Received: 24 March 2023 | Revised: 4 July 2023 | Accepted: 5 July 2023

DOI: 10.1111/ejn.16100

RESEARCH REPORT

EJN European Journal of Neuroscience FENS WILEY

# Arithmetic problem size modulates brain activations in females but not in males

**Nai-Feng Chen**[1] | **Ting-Ting Chang**[1,2]

[1]Department of Psychology, National Chengchi University, Taipei, Taiwan

[2]Research Center for Mind, Brain and Learning, National Chengchi University, Taipei, Taiwan

**Correspondence**
Ting-Ting Chang, Department of Psychology, National Chengchi University, No. 64, Sec. 2, ZhiNan Rd, Wenshan District, Taipei City 11605, Taiwan.
Email: ttchang@nccu.edu.tw

**Funding information**
National Science Council, Grant/Award Numbers: 110-2629-H-004-001, 111-2811-H-004-004

Edited by: John Foxe

**Abstract**

Numerous empirical studies have reported that males and females perform equally well in mathematical achievement. However, still to date, very limited is understood about the brain response profiles that are particularly characteristic of males and females when solving mathematical problems. The present study aimed to tackle this issue by manipulating arithmetic problem size to investigate functional significance using functional magnetic resonance imaging (fMRI) in young adults. Participants were instructed to complete two runs of simple calculation tasks with either large or small problem sizes. Behavioural results suggested that the performance did not differ between females and males. Neuroimaging data revealed that sex/gender-related patterns of problem size effect were found in the brain regions that are conventionally associated with arithmetic, including the left middle frontal gyrus (MFG), left intraparietal sulcus (IPS) and insula. Specifically, females demonstrated substantial brain responses of problem size effect in these regions, whereas males showed marginal effects. Moreover, the machine learning method implemented over the brain signal levels within these regions demonstrated that sex/gender is discriminable. These results showed sex/gender effects in the activating patterns varying as a function of the distinct math problem size, even in a simple calculation task. Accordingly, our findings suggested that females and males use two complementary brain resources to achieve equally successful performance levels and highlight the pivotal role of neuroimaging facilities in uncovering neural mechanisms that may not be behaviourally salient.

**KEYWORDS**
fMRI, logistic regression, machine learning, mental arithmetic, sex/gender effect

## 1 | INTRODUCTION

Over the past decades, empirical studies have reached the consensus that males and females perform equally well in arithmetic learning and mathematical achievement (Chang et al., 2022; Hyde, 2014). Using large-scale meta-analytical analyses approach over millions of global participants, multiple studies have shown that sex/gender effects in mathematical performance, regardless of the contents, are subtle and negligible (Hyde et al., 1990, 2008; Lindberg et al., 2010). In contrast to the considerable focus on behavioural examination, understanding whether there are distinct neural mechanisms that are particularly characteristic of each sex/gender in mathematical problem-solving is still relatively limited. In this study, we systematically investigate brain activations underlying mathematical problem solutions of each sex by applying functional magnetic resonance imaging (fMRI) techniques. Our primary goal is to report the phenomenal brain responses of each sex/gender using the fMRI technique rather than arguing whether sex/gender shows specificity. Because it is challenging to disentangle that the group-level differences identified between males and females are influenced by hereditary and genetic biological factors or learned from experience and environment, throughout the manuscript, we adopt the term 'sex/gender differences' (Eliot et al., 2021; Jordan-Young & Rumiati, 2012; Springer et al., 2012) to capture both the biological mechanisms and the psychosocial expression of maleness and feminineness.

During mathematical problem solution, accumulative neuroimaging studies have consistently identified several regions activated. The vast majority of neuroimaging studies addressing math problem-solving had emphasized arithmetic problem-solving and calculation skills. These identified regions primarily encompass the anterior insula (AI), dorsal anterior cingulate cortex (dACC), dorsal posterior parietal cortex (PPC) and dorsolateral prefrontal cortex (DLPFC; Arsalidou & Taylor, 2011; Chang et al., 2016; Houde et al., 2010; Menon et al., 2014; Ng et al., 2021). The PPC is considered to play the most crucial role in representing and manipulating quantitative information (Ansari, 2008; Cohen Kadosh et al., 2008; Dehaene et al., 2003). The AI coupling with dACC is often associated with subjective salience of external stimuli and in contributions to complex cognitive processes, including central executive function and affective processing (Menon, 2015b; Seeley et al., 2007). DLPFC, together with PPC, engaged in information retention and manipulation during working memory, manipulation of quantities over epochs, construction of problem solutions and decision-making (Chang et al., 2019; Menon, 2015a; Miller & Cohen, 2001; Petrides, 2005; Rottschy et al., 2012). Collectively, these regions jointly contribute to the core neural substrates of numerical problem-solving skills, ranging from simple number comparisons to complex arithmetic and problems that require mathematical reasoning across the essential learning stage (Chang et al., 2019; Cho et al., 2012; Rosenberg-Lee et al., 2011, 2015; Supekar & Menon, 2012).

The above-reviewed arithmetic-associated brain regions are modulated by numerical properties, for example, problem size (Chang et al., 2015, 2016; De Smedt et al., 2011; Metcalfe et al., 2013; Stanescu-Cosson et al., 2000). The problem size effect refers to the cognitive loading cost such that arithmetic problems with larger problem operands (e.g., $7 + 9$; $6 \times 8$) were responded less accurately and slower than problems with smaller operands (e.g., $2 + 3$; $2 \times 4$; Campbell & Xue, 2001; De Smedt et al., 2011; Stanescu-Cosson et al., 2000). The effect of problem size likely reflects the specificity of strategy usage in distinct problem types. In particular, small problems are usually solved by fast-retrieving arithmetic knowledge facts, while large problems are solved by reasoning through the process of multistep calculations (Barrouillet et al., 2008; Campbell & Xue, 2001; De Smedt et al., 2011). Aside from the behaviour findings, neural correlates of the problem size effect are also documented (De Smedt et al., 2011; Stanescu-Cosson et al., 2000). Stanescu-Cosson and colleagues demonstrated that large arithmetic problems engage more activations over the DLPFC and bilateral intraparietal sulcus (IPS). In contrast, small problems inversely engage stronger angular gyrus (AG) within the PPC than large problems (Stanescu-Cosson et al., 2000). Several other studies have also reported similar results with school-age children (Chang et al., 2015, 2016), with the exception that in children, it is the hippocampus rather than the AG that shows stronger activations for small problems (Cho et al., 2012; De Smedt et al., 2011). Together, these research findings have demonstrated clear strategy distinctions in the neural mechanisms involved in solving mathematical problems of varying sizes, with solving large problems relying on procedural computations and the allocation of working memory resources, whereas solving small problems relying on fast, rapid mathematical fact retrieval.

Despite the comparable level of performance, several behavioural studies have indicated that there are disparities in strategy utilization between males and females (Bailey et al., 2012; Carr & Davis, 2001; Gallagher et al., 2000; Hirnstein et al., 2009; Hong & Aqui, 2004; Voyer & Saunders, 2004; Zhu, 2007). In the early studies conducted on first graders, Carr and colleagues found that during free-choice strategy selection conditions, females tend to adopt procedural calculation strategies, whereas males are more likely to use retrieval strategies when solving arithmetic problems (Carr & Davis, 2001; Carr & Jessup, 1997). Furthermore, when the strategy was constrained to retrieval, females became less accurate (Carr & Davis, 2001). Bailey et al. (2012) reported that across first to sixth grades, males had higher preference for using retrieval strategy to solve single-digit addition problems. Gallagher et al. (2000) observed a similar tendency when administering Scholastic Assessment Test–Mathematics (SAT-M) to junior and senior high school students.

Critically, although overall correctness is equivalent among genders, female students were more likely to solve problems using school-taught algorithms, whereas male students had a higher chance to solve logical problems that required insights. Note that none of these above studies reported differences in overall error rates, indicating that males and females are likely adopting different strategies to achieve comparable performance.

In the current study, we attempt to systematically examine brain response profiles during mathematical problem solution of males and females by collecting fMRI data from adults who were proficient in general arithmetic problem-solving skills. In order to probe brain responses elicited by different problem-solving strategies while maintaining participants' compliance, we manipulate problem size as large and small problems because the above-reviewed arithmetic-associated brain regions, that is, the PFC, PPC, AI and dACC, had been consistently identified as a function of the problem size. Our study intended to systematically manipulate problem size and provide a genuine effect of problem size in interpreting sex/gender effect in the brain response profiles. Based on the existing mathematics-related assessment reports, we predicted that behaviour performances would not show differences in this task. Instead, we hypothesized that sex/gender differences would be observed in the mathematical learning-associated brain regions. More specifically, we expected that the fronto-parietal engagement elicited by large problems would be more enhanced in females because they tend to adopt more concrete, algorithmic calculations during a vital mathematical task. Males, in contrast, are expected to show less salient brain responses within the fronto-parietal regions, given that they are prone to solve mathematical problems using fast rote-fact retrieving, estimation and insight strategies (Bailey et al., 2012; Gallagher et al., 2000; Zhu, 2007). A machine learning logistic regression model will also be administered to ensure the discrimination of the brain responses between males and females.

