Exploring the Neural Representation of Elementary Math Concepts

Samuel Debray (samuel.debray@ens-paris-saclay.fr)
Cognitive Neuroimaging Unit (UNICOG), INSERM, CEA, CNRS, Université Paris-Saclay, NeuroSpin center
Bâtiment 145, CEA/SAC/DRF/Joliot, Point Courrier 156 F-91191 Gif-sur-Yvette, France

Daniela Valerio (daniela.valerio@cea.fr)
Cognitive Neuroimaging Unit (UNICOG), INSERM, CEA, CNRS, Université Paris-Saclay, NeuroSpin center
Bâtiment 145, CEA/SAC/DRF/Joliot, Point Courrier 156 F-91191 Gif-sur-Yvette, France

Alireza Karami (alireza.karami@cea.fr)
Cognitive Neuroimaging Unit (UNICOG), INSERM, CEA, CNRS, Université Paris-Saclay, NeuroSpin center
Bâtiment 145, CEA/SAC/DRF/Joliot, Point Courrier 156 F-91191 Gif-sur-Yvette, France

Maxime Cauté (maxime.caute@cea.fr)
Cognitive Neuroimaging Unit (UNICOG), INSERM, CEA, CNRS, Université Paris-Saclay, NeuroSpin center
Bâtiment 145, CEA/SAC/DRF/Joliot, Point Courrier 156 F-91191 Gif-sur-Yvette, France

Stanislas Dehaene (stanislas.dehaene@cea.fr)
Chaire Psychologie cognitive expérimentale, Collège de France, Université PSL
Cognitive Neuroimaging Unit (UNICOG), INSERM, CEA, CNRS, Université Paris-Saclay, NeuroSpin center
Bâtiment 145, CEA/SAC/DRF/Joliot, Point Courrier 156 F-91191 Gif-sur-Yvette, France

Abstract

Mathematical cognition engages a distributed network of brain regions, but the fine-grained organization of this network remains unclear. Using high-resolution 7T fMRI, we investigate how elementary math concepts from two domains – arithmetics (integers and fractions) and geometry (shapes) – are represented in the brain. We test whether these concepts are neurally organized not only by category but also according to shared numerical magnitude.

Behavioral similarity judgments reveal consistent cross-category associations between concepts with equivalent magnitudes (e.g. "three" and "triangle"). In the brain, we observe distinct activation patterns for arithmetic versus geometric items, and further differentiation between integers and fractions – refining prior accounts of the math network's structure. Although direct magnitude-based correspondence was not detected in neural similarity patterns, semantic distances derived from GloVe embeddings significantly predict both behavioral judgments and neural representations in parietal and temporal regions. These findings offer new insights into how mathematical concepts are structured and encoded in the human brain.

Keywords: mathematical cognition; 7T fMRI; neural vector representation; representational similarity analysis; GloVe

Introduction

Amalric and Dehaene (2016, 2019) identified a coarse-grained brain network associated with mathematical processing. Leveraging high-resolution 7T fMRI, this study seeks to refine and better characterize this network by examining its structure with greater specificity. Our investigation focuses on fundamental math concepts, particularly numbers and geometric shapes. Given the potential challenges in deriving distinct vector embeddings for each concept, we adopt an alternative approach by assessing *magnitude-based correspondence*.

Specifically, we hypothesize that neural embeddings carry information not only about category membership but also about shared magnitude – an idea akin to the concept of second-order isomorphism coined by Shepard and Chipman (1970). For example, both "three" and "triangle" imply a magnitude of three, despite belonging to different conceptual categories. We test whether such cross-category magnitude-based correspondences are reflected in behavioral similarity judgments and brain activation patterns, aiming to bridge high-level semantic organization with localized neural coding of math concepts.

Methods

Eighteen healthy participants took part in a two-run fMRI experiment where they were asked to judge the conceptual similarity between written elementary math concepts belonging to three different categories: *integers* (from zero to six), *fractions* ("half", "third" and "fourth") and *shapes* ("segment", "triangle", "square", "pentagon" and "hexagon"). A summary of the task is given in fig. 1.

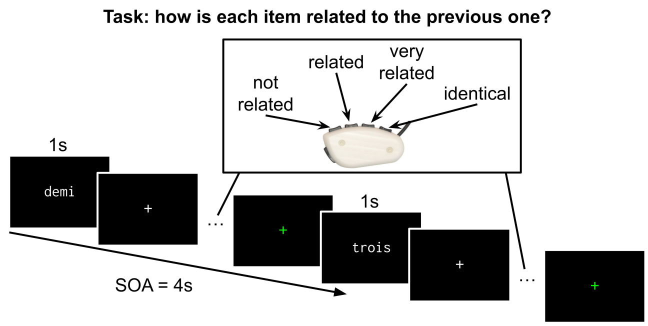

Figure 1: fMRI task

Time series of both runs were then entered into multiple linear regression with predictors of interest including the onsets for the math items, the button presses, and participant RT.

