et al., 2014), which can be a source of individual variability. Showing a representational similarity between individuals would support the existence of a shared brain mechanisms underlying mathematical problems with seemingly different structures.

Although the current study paves the way toward quantitative modelling of complex mathematical problems, we acknowledge that it has several limitations. First, because we chose experimental stimuli from existing datasets (Roy et al., 2015; Roy & Roth, 2015), we could not perfectly balance the number of letters and digit sizes. Although we excluded such effects by subtracting baseline data determined using the RW condition with the same visual stimuli (Figure 3b) and by explicitly regressing out features of noninterest (Figure S4), the between-condition difference may still have had some influence on the representational organization of the tested math problems.

Second, although we validated our results for each subject (Figure S2) and the test sample size (N = 465 for the ME and N = 605 for the MW conditions), as well as the number of subjects, were consistent with those of previous studies (Horikawa et al., 2020; Huth et al., 2016; Kay, Naselaris, et al., 2008; Nakai & Nishimoto, 2020; Nishimoto et al., 2011), the relatively small number of
subjects tested may limit the generalizability of our results to the broader population and other age groups.

Third, the categorical model of mathematical operations (i.e., operator features) may not fully capture the complex processes involved in math problem solving, especially the quantity information used in the problems. Indeed, the numerical problem size was not controlled across problems in the current experiment (see Section 2). However, it is unlikely that the numerical information is the sole factor accounting for the format invariance since we observed an opposite tendency between the problem size (Single-operators > Double-operators) and the behaviour and neural results (Figures S1 and S3). In addition, the subtraction analysis with the control RW condition (Figure 3c) indicates that the current findings are independent of numerical symbols since the RW condition contains the same numerical stimuli as the other conditions. Although further studies are warranted to fully clarify the individual differences and detailed subprocesses involved in mathematical problems, based on the current results, we conclude that the bilateral IPS network subserves format-invariant representations of mathematical problems.

AUTHOR CONTRIBUTIONS
Tomoya Nakai designed the study, wrote the manuscript and collected and analysed the data. Shinji Nishimoto helped design the study and write the manuscript.

ACKNOWLEDGEMENTS
We thank MEXT/JSPS KAKENHI (grant numbers JP17K13083 and JP18H05091 in #4903 [Evolinguistics] for T.N. and JP15H05311 and JP18H05522 for S.N.) as well as JST CREST (JPMJCR18A5), ERATO (JPMJER1801) (for S.N.) and H2020 Marie Skłodowska-Curie Actions (grant number 101023033 for T.N.) for their partial financial support of this study. The funders played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTEREST STATEMENT
The authors declare no competing interests.

DATA AVAILABILITY STATEMENT
The source data and analysis code used in the current study are available from Zenodo (https://doi.org/10.5281/zenodo.7572919).

ORCID
Tomoya Nakai https://orcid.org/0000-0001-5225-0894

REFERENCES
Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*, 289–300.

Bulté, J., De Smedt, B., & Op de Beeck, H. P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage, 87*, 311–322. https://doi.org/10.1016/j.neuroimage.2013.10.049

Bulté, J., De Smedt, B., & Op de Beeck, H. P. (2015). Visual number beats abstract numerical magnitude: Format-dependent representation of Arabic digits and dot patterns in human parietal cortex. *Journal of Cognitive Neuroscience, 27*, 1376–1387. https://doi.org/10.1162/jocn_a_00787

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology, 4*, e125. https://doi.org/10.1371/journal.pbio.0040125

Chang, T.-T., Lung, T.-C., Ng, C.-T., & Metcalfe, A. W. S. (2019). Fronto-insular-parietal network engagement underlying arithmetic problem solving. *Human Brain Mapping, 40*, 1927–1941. https://doi.org/10.1002/hbm.24502

Cohen Kadosh, R., Henik, A., Rubinsten, O., Mohr, H., Dori, H., van de Ven, V., Zorzi, M., Hendler, T., Goebel, R., & Linden, D. E. J. (2005). Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia, 43*, 1238–1248. https://doi.org/10.1016/j.neuropsychologia.2004.12.017

Çukur, T., Huth, A. G., Nishimoto, S., & Gallant, J. L. (2013). Functional subdomains within human FFA. *The Journal of Neuroscience, 33*, 16748–16766. https://doi.org/10.1523/JNEUROSCI.1259-13.2013

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience, 16*, 763–770. https://doi.org/10.1038/nn.3381