## 2 | METHODS

### 2.1 | Participants

Seventy-five adults (38 females and 37 males) were recruited from local educational institutions in Taipei City, Taiwan. Among the participating adults, four had excessive head movements (for the movement exclusion criteria, see Section 2.4), resulting in the final sample of 71 participants (36 females; age range 18.93–29.09 years, $M = 23.04$, $SE = .28$). Mean ages did not differ between females ($M = 22.60$, $SE = .37$) and males ($M = 23.49$, $SE = .41$) ($t_{(69)} = 1.62$, $p = .11$, 95% confidence interval [CI] = [−1.99, .21], $d = .38$). Because no previous studies had adopted the same design to investigate sex/gender effect using fMRI experiments, our estimation of power analysis is based on previous studies examining behavioural problem size effect using arithmetic tasks, which had effect sizes ranging from 1.53 to 1.72 (Tiberghien et al., 2019). To reach the power > 80% at Bonferroni-corrected $\alpha = .05$, the sample size shall be no less than 6. The observed effect size and power from the current study are presented in Table S2. All power analyses were conducted in G*Power 3.1.9 (Faul et al., 2009).

All participants were right-handed with no reported history of psychiatric or neurological disorders and had a normal or corrected-to-normal vision. All participants had comparable educational status (undergraduate or graduate students). Fifty-six of the participants (28 females) completed an arithmetic assessment prior to the fMRI scan using an arithmetic test similar to the French Kit (Ekstrom et al., 1976) and our previous study (Chang et al., 2018). During the test, participants were instructed to solve a mixture of single- and two-digit addition and subtraction problems as quickly and accurately as possible. The mean accuracy of this test did not differ between females ($M = .730$, $SE = .022$) and males ($M = .755$, $SE = .027$; $t_{(54)} = .703$, $p = .485$, 95% CI = [−.094, .045], $d = .188$). Informed written consent was obtained from each participant. All participants were volunteers and were treated according to the Helsinki Declaration guidelines. All study protocols were approved by the National Chengchi University Review Board.

### 2.2 | Experimental design

All participants were instructed to complete a two-operand arithmetic task composed of addition or subtraction during fMRI scanning, in which they responded the correctness of the presented problem. The problem selection criteria and problem size definition followed the procedure described by Campbell and Xue (2001) and De Smedt et al. (2011), whereby problem stimuli consisted of combinations of single-digit operands from 2 to 9, with the exclusion of tie problems (e.g., 5 + 5) and operand containing 1 or 0. These criteria led to a total of 112 unique problems (56 for each operation). For addition, in the large problem condition, the product of the two operands was larger than 25 (e.g., 8 + 7); in the small problem condition, the product of the two operands was smaller than or equal to 25 (e.g., 5 + 2). For subtraction, stimuli were inverses of addition problems.

The whole set of 112 problems was divided into two functional runs. The sequence of the two runs was counterbalanced between the participants. The presentation sequence of each trial in each run was randomized across participants. As shown in Figure S1, each trial began with a '*' sign as a fixation for 500 ms, followed by the presentation of a problem for 3000 ms in the centre. Next, the corresponding answer to the problem was displayed for 1000 ms. During this period, participants were asked to determine the correctness of the showing answer by pressing one of two keys based on their answer; 50% of the trials were correct (e.g., '6 + 3 = 9'), and the other 50% were incorrect (e.g., '6 + 3 = 8'). The incorrect answers differed by ±1 or ±2 from the correct ones. The screen was then blank for 750 ms. Afterwards, the screen remained blank for a jittered inter-trial interval between 2 and 5 s. Each of the two runs lasted 8.3 min.

## 2.3 | fMRI data acquisition

Neuroimaging data were acquired using a Siemens MAGNETOM Skyra 3 T scanner at National Chengchi University in Taipei City, Taiwan. Head movement was minimized during the scan using cushions placed around the head of each participant. T2* weighted echo-planar sequences were employed with the following parameters: TR = 2 s, TE = 30 ms, flip angle = 90°, 36 ascending axial slices with slice thickness = 4 mm, field of view = 220 × 220 $mm^2$, matrix size = 64 × 64, providing an in-plane spatial resolution of 3.4 mm, and the total number of volumes = 250. In the same scan session, high-resolution T1-weighted MRI sequences were acquired for each participant to aid localization of functional data, with the following parameters: TR = 3500 ms, TE = 3.37 ms, TI = 1100 ms, flip angle = 7°, field of view = 256 × 256 $mm^2$, matrix size = 256 × 256, resulting in resolution of 1 × 1 × 1 $mm^3$, number of excitations = 1 and 192 slices in sagittal plane.

## 2.4 | fMRI data preprocessing

SPM12 (http://www.fil.io.ucl.ac.hk/spm) was used for preprocessing of fMRI data. All functional images were corrected prior to statistical analysis for errors in slice timing, realigned to the first image of each run to correct for head motion, coregistered to each of the individual participant's structural scans, normalized to standard stereotaxic space (based on the Montreal Neurological Institute [MNI] coordinate system) and smoothed with a 6-mm full-width half-maximum (FWHM) Gaussian kernel to decrease spatial noise. Participants with movement more than 3 mm in translational directions and 3° in rotational directions were excluded from further analyses. The average movements of the final participants were .34 ($SE$ = .01), .47 ($SE$ = .03) and .95 ($SE$ = .05) mm in the $x$, $y$ and $z$ directions, and .84° ($SE$ = .05), .34° ($SE$ = .02) and .28° ($SE$ = .01) of roll, pitch and yaw, respectively.

## 2.5 | Individual- and group-level analyses

Statistical analysis was performed on both individual- and group-level data using the general linear model (GLM) implemented in SPM12. Individual subject analyses were first performed by applying GLM that modelled the correctly responded trials as regressors and convolved with a canonical haemodynamic response function to model the expected BOLD signal. Incorrectly responded trials, the epoch participants made responses and the six motion parameters generated in the SPM12 realignment procedure were included as regressors of no interest. Voxel-wise $t$-maps for the contrast between large and small problems from the individual level were generated and entered into subsequent group-level analyses: (i) One-way $t$-contrasts were conducted to identify areas of significant problem size effect (large vs. small problems) of overall participants, (ii) $t$-contrasts of problem size effect for each sex/gender group and (iii) between-group $t$-tests were used to compare activation differences between sexes/genders. All significant results were determined according to a voxel-wise height threshold of $p < .005$ uncorrected and a multiple comparison correction at a spatial-extent threshold of familywise error (FWE) $p < .05$ after grey matter (GM) masking.

## 2.6 | Logistic regression and cross-validation analysis

We conducted a machine learning approach with logistic regression and cross-validation analyses to investigate the robustness of the observed sex/gender differences in the problem size of brain activation level. Because neural correlates associated with arithmetic problems predominantly include fronto-insular-parietal regions (Arsalidou & Taylor, 2011; Chang et al., 2019; De Smedt et al., 2011; Houde et al., 2010), we conducted logistic regression and cross-validations within these regions using regions of interest (ROIs) approach. In order to avoid inflated correlations produced by deriving ROIs from the same dataset (Vul et al., 2009), we defined ROIs using a meta-analysis based on the approach of our previous work (Chang et al., 2018, 2019). Specifically, a

EJN European Journal of Neuroscience FENS


Bayesian meta-analysis of the reverse inference mask available in Neurosynth (Yarkoni et al., 2011) was conducted using the search term 'arithmetic', resulting in a total of 96 studies generated. A false discovery rate (FDR) adjusted *p*-value of .01 was applied to produce the association test map. The coordinates with peak *z*-scores with clusters exceeding 50 voxels on the association test map were identified using the xjView toolbox (www.alivelearn.net/xjview) and selected for further analyses. The resulting brain maps encompassed the left IPS (peak at [−28, −60, 44]), the right IPS [30, −64, 46], the left insula [−22, 22, 2], the left middle frontal gyrus (MFG) [−26, 10, 54], the left inferior frontal gyrus (IFG) [−50, 10, 26] and the left superior frontal gyrus (SFG) [−4, 14, 54]. In subsequent logistic regression analyses, a 10-mm radius sphere (515 voxels with voxel size = 4120 mm$^3$) centred on each of these six identified peak coordinates was created using MarsBaR (http://marsbar.sourceforge.net/) as selected ROIs. Estimated beta values of activation level differences between large and small problems extracted from these ROIs were then entered into the following logistic regression model to classify participants based on their sex/gender.