We also ran representational similarity analyses (Kriegeskorte et al., 2008) on the beta maps extracted for each math item. Representational dissimilarity matrices (RDMs) were computed using Pearson correlation distance, and their averages across participants were compared against a set of theoretical RDMs using multiple regression.

Results

Magnitude-based Correspondence in Behavior

The behavioral RDM (fig. 2C) was entered into a multiple linear regression with predictors including indicators for integers, fractions, shapes, and magnitude-based correspondence between these (e.g. "three"–"triangle"). We found an effect of all three possible kinds of correspondence: integers–fractions, integers–shapes and fractions–shapes. These are visible on fig. 2C as the small diagonals on the dissimilarity matrix.

Distinction Between Math Categories in fMRI Data

Inherent to the design of the experiment was the fact that number words are consistently shorter than names of geometric shapes. We show on fig. 2A the linear contrast for items length, activating regions associated with visual processing in the occipital cortex.

At the univariate level, fig. 2A shows clusters of brain activation for the numbers (i.e. integers and fractions) versus geometric shapes contrast that are not lit up by the effect of the length of the items. This is consistent with previous findings that arithmetics and geometry elicit slightly different parts of the math network (Amalric & Dehaene, 2016). In addition, the resolution of 7T fMRI enabled us to find differences in activations between integers and fractions. This, to our knowledge, is a novel finding, and a first step towards identifying cortical patches for individual math concepts.

We also performed multivariate pattern analysis to probe both effects of categories and of magnitude-based correspondence. We restricted this analysis to six regions of interest that