De Angelis, V., De Martino, F., Moerel, M., Santoro, R., Hausfeld, L., & Formisano, E. (2018). Cortical processing of pitch: Model-based encoding and decoding of auditory fMRI responses to real-life sounds. *NeuroImage, 180*, 291–300. https://doi.org/10.1016/j.neuroimage.2017.11.020

De Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *The Journal of Neuroscience, 37*, 6539–6557. https://doi.org/10.1523/JNEUROSCI.3267-16.2017

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex, 33*, 219–250. https://doi.org/10.1016/S0010-9452(08)70002-9

Eger, E., Sterzer, P., Russ, M. O., Giraud, A.-L., & Kleinschmidt, A. (2003). A supramodal number representation in human intra-parietal cortex. *Neuron, 37*, 719–725. https://doi.org/10.1016/S0896-6273(03)00036-9

Emerson, R. W., & Cantlon, J. F. (2012). Early math achievement and functional connectivity in the fronto-parietal network. *Developmental Cognitive Neuroscience, 2*(Suppl 1), S139–S151. https://doi.org/10.1016/j.dcn.2011.11.003

Fuchs, L. S., Compton, D. L., Fuchs, D., Powell, S. R., Schumacher, R. F., Hamlett, C. L., Verner, E., Namkung, J. M., & Vukovic, R. K. (2012). Contributions of
domain-general cognitive resources and different forms of arithmetic development to pre-algebraic knowledge. Developmental Psychology, 48, 1315–1326. https://doi.org/10.1037/a0027475

Gao, J. S., Huth, A. G., Lescroart, M. D., & Gallant, J. L. (2015). Pycortex: An interactive surface visualizer for fMRI. Frontiers in Neuroinformatics, 9, 23. https://doi.org/10.3389/fninf.2015.00023

Horiwata, K., Cowen, A. S., Keltner, D., & Kamilani, Y. (2020). The neural representation of visually evoked emotion is high-dimensional, categorical, and distributed across transmodal brain regions. iScience, 23, 101060. https://doi.org/10.1016/j.isci.2020.101060

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. Nature, 532, 453–458. https://doi.org/10.1038/nature17637

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron, 76, 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014

Kay, K. N., David, S. V., Prenger, R. J., Hansen, K. A., & Gallant, J. L. (2008). Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. Human Brain Mapping, 29, 142–156. https://doi.org/10.1002/hbm.20379

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. Nature, 452, 352–355. https://doi.org/10.1038/nature06713

Kiremitçi, I., Yilmaz, Ö., Çelik, E., Shahdloo, M., Huth, A. G., & Çukur, T. (2021). Attentional modulation of hierarchical speech representations in a multitalker environment. Cerebral Cortex, 31, 4986–5005. https://doi.org/10.1093/cercor/bhab136

Koide-Majima, N., Nakai, T., & Nishimoto, S. (2020). Distinct dimensions of emotion in the human brain and their representation on the cortical surface. NeuroImage, 222, 117258. https://doi.org/10.1016/j.neuroimage.2020.117258

Kutter, E. F., Boström, J., Elger, C. E., Nieder, A., & Mormann, F. (2022). Neuronal codes for arithmetic rule processing in the human brain. Current Biology, 32, 1275–1284. https://doi.org/10.1016/j.cub.2022.01.054

Lyons, I. M., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and non-symbolic numbers in the human brain. Human Brain Mapping, 36, 475–488. https://doi.org/10.1002/hbm.22641

Lyons, I. M., & Beilock, S. L. (2018). Characterizing the neural coding of symbolic quantities. NeuroImage, 178, 503–518. https://doi.org/10.1016/j.neuroimage.2018.05.062

Mandal, S., & Naskar, S. K. (2019). Solving arithmetic mathematical word problems: A review and recent advancements. In Information technology and applied mathematics (pp. 95–114). Springer Singapore. https://doi.org/10.1007/978-981-10-7590-2_7

Nakai, T., Koide-Majima, N., & Nishimoto, S. (2021). Correspondence of categorical and feature-based representations of music in the human brain. Brain and Behavior: A Cognitive Neuroscience Perspective, 11, e01936. https://doi.org/10.1002/brb3.1936

Nakai, T., Nakatani, H., Hosoda, C., Nonaka, Y., & Okanoya, K. (2017). Sense of accomplishment is modulated by a proper level of instruction and represented in the brain reward system. PLoS ONE, 12, e0168661. https://doi.org/10.1371/journal.pone.0168661

Nakai, T., & Nishimoto, S. (2020). Quantitative models reveal the organization of diverse cognitive functions in the brain. Nature Communications, 11, 1142. https://doi.org/10.1038/s41467-020-14913-w