A multiple logistic regression model was built and verified using the forward stepwise method based on the Akaike information criterion (AIC) (Akaike, 1974) selection and the probability of the Wald statistic. The AIC measures the trade-off between the uncertainty in a model and the number of predictor variables in the model. Lower AIC values imply better prediction of sex/gender labels, as they explain the greatest amount of variation in the response variable with the least amount of predictor variables. The forward stepwise logistic regression starts with a null model, adds the most contributed variables one by one and ends with a model that picks the best variables for an optimal solution. In the current study, the beta value differences between large and small problems generated from the six Neurosynth ROIs were considered as predictive variables. The optimal subset of variables related to sex label discrimination could be determined by utilizing the forward stepwise selection method.

Finally, we evaluated the classification accuracy using the leave-one-out cross-validation (LOOCV) method, which is a commonly used model estimation approach suitable for limited sample sizes (Arlot & Celisse, 2010). LOOCV is an exhaustive holdout splitting approach that involves training the model on all available data points except one, and the excluded data point is used for evaluation. This process is repeated for each data point in the dataset, with each point being left out once and used for evaluation. The utilization of LOOCV enables us to maximize the use of available data, which in turn can produce estimates that are, on average, correct and unbiased. This is particularly advantageous in small sample sizes where the use of all available data is crucial for achieving reliable and accurate model performance. In our study, we applied the LOOCV method to all 71 subjects, with all but one participant's data used to train the model (training data = 70) and the remaining participant's data used for testing (testing data = 1). We repeated this procedure for each data point in the dataset to obtain the overall classification accuracy. Subsequently, we constructed a receiver-operating characteristic (ROC) curve using the probability thresholds with corresponding data points (sensitivity, 1 − specificity) and calculated the area under the curve (AUC). The modelling and statistical analyses were implemented using R packages 'caret' and 'MLeval'.

### 2.7 | Voxel-based morphometry (VBM) analysis

In order to investigate whether the sex/gender effects in brain activation are associated with anatomical differences, brain anatomy was examined using VBM analysis. The CAT12 toolbox (CAT12; http://dbm.neuro.uni-jena.de/cat12) implemented in the SPM12 software was used to process the T1-weighted images. Structural T1-weighted images of each participant were first converted into the Neuroimaging Informatics Technology Initiative (NIFTI) format through SPM12. The images were then preprocessed with the standard default procedure recommended in the CAT12 manual. The preprocessing steps included skull stripping, segmentation into GM, white matter (WM) and cerebrospinal fluid (CSF), followed by spatial normalization to the DARTEL template in the MNI space with 1.5-mm cubic resolution. The quality of the images was assessed with the built-in image quality rating and manually visual check. Finally, images were smoothed using a 6-mm FWHM isotropic Gaussian kernel. In addition, the total intracranial volume (TIV), which is the sum of GM, WM and CSF volumes in the native space, was also estimated.

## 3 | RESULTS

### 3.1 | Behavioural results

The mean accuracy and reaction time for each problem condition for each participant were computed and analysed using repeated-measures analysis of variance (ANOVA) with problem size (small, large) as within-subject factor and sex (female, male) as a between-subjects factor. For the accuracy (Figure 1a), as predicted, there was a significant main effect of problem size, showing that participants performed more accurately on small-size problems than large-size problems (96.8% vs. 95.6%, $F[1,69] = 9.334$, MSE $= .005$, $p = .003$, $\eta^2 = .016$). No differences between males and females were observed when performing the arithmetic task (95.4% vs. 96.9%, $F[1,69] = 1.931$, MSE $= .008$, $p = .169$, $\eta^2 = .024$), nor did the interaction effect between sex and problem size was significant ($F[1,69] = .618$, MSE $< .001$, $p = .434$, $\eta^2 = .001$), indicating that females and males perform equally well on this current simple single-digit calculation task.

Regarding the reaction time analysis (Figure 1b), likewise, there were main effects of problem size, showing that participants responded faster to small-size problems than to large-size problems (640 vs. 650 ms, $F[1,69] = 9.969$, MSE $= 3901$, $p = .002$, $\eta^2 = .002$). No difference was observed between males and females (637 vs. 652 ms, $F[1,69] = .227$, MSE $= 7982$, $p = .635$, $\eta^2 = .003$), nor did the interaction effect between sex and problem size reach significance ($F[1,69] = .244$, MSE $= 96$, $p = .623$, $\eta^2 = .001$).

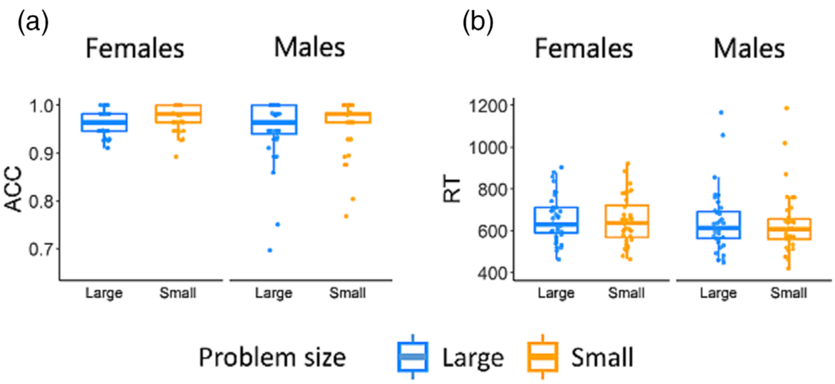


**FIGURE 1** Accuracy (ACC) and reaction times (RT) of the arithmetic task in females and males. (a) Participants performed more accurately to small-size problems than large-size problems. No significant difference between sexes was found, and sex and problem size did not interact significantly. (b) Participants responded faster to small-size problems than large-size problems. No significant difference between sexes was found, nor was the interaction effect between sex and problem size significant.

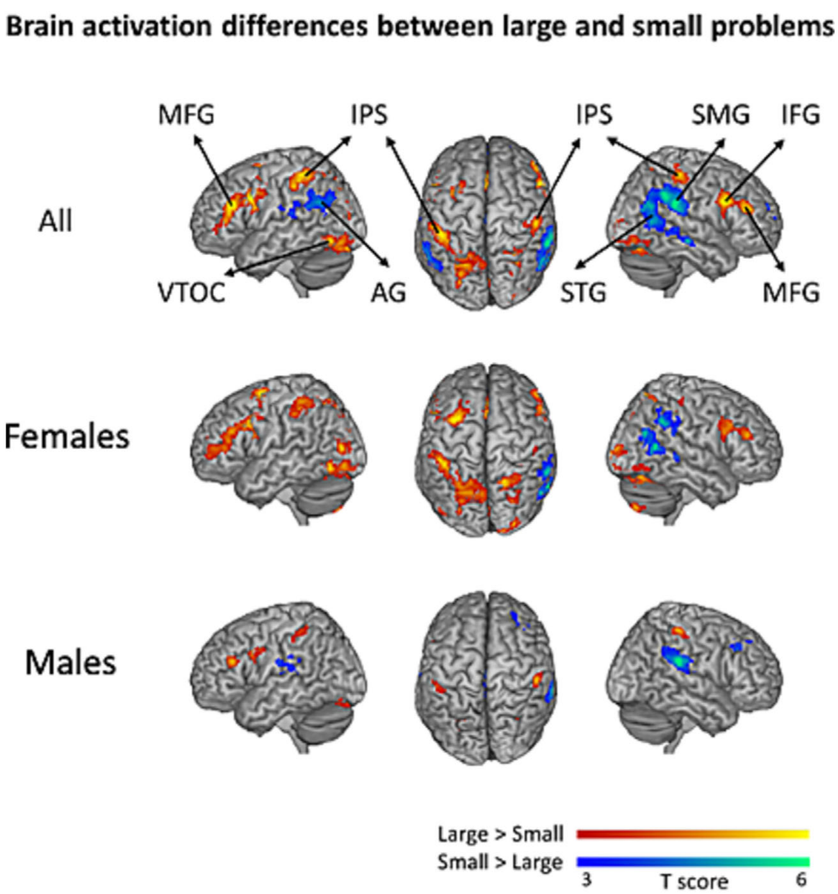


**FIGURE 2** Brain regions that showed different activation levels between large- and small-size problem-solving across overall participants (upper panel), in females (middle panel) and in males (lower panel). Activations in fronto-parietal regions including middle frontal gyrus (MFG), inferior frontal gyrus (IFG) and intraparietal sulcus (IPS) were greater in large problems than small problems. On the other hand, activations in angular gyrus (AG), supramarginal gyrus (SMG) and superior temporal gyrus (STG) were higher in small problems than large problems. VTOC, ventrotemporal occipital cortex.