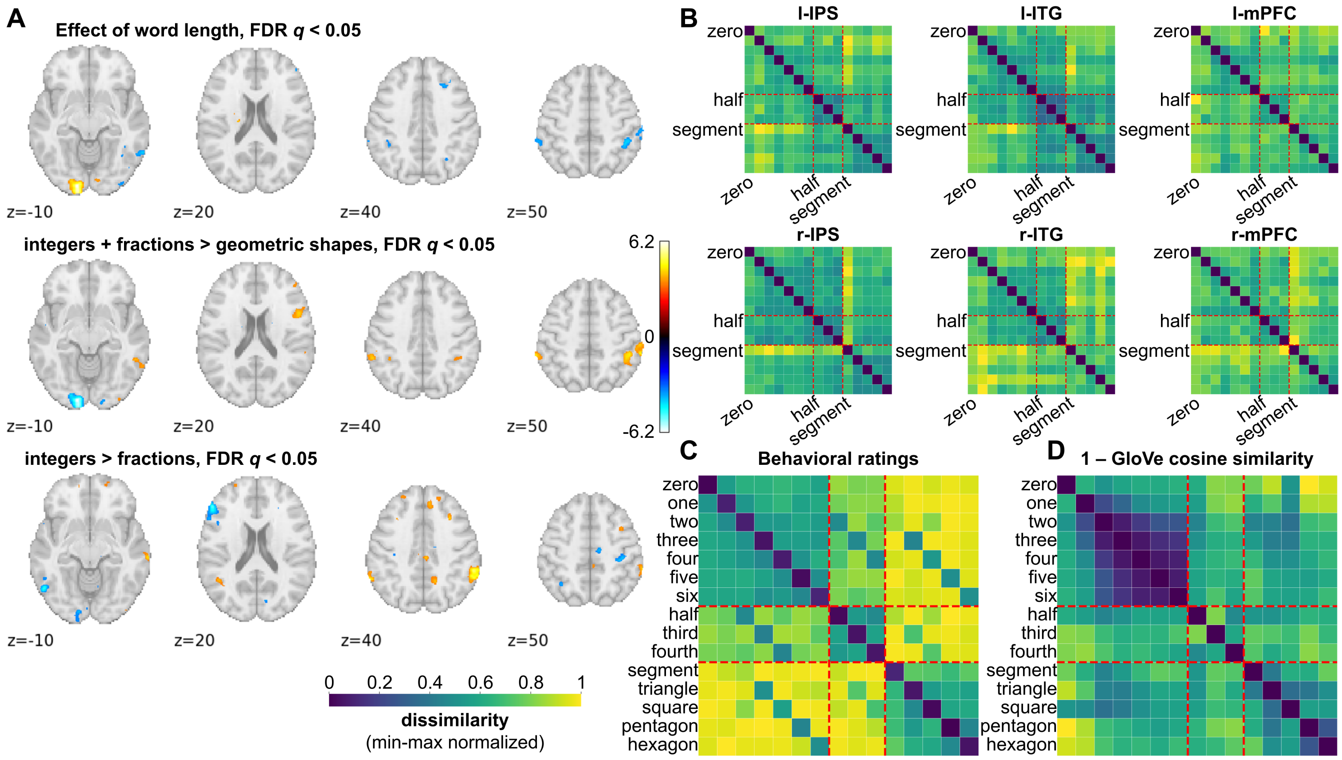

Figure 2: **(A)** Second level contrasts: effect of the length of the fifteen math items, contrast for numbers (integers and fractions) versus geometric shapes, and contrast for integers versus fractions. FDR corrected at $q < .05$. **(B)** Brain RDMs in six math-related regions of interest: bilateral IPS, ITG and mPFC. Averaged across participants. **(C)** Average behavioral dissimilarity matrix collected using the fMRI task. **(D)** RDM obtained from GloVe embeddings using cosine distance.

are known to be involved in mathematical cognition (Amalric & Dehaene, 2016): bilateral intra-parietal sulci (IPS), inferio-temporal gyri (ITG) and medial prefrontal cortices (mPFC). The brain RDMs in these regions are shown on fig. 2B. We again found various effects of categories (integers, fractions, numbers and shapes), but still no trace of magnitude-based correspondence.

Additional Modeling of Brain RDMs using GloVe

As the theoretical RDMs based on magnitude-based correspondence did not reliably predict brain RDMs, we instead used GloVe (Pennington et al., 2014), an unsupervised learning algorithm that generates word embeddings by analyzing word co-occurrence statistics from a corpus. More specifically, we used GloVe cosine distances (fig. 2D, vectors from Debray and Dehaene, 2025), as GloVe has been shown to capture similar patterns of correspondence in other contexts (Debray & Dehaene, 2025; Mikolov et al., 2013).

We found that GloVe cosine distance was a good predictor of item dissimilarity judged by participants ($R^2 = 0.6$, $F(3,221) = 107.38$, $p < .001$). This is a bit more than the 47% of explainable variance reported by Debray and Dehaene (2025), which is probably due to the fact that we only consider a small subset of elementary math items.

Additionally, GloVe RDM significantly predicted brain RDMs in temporal and frontal regions (one-sample t-test on individual betas, l-ITG: $t(17) = 2.6$, $p = 0.02$, mean $\beta = 0.03$; r-ITG: $t(17) = 3.33$, $p = 0.004$, mean $\beta = 0.05$; l-mPFC: $t(17) = 3.05$, $p = 0.007$, mean $\beta = 0.02$; r-mPFC: $t(17) = 2.74$, $p = 0.01$, mean $\beta = 0.02$). This suggests that these embeddings might capture mathematical structure not evident in the theoretical models of categories and magnitude-based correspondence.

Conclusions

Our findings demonstrate that high-field fMRI can reveal fine distinctions within the brain's math network, including category-specific activations for arithmetic versus geometry, and for integers versus fractions. Behavioral data indicate that people intuitively relate math concepts across categories based on shared numerical magnitude, but this structure was not detected in multivariate neural patterns across core math-related regions.

However, the fact that GloVe embeddings predicted both behavioral and brain similarity patterns suggests that these may capture latent structure that current neural analyses do not. Thus, more research is needed to probe effects of magnitude-based correspondence in brain activation patterns.

Acknowledgments

We thank Christophe Pallier for his valuable scientific input on this project; Antonio Moreno for sharing his stimuli for the localizer; Thomas Dighiero-Brecht, Manon Pietrantoni, Bosco Taddei, Antoine Grigis and Alexander Paunov for sharing scripts and insights on data analysis; Minye Zhan for helping us set up the fMRI acquisition sequences; and the MRI technicians, the doctors and the nurses from NeuroSpin's Biomedical Research Imaging Cell for their support with participant recruitement and fMRI acquisitions.

This work was supported by INSERM, CEA, Collège de France, Université Paris-Saclay, a PhD grant from École Normale Supérieure Paris-Saclay to Sa.D. and an ERC grant "MathBrain" (ERC2022-ADG 101095866) to St.D.

References

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences*, *113*(18). https://doi.org/10.1073/pnas.1603205113

Amalric, M., & Dehaene, S. (2019). A distinct cortical network for mathematical knowledge in the human brain. *NeuroImage*, *189*. https://doi.org/10.1016/j.neuroimage.2019.01.001

Debray, S., & Dehaene, S. (2025). Mapping and modeling the semantic space of math concepts. *Cognition*, *254*. https://doi.org/10.1016/j.cognition.2024.105971

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://www.frontiersin.org/articles/10.3389/neuro.06.004.2008

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality [arXiv:1310.4546 [cs, stat]]. http://arxiv.org/abs/1310.4546

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. https://doi.org/10.3115/v1/D14-1162

Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, *1*(1). https://doi.org/10.1016/0010-0285(70)90002-2