Nakai, T., & Okanoya, K. (2018). Neural evidence of cross-domain structural interaction between language and arithmetic. Scientific Reports, 8, 12873. https://doi.org/10.1038/s41598-018-31279-8

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

Nakai, T., & Sakai, K. L. (2014). Neural mechanisms underlying the computation of hierarchical tree structures in mathematics. PLoS ONE, 9, e111439. https://doi.org/10.1371/journal.pone.011439

Nakai, T., Yamaguchi, H. Q., & Nishimoto, S. (2021). Convergence of modality invariance and attention selectivity in the cortical semantic circuit. Cerebral Cortex, 31, 4825–4839. https://doi.org/10.1093/cercor/bhab125

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. NeuroImage, 56, 400–410. https://doi.org/10.1016/j.neuroimage.2010.07.073

Newman, S. D., Willoughby, G., & Bruce, B. (2011). The effect of problem structure on problem-solving: An fMRI study of word versus number problems. Brain Research, 1410, 77–88. https://doi.org/10.1016/j.brainres.2011.06.053

Ng, C.-T., Lung, T.-C., & Chang, T.-T. (2021). Operation-specific lexical consistency effect in fronto-insular-parietal network during word problem solving. Frontiers in Human Neuroscience, 15, 631438. https://doi.org/10.3389/fnhum.2021.631438

Nishida, S., Blanc, A., Maeda, N., Kado, M., & Nishimoto, S. (2021). Behavioral correlates of cortical semantic representations modeled by word vectors. PLoS Computational Biology, 17, e1009138. https://doi.org/10.1371/journal.pcbi.1009138

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. Current Biology, 21, 1641–1646. https://doi.org/10.1016/j.cub.2011.08.031

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. Neuron, 53, 293–305. https://doi.org/10.1016/j.neuron.2006.11.022

Powell, S. R., & Fuchs, L. S. (2014). Does early algebraic reasoning differ as a function of Students’ difficulty with calculations versus word problems? Learning Disabilities Research and Practice, 29, 106–116. https://doi.org/10.1111/ldrp.12037

Prabhakaran, V., Rypma, B., & Gabrieli, J. D. (2001). Neural substrates of mathematical reasoning: A functional magnetic resonance imaging study of neocortical activation during performance of the necessary arithmetic operations test. Neuropsychology, 15, 115–127. https://doi.org/10.1037/0894-4105.15.1.115

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

Rosenberg-Lee, M., Ashkenazi, S., Chen, T., Young, C. B., Geary, D. C., & Menon, V. (2015). Brain hyper-connectivity and operation-specific deficits during arithmetic problem solving in children with developmental dyscalculia. *Developmental Science*, 18, 351–372. https://doi.org/10.1111/desc.12216

Roy, S. & Roth, D. (2015). Solving general arithmetic word problems. In EMNLP. pp. 1743–1752.

Roy, S., Vieira, T., & Roth, D. (2015). Reasoning about quantities in natural language. *Transactions of the Association for Computational Linguistics*, 3, 1–13. https://doi.org/10.1162/tacl_a_00118

Schäfer, T. & Ecker, C. (2020). fsbrain: An R package for the visualization of structural neuroimaging data. bioRxiv, https://doi.org/10.1101/2020.09.18.302935

Shahdloo, M., Çelik, E., Ürgen, B. A., Gallant, J. L., & Çukur, T. (2022). Task-dependent warping of semantic representations during search for visual action categories. *The Journal of Neuroscience*, 42, 6782–6799. https://doi.org/10.1523/JNEUROSCI.1372-21.2022

Sohn, M.-H., Goode, A., Koedinger, K. R., Stenger, V. A., Fissell, K., Carter, C. S., & Anderson, J. R. (2004). Behavioral equivalence, but not neural equivalence—Neural evidence of alternative strategies in mathematical thinking. *Nature Neuroscience*, 7, 1193–1194. https://doi.org/10.1038/nn1337

Tsang, J. M., Dougherty, R. F., Deutsch, G. K., Wandell, B. A., & Ben-Shachar, M. (2009). Frontoparietal white matter diffusion properties predict mental arithmetic skills in children. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 22546–22551. https://doi.org/10.1073/pnas.0906094106

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

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Nakai, T., & Nishimoto, S. (2023). Quantitative modelling demonstrates format-invariant representations of mathematical problems in the brain. *European Journal of Neuroscience*, 57(6), 1003–1017. https://doi.org/10.1111/ejn.15925