## 3.2 | Brain imaging results

### 3.2.1 | Brain responses that showed differences between large and small problems

We first identified brain regions showing response differences associated with arithmetic problem size by contrasting the neural correlates of large and small problems in the pooled group of males and females. The resulting problem size effects on brain activation are presented in Figure 2a. Relative to small problems, large problems activated stronger in the bilateral MFG extending to adjacent IFG and medial frontal gyrus in the prefrontal cortex (PFC), as well as bilateral IPS in the PPC. Additional clusters were also found in the ventrotemporal occipital cortex (VTOC), including bilateral lingual gyrus (LG), fusiform gyrus (FG) and calcarine. In contrast to large problems, small problems were activated more in the bilateral supramarginal gyrus (SMG) and the left AG in the PPC, medial PFC, posterior cingulate cortex and bilateral superior temporal gyrus (STG) (Table 1 shows the detailed results of the peak coordinates in each cluster).

**TABLE 1** Main effect of problem size on brain activation.

| Region | Corrected $p_{FWE}$ | # of voxels | Peak $T$-score | Peak MNI coordinates | | |
|---|---|---|---|---|---|---|
| | | | | $x$ | $y$ | $z$ |
| **Large > small** | | | | | | |
| R lingual gyrus | <.001 | 11,144 | 10.07 | 18 | −82 | −8 |
| R calcarine | | | 9.48 | 18 | −86 | 0 |
| L lingual gyrus | | | 9.30 | −16 | −88 | −10 |
| L fusiform gyrus | | | 7.12 | −24 | −78 | −12 |
| R intraparietal sulcus | | | 6.59 | 28 | −56 | 44 |
| L intraparietal sulcus | <.001 | 3416 | 8.17 | −28 | −64 | 40 |
| L inferior parietal lobule | | | 7.52 | −32 | −52 | 42 |
| L middle occipital gyrus | | | 5.61 | −30 | −74 | 22 |
| L inferior frontal gyrus | <.001 | 2819 | 7.62 | −46 | 10 | 30 |
| L middle frontal gyrus | | | 6.80 | −48 | 30 | 24 |
| L medial frontal gyrus | <.001 | 1174 | 8.16 | −6 | 16 | 48 |
| R medial frontal gyrus | | | 6.14 | 6 | 20 | 44 |
| R cingulate gyrus | | | 4.45 | 12 | 22 | 36 |
| R inferior frontal gyrus | <.001 | 961 | 6.09 | 56 | 14 | 28 |
| R middle frontal gyrus | | | 5.49 | 46 | 36 | 20 |
| **Small > large** | | | | | | |
| R supramarginal gyrus | <.001 | 1959 | 5.92 | 62 | −36 | 26 |
| R middle temporal gyrus | | | 5.55 | 62 | −22 | −12 |
| R superior temporal gyrus | | | 5.03 | 58 | −60 | 22 |
| R medial frontal gyrus | <.001 | 1143 | 5.24 | 6 | 56 | −2 |
| R superior frontal gyrus | | | 4.07 | 20 | 56 | 24 |
| L superior frontal gyrus | | | 4.05 | −10 | 58 | 2 |
| L medial frontal gyrus | | | 4.03 | −4 | 62 | 8 |
| L superior temporal gyrus | <.001 | 843 | 4.66 | −52 | −28 | 14 |
| L insula | | | 4.62 | −46 | −18 | 10 |
| L supramarginal gyrus | | | 3.61 | −58 | −36 | 26 |
| R posterior cingulate gyrus | .012 | 406 | 5.03 | 8 | −32 | 44 |
| L cingulate gyrus | | | 3.93 | −2 | −22 | 42 |
| L angular gyrus | .016 | 385 | 4.72 | −58 | −58 | 28 |

Abbreviations: FWE, familywise error; L, left; MNI, Montreal Neurological Institute; R, right.

### 3.2.2 | Females exhibited larger problem size effects than males

We then examined problem size effect of each sex/gender. As shown in Figure 2 and Table S1, females exhibited positive problem size effect (large–small problems) in the bilateral MFG, IPS, VTOC and the dACC. Small problems activated stronger right STG and SMG than large problems. In contrast, for males, large problems showed stronger activations than small problems only in the left MFG, the right IPS and the bilateral LG to a lesser extent. Negative problem size effects were observed in the bilateral SMG.

To investigate whether males and females show differences when processing large and small problems, we examined brain areas that showed problem size by sex interaction. This analysis revealed significant differences in the left MFG, IPS and the right dACC (Figure 3; Table 2 reveals detailed results of the peak coordinates in each cluster). Further analysis of the averaged beta weights of each significant cluster revealed that the interaction effect was driven by the brain response cost (complex–simple) being more prominent in females (Figure 3). Specifically, females showed stronger activations for complex than simple problems in the left MFG ($t_{(35)} = 4.006$, $p < .001$, 95% CI = [.120, .366], $d = .677$), the left IPS ($t_{(35)} = 6.600$, $p < .001$, 95% CI = [.371, .701], $d = 1.116$) and the right dACC ($t_{(35)} = 4.332$, $p < .001$, 95% CI = [.098, .272], $d = .732$). Males, on the contrary, showed a minimal or null effect of problem size in these brain regions (left MFG: $t_{(34)} = -2.033$, $p = .05$, 95% CI = [−.228, −.001], $d = .349$; left IPS: $t_{(34)} = 1.699$, $p = .099$, 95% CI = [−.022, .246], $d = .291$; and right dACC: $t_{(34)} = -1.713$, $p = .096$, 95% CI = [−.124, .011], $d = .294$).

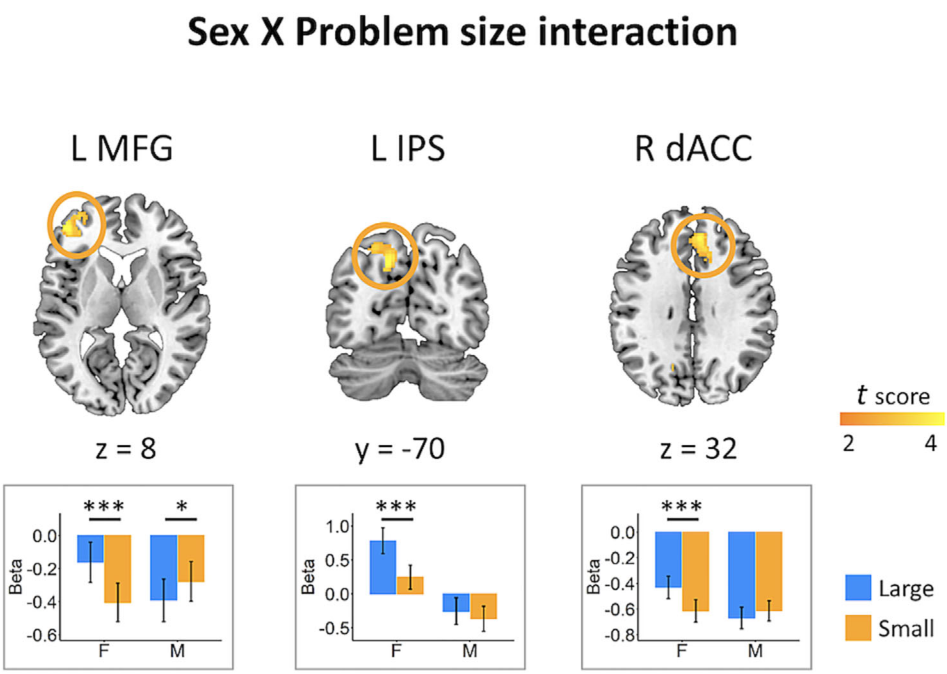


**FIGURE 3** Statistical maps illustrating regions activated for sex and problem size interaction effects. A problem size effect was evident in females (F), with greater activation in large problems than small problems, whereas a problem size effect was negligible in males (M). Error bars represent standard errors. *$p < .05$, ***$p < .001$. L IPS, left intraparietal sulcus; L MFG, left middle frontal gyrus; R dACC, right dorsal anterior cingulate cortex.

**TABLE 2** Sex differences in brain activation in the mental arithmetic task.

| Region | Corrected $p_{FWE}$ | # of voxels | Peak $T$-score | Peak MNI coordinates $x$ | $y$ | $z$ |
|---|---|---|---|---|---|---|
| **Problem size effects** | | | | | | |
| Females > males | | | | | | |
| R dorsal anterior cingulate gyrus | .006 | 447 | 4.06 | 6 | 40 | 32 |
| L intraparietal sulcus | .006 | 444 | 4.33 | −14 | −70 | 40 |
| L middle frontal gyrus | .039 | 314 | 4.29 | −44 | 48 | 8 |
| Males > females | | | | | | |
| No significant clusters | | | | | | |

Abbreviations: FWE, familywise error; L, left; MNI, Montreal Neurological Institute; R, right.

To validate that the sex/gender differences in the functional brain activations were not resulted from fundamental neuroanatomical differences, we conducted a mix-designed two-way analysis of covariance (ANCOVA) with sex/gender group as a between-participant factor and problem size as a within-participant factor while covarying out TIVs and regional GM volumes in the dACC, left IPS and left MFG, the regions identified in the functional activation analysis that demonstrated the sex/gender differences in the problem size effect. The sex/gender by problem size interaction effect remained significant in all the three regions (dACC: $p = .001$, $\eta^2 = .005$; left IPS: $p = .017$, $\eta^2 = .002$; and left MFG: $p = .008$, $\eta^2 = .003$). Overall, our results suggested that sex-related differences in fronto-parietal activation are not related to differences in underlying neuroanatomy.

### 3.2.3 | The robustness of sex/gender effect in the brain responses

We then examined the robustness of the sex/gender effect in the brain activity levels using a logistic regression function. To avoid inflated correlations produced by deriving ROIs from the same dataset (Vul et al., 2009), the averaged beta values of large and small problem sizes for each participant were extracted from the six ROIs defined by meta-analysis, as introduced in Section 2.6. The selected ROIs were highly overlapped with the activation level maps generated from the one-sample *t*-test on the contrast of large minus small problems on the data from pooled males and females together (Figure S2). We conducted a binary logistic regression classifier to categorize males and females using the estimated activation level difference between large and small problems extracted from the aforementioned unbiased ROIs. The results showed that among the six ROIs, the logistic coefficients were significant in the left insula ($\beta = 6.089$, $p = .011$, odds ratio = 441.099, 95% CI = [1.708, 11.167]), the left MFG ($\beta = 2.677$, $p = .006$, odds ratio = 14.541, 95% CI = [.918, 4.761]) and the left IPS ($\beta = 2.094$, $p = .005$, odds ratio = 8.116, 95% CI = [.723, 3.688]) (Figure 4). These results indicated that female participants had a higher probability of exhibiting greater degrees of activation level difference between large and small problems within the left insula, MFG and IPS (Figure 4).

In an attempt to build a logistic model that best describes the sex differences in problem size, the stepwise forward method was then performed with the variables based on the estimated activation level difference between large and small problems extracted from the selected ROIs. Of the six aforementioned ROIs, the optimal subset of variables related to sex label discrimination using the forward stepwise selection method resulted in five ROIs: right IPS, left IPS, left IFG, left MFG and left insula. The AIC of the final model consisting of the five ROIs is 77.467, with the logistic coefficients for the right IPS ($\beta = -3.682$, Wald's $\chi^2 = 10.4$), the left IPS ($\beta = 6.228$, Wald's $\chi^2 = 8.7$), the left MFG ($\beta = 3.482$, Wald's $\chi^2 = 4.9$) and the left IFG ($\beta = -2.537$, Wald's $\chi^2 = 4.2$) showing statistical significance at the .05 level (Figure 5a). All estimation parameters are reported in Table 3. The overall model classification accuracy was then evaluated using an LOOCV procedure. Figure 5b illustrates the machine learning method's analysing strategy and the model evaluation procedures. The LOOCV method was employed to evaluate the classification

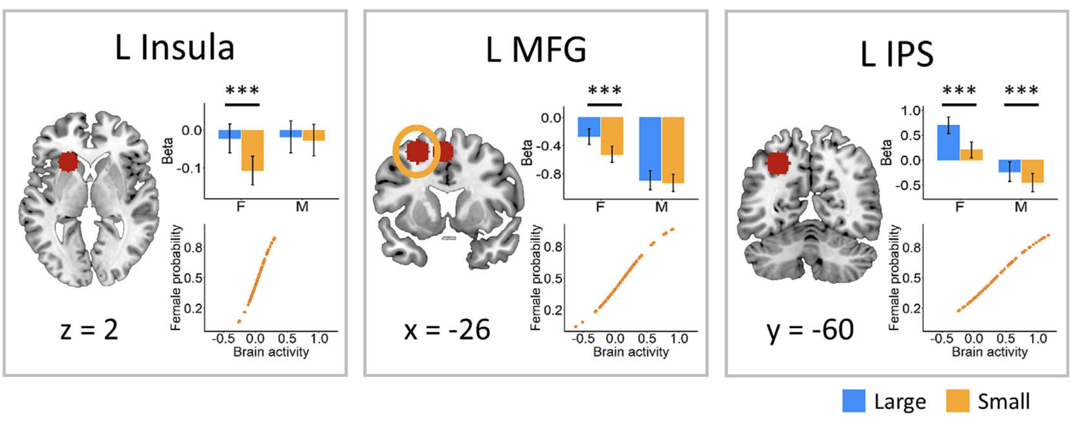


**FIGURE 4** Logistic regression models applied in each region of interest to classify participants' sex. Within each frame, the top-right bar plots revealed that problem size effects were only observed in females (F) over left (L) insula and left middle frontal gyrus (L MFG) while the effects were found in both females and males (M) over left intraparietal sulcus (L IPS). The bottom-right sigmoid function plots of each frame indicated that participants' sex (*y* axis) could be classifiable based on brain activation level differences between large and small problems (*x* axis).

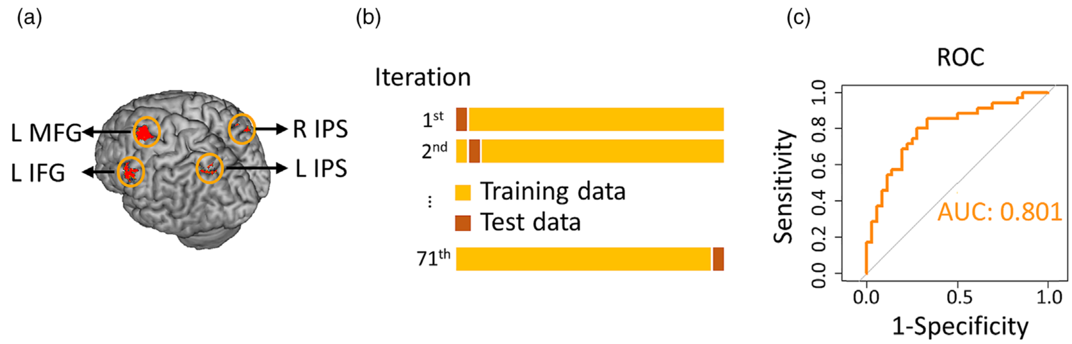


**FIGURE 5** Multiple logistic regression and leave-one-out cross-validation (LOOCV) applied on the designated regions of interest (ROIs). (a) Final multiple logistic model explaining sex differences in problem complexity by looking at the estimated activation levels between large and small problems extracted from the selected ROIs. In the final five-variant model, predictors of right intraparietal sulcus (R IPS), left intraparietal sulcus (L IPS), left middle frontal gyrus (L MFG) and left inferior frontal gyrus (L IFG) contributed significantly to female/male differentiation during the arithmetic task. (b) Multiple logistic regression model evaluation. In the original dataset, the prediction variables were beta value differences in five ROIs: right IPS, left IPS, left IFG, left MFG and left insula. LOOCV method was then applied to all 71 subjects by training the model on all but one participant's data and testing on the remaining participant's data. This procedure was repeated for each participant in the dataset to obtain the overall classification accuracy. (c) Receiver-operating characteristic (ROC) curve and overall model performances for prediction of sex labels on the test set. AUC indicates area under the curve.

**TABLE 3** Multiple logistic regression model results for predicting biological sex during simple and complex arithmetic task.

| Term | $\beta$ | *SE* | Wald's $\chi^2$ (df = 1) | *p*-value | OR | 95% CI |
|---|---|---|---|---|---|---|
| Intercept | −.69 | .45 | 2.3 | .128 | .50 | [−1.61, .18] |
| R IPS | **−3.68** | **1.14** | **10.4** | **.012** | **.03** | **[−6.20, −1.67]** |
| L IPS | **6.23** | **2.11** | **8.7** | **.003** | **506.68** | **[2.50, 10.88]** |
| L MFG | **3.48** | **1.58** | **4.9** | **.027** | **32.53** | **[.68, 6.94]** |
| L IFG | **−2.54** | **1.24** | **4.2** | **.040** | **.08** | **[−5.13, .21]** |
| L insula | 4.88 | 3.52 | 1.9 | .165 | 132.22 | [−1.75, 12.32] |

*Note*: Significant variables are in bold.
Abbreviations: CI, confidence interval; IFG, inferior frontal gyrus; IPS, intraparietal sulcus; L, left; MFG, middle frontal gyrus; OR, odds ratio; R, right.

accuracy in our study, which resulted in an overall accuracy of 73% ($\kappa = .46$). The model's sensitivity and specificity were found to be 66% and 64%, respectively, while the AUC was calculated to be 80%. The ROC curve generated from the test set and the corresponding AUC are presented in Figure 5d.

## 4 | DISCUSSION

In this study, we investigate sex/gender effect in brain response profiles underlying arithmetic problem-solving. We aim to probe distinct math problem-solving strategies by directly manipulating problem size. As predicted, we did not observe any behavioural performance differences between females and males. However, the sex/gender-related effects in brain activation level varied depending on problem size in the left MFG, IPS and the right dACC. This interaction is manifested by females showing greater problem size effect of brain activation level than males. Furthermore, machine learning algorithm confirmed the robustness of the distinct brain response profile in each sex/gender. Importantly, these sex/gender effects in the functional activations were independent of differences in the neuroanatomical structure. Below, we discuss the implications of our findings.

### 4.1 | Problem size effect in fronto-parietal circuits is more salient in females

The key finding of the current study is that the brain activates to a different level towards problem size between females and males in the left MFG, the left IPS and dACC. Specifically, within these regions, females exhibited stronger activations for large problems than small problems, whereas males displayed more negligible problem size effects within these regions. The IPS has been identified as playing a crucial role in quantity representation (Arsalidou & Taylor, 2011; Dehaene et al., 2003) and has been suggested to reflect the use of quantity-based procedure strategies in mathematics (Stanescu-Cosson et al., 2000). The MFG has been associated with complex and effortful tasks involving quantity manipulation (Chang et al., 2015; Menon et al., 2000; Wu et al., 2009). The IPS and MFG comprise the major nodes of the central executive network (CEN). On the other hand, the dACC coupling with the AI forms the major components of the salience network (SN) that serves as a major causal hub in problem-solving and functions as integrating and directing salient stimuli and initiating control signals (Seeley et al., 2007; Menon, 2015b). Functional imaging studies have identified that the CEN and SN jointly play mandatory and unique role in arithmetic processing. The problem size effect within the CEN and SN reflects the demand for quantity processing and attentional control of large problem relative to small problems (Chang et al., 2016; De Smedt et al., 2011; Polspoel et al., 2019; Stanescu-Cosson et al., 2000; Tiberghien et al., 2019). Self-report trial-based strategy assessment revealed that problems solved by procedural calculation recruited stronger activation levels within these fronto-parietal regions (Grabner et al., 2009). Our results showed that problem size effects in these regions are more prominent in females. We thus suspect that these findings might implicate the different problem-solving strategies used by each sex during calculation (Bailey et al., 2012; Gallagher et al., 2000; Quinn & Spencer, 2001; Zhu, 2007).

Several studies have reported that females use retrieval strategies less frequently than males to solve arithmetic problems (Bailey et al., 2012; Carr & Davis, 2001). In a longitudinal study that followed students from second to fourth grades, Carr and Alexeev (2011) found that even when controlling for accuracy and fluency, females showed a slower rate than males in reducing their use of manipulative strategies use, such as counting on fingers. These results suggested that females are more likely than males to adopt procedural calculations, even when their performance is comparable. Bailey et al. (2012) further interpreted the sex/gender specificity in problem-solving strategy as a product of personalities. Specifically, females are generally less risk-taking than males. Instead of bearing the chance to make a mistake in rapidly retrieving answer, females are more likely to choose manipulative strategies that are slower but usually more precise. Consistent with this view, our results showed that females engage in a greater level of fronto-insular-parietal circuits even while solving simple arithmetic problems, indicating a biological basis of the characteristic of each sex/gender in the problem-solving strategies.

Another possibility we suspect that can attribute sex/gender profile is the attitude towards mathematics (Di Martino & Zan, 2011). Negative emotional reactions—math anxiety—can be elicited when dealing with math-related situations (Ashcraft & Ridley, 2005). It can be triggered even when solving simple arithmetic problems, especially during timed conditions (Caviola et al., 2017) and when tasks increase in complexity (Ashcraft & Krause, 2007). Negative correlations between math anxiety and math achievement have also been reported in a wide range of students (Hembree, 1990; Ng et al., 2022). Even though females generally perform equivalently well with males in mathematical achievement, females are reported to show higher self-report math anxiety (Devine et al., 2012; Else-Quest et al., 2010; Ferguson et al., 2015; Hembree, 1990; Lau et al., 2022; Maloney et al., 2012) that can likely be attributed to social–cultural or emotional factors (Beilock et al., 2007; Bieg et al., 2015). The sex-specific math anxiety profile remained even when general anxiety was controlled (Devine et al., 2012; Goetz et al., 2013). Math anxiety is also often comorbid with limited working memory (Ashcraft & Kirk, 2001; Ashcraft & Krause, 2007; Ramirez et al., 2013). One empirical example provided by Ashcraft and Kirk (2001) reported that highly math-anxious college students were less accurate in performing addition problems with carry operation only when implementing a secondary task that required a high working memory load. Consequently, the impact of math anxiety on learning can possibly be due to the disturbance of working memory strategies while performing mathematical tasks. Consistently, highly math-anxious participants showed more enhanced engagement of the fronto-parietal cortices, including the IPS and MFG (Supekar et al., 2015). Therefore, it is possibly the exceedingly high math anxiety that up-regulates fronto-parietal engagement in females during the timed calculation task. This interpretation, however, is admittedly speculative. Further direct assessments on the relationship between math anxiety levels and brain response profiles of each sex/gender are still needed.

### 4.2 | Males and females showed a comparable level of behaviour performance

It is noteworthy to highlight that our study is in line with the majority of the previous research that behaviour assessment (cf. Chang et al., 2022; Hyde, 2014) and task-dependent neuroimaging studies (Keller & Menon, 2009; Pletzer, 2016) have reported no differences between females and males in either accuracy or reaction times.

These results demonstrated that while males and females can utilize distinct cognitive strategies and exhibit different brain response profiles, they are capable of achieving a comparable level of behavioural performance.

Note that our task was designed in order to ensure participants had enough time to obtain each problem solution and, in the meantime, to avoid motor responses contaminating the brain activation level towards responding to numerical problems, as the supplementary motor area (SMA) is consistently activated during arithmetic problem-solving (Menon et al., 2014). Therefore, the presentation duration of the stimuli was long as 3 s, and participants were instructed to wait to make a verification response until the problem offset. Consequently, one may argue that our behavioural results needed to be validated for indexing the actual response time. We argue that given so our results still inherent conventional problem size effect in the measurements of accuracy, response latencies and brain response profiles (De Smedt et al., 2011; Stanescu-Cosson et al., 2000), suggesting that the task design has sufficient loading to differentiate the processes between distinct conditions even when the performance reaches high as ceilings. The current task design should thus be sensitive enough to provide behaviour-independent evidence of examination of brain functional organization.

## 4.3 | Discrepancies with previous studies

Note that there are discrepancies between our results and other fMRI studies, for example, Keller and Menon (2009). In that study, Keller and Menon (2009) compared sex/gender differences in the brain activations while adult participants calculated three-operand addition and subtraction problems. The results showed that males engaged in a greater level of IPS, AG, lingual and parahippocampal gyri, whereas no regions showed greater functional activation in females than males. The discrepancies can likely be resulted from the varied task design. In the study conducted by Keller and Menon (2009), a three-operand mixed operation task was implemented and compared with a number identification task, resulting in greater dorsal and ventral-stream activations in males. The multistep and multi-operation calculation task can consume more working memory load and require the engagement of multiple strategies, making it difficult to disentangle the sex effect resulting from problem size or problem operation.

Our results are also inconsistent with Pletzer (2016). In that study, Pletzer examined the brain response patterns of young adults as they performed subtraction and multiplication tasks. Participants showed stronger IPS activations for subtraction and greater AG activations for multiplication tasks. Interestingly, this operation effect was only observed in males but not in females, suggesting that females showed less differentiation between specific numerical problems. The task problem included two-digit subtraction with single-digit multiplication. Given that decomposition and transformation strategies are frequently reported in solving multi-digit subtraction with borrowing (LeFevre et al., 2006), it may be necessary to use a combination of strategies and to engage higher order of attention. Moreover, it is worth noting that the problem size is much larger in their subtraction task than ours, and the operation effects may be confounded with the problem size effects. Given that both the operations and problem sizes are distinct from Keller and Menon (2009) and Pletzer (2016) as well as our studies, it is challenging to directly generalize the results. Importantly, the discrepancies between these studies have highlighted the importance of continuing to investigate context-dependent sex/gender specificity.

## 4.4 | Implications for using neuroimaging studies to understand sex/gender difference

The current findings shed light on the limitations of using behavioural performance as a sole measure to characterize brain configuration. This is illustrated by females and males engaging distinct brain response profiles despite having similar behavioural performances. Insomuch of this assumption, it can be doubted that the previous observations of null results on sex/gender differences are likely underestimated. Behavioural assessments may not always secure such a level of cognitive processes. As a result, neuroimaging facilities, in contrast, have a strong potential to provide useful knowledge that is unseen in behavioural outcomes alone. Therefore, it is of crucial importance to provide unique perspectives using state-of-the-art neuroimaging techniques to characterize the brain mechanisms of each sex/gender.

In view of our findings, it should be noted that males and females engage different response profiles of neural resources to maintain parallel performance. Such patterns may be regulated by problem-solving strategies and various psychosocial factors (Taddei et al., 2022), which may pose greater challenges to females than males. It is, therefore, crucial to consider these factors when designing interventions aimed at supporting female students in math-related fields. We propose that efforts in instructional practices should prioritize assisting these psychological aspects by promoting positive math attitudes and encouraging the adoption of more efficient strategies.

### 4.5 | Limitations

While our study proposes a robust sex/gender difference in the neural correlates of arithmetic problem-solving, we acknowledge that there are certain limitations that must be considered. First, we did not measure the individual-level strategy assessment and math anxiety levels. Further studies are warranted to validate the relationship between the brain functional activation profile and the strategy used by each individual as well as the math anxiety level. Second, our task does not measure the immediate response made by participants. Given that gender effect in math performance can be enlarged under timed conditions (Tsui & Mazzocco, 2006), further investigations with control and manipulation on timing are still expected.

### 4.6 | Conclusions

Over the past decades, cognitive and neural imaging studies have gained considerable insight into uncovering sex/gender differences in the mechanisms of learning. This work has led to advances in exploring the biological underpinnings of individual differences. However, direct manipulation of problem types during mathematical problem-solving had yet to be systematically investigated. Our study emphasizes the importance of a linear task design in probing brain response profiles. Our findings revealed that, for the first time, problem size effects were markedly more prominent for females than males. The robustness of the sex/gender effect was confirmed using a machine learning approach to characterize each individual's biological sex label. In line with the majority of the literature, our results suggested that females and males accomplish equivalently successful mathematical achievements by taking distinct neural mechanisms. Further questions are raised, such as the effect of problem type, strategy selection and developmental progression. Future studies investigating potential neural mechanisms of when and how certain factors influence children's developing mathematical knowledge would improve the quality of school instruction and methods of teaching mathematics.

## AUTHOR CONTRIBUTIONS

**Nai-Feng Chen:** Data curation; formal analysis; investigation; methodology; visualization; writing — original draft. **Ting-Ting Chang:** Conceptualization; funding acquisition; investigation; methodology; project administration; supervision; validation; writing — original draft; writing — review and editing.

## ACKNOWLEDGEMENTS

We thank Dr. Tzu-Ting Lin for her valuable feedback and suggestions. This research was supported by grants from the National Science Council (grant numbers 110-2629-H-004-001 and 111-2811-H-004-004).

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ejn.16100.

## DATA AVAILABILITY STATEMENT

Data are available on request because of privacy/ethical restrictions.

## ORCID

*Ting-Ting Chang* https://orcid.org/0000-0002-5943-1396

## REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews. Neuroscience*, *9*, 278–291. https://doi.org/10.1038/nrn2334

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79. https://doi.org/10.1214/09-SS054

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

Ashcraft, M., & Ridley, K. S. (2005). Math anxiety and its cognitive consequences: A tutorial review. In *Handbook of mathematical cognition* (pp. 315–327). Psychology Press.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology. General*, *130*(2), 224–237. https://doi.org/10.1037/0096-3445.130.2.224

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*, *14*, 243–248. https://doi.org/10.3758/BF03194059

Bailey, D. H., Littlefield, A., & Geary, D. C. (2012). The codevelopment of skill at and preference for use of retrieval-based processes for solving addition problems: Individual and sex differences from first to sixth grades. *Journal of Experimental Child Psychology*, *113*, 78–92. https://doi.org/10.1016/j.jecp.2012.04.014

Barrouillet, P., Mignon, M., & Thevenot, C. (2008). Strategies in subtraction problem solving in children. *Journal of Experimental Child Psychology*, *99*, 233–251. https://doi.org/10.1016/j.jecp.2007.12.001

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*, 256–276. https://doi.org/10.1037/0096-3445.136.2.256

Bieg, M., Goetz, T., Wolter, I., & Hall, N. (2015). Gender stereotype endorsement differentially predicts girls' and boys' trait-state discrepancy in math anxiety. *Frontiers in Psychology*, *6*, 1404. https://doi.org/10.3389/fpsyg.2015.01404

Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology. General*, *130*, 299–315. https://doi.org/10.1037/0096-3445.130.2.299

Carr, M., & Alexeev, N. (2011). Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology*, *103*, 617–631. https://doi.org/10.1037/a0023864

Carr, M., & Davis, H. (2001). Gender differences in arithmetic strategy use: A function of skill and preference. *Contemporary Educational Psychology*, *26*, 330–347. https://doi.org/10.1006/ceps.2000.1059

Carr, M., & Jessup, D. (1997). Gender differences in first-grade mathematics strategy use: Social and metacognitive influences. *Journal of Educational Psychology*, *89*, 318–328. https://doi.org/10.1037/0022-0663.89.2.318

Caviola, S., Carey, E., Mammarella, I. C., & Szucs, D. (2017). Stress, time pressure, strategy selection and math anxiety in mathematics: A review of the literature. *Frontiers in Psychology*, *8*, 1488. https://doi.org/10.3389/fpsyg.2017.01488

Chang, T. T., Chen, N. F., & Fan, Y. T. (2022). Uncovering sex/-gender differences of arithmetic in the human brain: Insights from fMRI studies. *Brain and Behavior: a Cognitive Neuroscience Perspective*, *12*, e2775. https://doi.org/10.1002/brb3.2775

Chang, T. T., Lee, P. H., & Metcalfe, A. W. S. (2018). Intrinsic insula network engagement underlying children's reading and arithmetic skills. *NeuroImage*, *167*, 162–177. https://doi.org/10.1016/j.neuroimage.2017.11.027

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

Chang, T.-T., Metcalfe, A. W., Padmanabhan, A., Chen, T., & Menon, V. (2016). Heterogeneous and nonlinear development of human posterior parietal cortex function. *NeuroImage*, *126*, 184–195. https://doi.org/10.1016/j.neuroimage.2015.11.053

Chang, T. T., Rosenberg-Lee, M., Metcalfe, A. W., Chen, T., & Menon, V. (2015). Development of common neural representations for distinct numerical problems. *Neuropsychologia*, *75*, 481–495. https://doi.org/10.1016/j.neuropsychologia.2015.07.005

Cho, S., Metcalfe, A. W., Young, C. B., Ryali, S., Geary, D. C., & Menon, V. (2012). Hippocampal-prefrontal engagement and dynamic causal interactions in the maturation of children's fact retrieval. *Journal of Cognitive Neuroscience*, *24*, 1849–1866. https://doi.org/10.1162/jocn_a_00246

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, *84*, 132–147. https://doi.org/10.1016/j.pneurobio.2007.11.001

De Smedt, B., Holloway, I. D., & Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. *NeuroImage*, *57*, 771–781. https://doi.org/10.1016/j.neuroimage.2010.12.037

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*, 487–506. https://doi.org/10.1080/02643290244000239

Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions*, *8*, 33. https://doi.org/10.1186/1744-9081-8-33

Di Martino, P., & Zan, R. (2011). Attitude towards mathematics: A bridge between beliefs and emotions. *ZDM - International Journal on Mathematics Education*, *43*, 471–482. https://doi.org/10.1007/s11858-011-0309-6

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Education Testing Service.

Eliot, L., Ahmed, A., Khan, H., & Patel, J. (2021). Dump the "dimorphism": Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neuroscience and Biobehavioral Reviews*, *125*, 667–697. https://doi.org/10.1016/j.neubiorev.2021.02.026

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127. https://doi.org/10.1037/a0018053

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ferguson, A. M., Maloney, E. A., Fugelsang, J., & Risko, E. F. (2015). On the relation between math and spatial ability: The case of math anxiety. *Learning and Individual Differences*, *39*, 1–12. https://doi.org/10.1016/j.lindif.2015.02.007

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, *75*, 165–190. https://doi.org/10.1006/jecp.1999.2532

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science*, *24*, 2079–2087. https://doi.org/10.1177/0956797613486989

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

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, *21*, 33–46. https://doi.org/10.2307/749455

Hirnstein, M., Bayer, U., & Hausmann, M. (2009). Sex-specific response strategies in mental rotation. *Learning and Individual Differences*, *19*, 225–228. https://doi.org/10.1016/j.lindif.2008.11.006

Hong, E., & Aqui, Y. (2004). Cognitive and motivational characteristics of adolescents gifted in mathematics: Comparisons among students with different types of giftedness. *The Gifted Child Quarterly*, *48*, 191–201. https://doi.org/10.1177/001698620404800304

Houde, O., Rossi, S., Lubin, A., & Joliot, M. (2010). Mapping numerical processing, reading, and executive functions in the developing brain: An fMRI meta-analysis of 52 studies including 842 children. *Developmental Science*, *13*, 876–885. https://doi.org/10.1111/j.1467-7687.2009.00938.x

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, *65*, 373–398. https://doi.org/10.1146/annurev-psych-010213-115057

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139–155. https://doi.org/10.1037/0033-2909.107.2.139

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*, 494–495. https://doi.org/10.1126/science.1160364

Jordan-Young, R., & Rumiati, R. I. (2012). Hardwired for sexism? Approaches to sex/gender in neuroscience. *Neuroethics*, *5*, 305–315. https://doi.org/10.1007/s12152-011-9134-4

Keller, K., & Menon, V. (2009). Gender differences in the functional and structural neuroanatomy of mathematical cognition. *NeuroImage*, *47*, 342–352. https://doi.org/10.1016/j.neuroimage.2009.04.042

Lau, N. T. T., Hawes, Z., Tremblay, P., & Ansari, D. (2022). Disentangling the individual and contextual effects of math anxiety: A global perspective. *Proceedings of the National Academy of Sciences*, *119*, e2115855119. https://doi.org/10.1073/pnas.2115855119

LeFevre, J.-A., Smith-Chant, B. L., Fast, L., Skwarchuk, S.-L., Sargla, E., Arnup, J. S., Penner-Wilger, M., Bisanz, J., & Kamawar, D. (2006). What counts as knowing? The development of conceptual and procedural knowledge of counting from kindergarten through Grade 2. *Journal of Experimental Child Psychology*, *93*, 285–303. https://doi.org/10.1016/j.jecp.2005.11.002

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, *136*, 1123–1135. https://doi.org/10.1037/a0021276

Maloney, E. A., Waechter, S., Risko, E. F., & Fugelsang, J. A. (2012). Reducing the sex difference in math anxiety: The role of spatial processing ability. *Learning and Individual Differences*, *22*, 380–384. https://doi.org/10.1016/j.lindif.2012.01.001

Menon, V. (2015a). Large-scale functional brain organization. In A. W. Toga (Ed.), *Brain mapping: An encyclopedic reference*. Elsevier, Academic Press. https://doi.org/10.1016/B978-0-12-397025-1.00024-5

Menon, V. (2015b). Salience network. In A. W. Toga (Ed.), *Brain mapping: An encyclopedic reference* (pp. 597–611). Elsevier, Academic Press. https://doi.org/10.1016/B978-0-12-397025-1.00052-X

Menon, V., Cohen Kadosh, R., & Dowker, A. (2014). Arithmetic in the child and adult brain. In R. Cohen Kadosh & A. Dowker (Eds.), *Handbook of mathematical cognition*. Oxford University Press.

Menon, V., Rivera, S. M., White, C. D., Glover, G. H., & Reiss, A. L. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. *NeuroImage*, *12*, 357–365. https://doi.org/10.1006/nimg.2000.0613

Metcalfe, A. W., Ashkenazi, S., Rosenberg-Lee, M., & Menon, V. (2013). Fractionating the neural correlates of individual working memory components underlying arithmetic problem solving skills in children. *Developmental Cognitive Neuroscience*, *6*, 162–175. https://doi.org/10.1016/j.dcn.2013.10.001

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Ng, C. T., Chen, Y. H., Wu, C. J., & Chang, T. T. (2022). Evaluation of math anxiety and its remediation through a digital training program in mathematics for first and second graders. *Brain and Behavior: a Cognitive Neuroscience Perspective*, *12*, e2557. https://doi.org/10.1002/brb3.2557

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

Petrides, M. (2005). Lateral prefrontal cortex: Architectonic and functional organization. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*, 781–795. https://doi.org/10.1098/rstb.2005.1631

Pletzer, B. (2016). Sex differences in number processing: Differential systems for subtraction and multiplication were confirmed in men, but not in women. *Scientific Reports*, *6*, 39064. https://doi.org/10.1038/srep39064

Polspoel, B., De Visscher, A., Vandermosten, M., Vogel, S. E., Grabner, R. H., & De Smedt, B. (2019). The neural substrates of the problem size and interference effect in children's multiplication: An fMRI study. *Brain Research*, *1714*, 147–157. https://doi.org/10.1016/j.brainres.2019.03.002

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, *57*, 55–71. https://doi.org/10.1111/0022-4537.00201

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, *14*, 187–202. https://doi.org/10.1080/15248372.2012.664593

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

Rosenberg-Lee, M., Chang, T. T., Young, C. B., Wu, S., & Menon, V. (2011). Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: A cytoarchitectonic mapping study. *Neuropsychologia*, *49*, 2592–2608. https://doi.org/10.1016/j.neuropsychologia.2011.04.035

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, *60*, 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience*, *27*, 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Springer, K. W., Mager Stellman, J., & Jordan-Young, R. M. (2012). Beyond a catalogue of differences: A theoretical frame and good practice guidelines for researching sex/gender in human health. *Social Science & Medicine*, *74*, 1817–1824. https://doi.org/10.1016/j.socscimed.2011.05.033

Stanescu-Cosson, R., Pinel, P., van De Moortele, P. F., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain*, *123*(11), 2240–2255. https://doi.org/10.1093/brain/123.11.2240

Supekar, K., Iuculano, T., Chen, L., & Menon, V. (2015). Remediation of childhood math anxiety and associated neural circuits through cognitive tutoring. *The Journal of Neuroscience*, *35*, 12574–12583. https://doi.org/10.1523/JNEUROSCI.0786-15.2015

Supekar, K., & Menon, V. (2012). Developmental maturation of dynamic causal control signals in higher-order cognition: A neurocognitive network model. *PLoS Computational Biology*, *8*, e1002374. https://doi.org/10.1371/journal.pcbi.1002374

Taddei, M., Bulgheroni, S., Riva, D., & Erbetta, A. (2022). Task-related functional neuroimaging contribution to sex/gender differences in cognition and emotion during development. *Journal of Neuroscience Research*, *101*, 575–603. https://doi.org/10.1002/jnr.25143

Tiberghien, K., De Smedt, B., Fias, W., & Lyons, I. M. (2019). Distinguishing between cognitive explanations of the problem size effect in mental arithmetic via representational similarity analysis of fMRI data. *Neuropsychologia*, *132*, 107120. https://doi.org/10.1016/j.neuropsychologia.2019.107120

Tsui, J. M., & Mazzocco, M. M. M. (2006). Effects of math anxiety and perfectionism on timed versus untimed math testing in mathematically gifted sixth graders. *Roeper Review*, *29*, 132–139. https://doi.org/10.1080/02783190709554397

Voyer, D., & Saunders, K. A. (2004). Gender differences on the mental rotations test: A factor analysis. *Acta Psychologica*, *117*, 79–94. https://doi.org/10.1016/j.actpsy.2004.05.003

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*, 274–290. https://doi.org/10.1111/j.1745-6924.2009.01125.x

Wu, S. S., Chang, T. T., Majid, A., Caspers, S., Eickhoff, S. B., & Menon, V. (2009). Functional heterogeneity of inferior parietal cortex during mathematical cognition assessed with cytoarchitectonic probability maps. *Cerebral Cortex*, *19*, 2930–2945. https://doi.org/10.1093/cercor/bhp063

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

Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: A review of literature. *International Education Journal*, *8*(2), 187–203.

## 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:** Chen, N.-F., & Chang, T.-T. (2023). Arithmetic problem size modulates brain activations in females but not in males. *European Journal of Neuroscience*, *58*(5), 3299–3314. https://doi.org/10.1111/ejn.16100