# Which Neural Networks model human symbolic shape perception?

**Maxence Pajot (maxence.pajot@gmail.com)**
Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France

**Théo Morfoisse**
Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France

**Mathias Sablé-Meyer**
Sainsbury Wellcome Centre for Neural Circuits and Behaviour, University College London, London, UK

**Yair Lakretz**
Laboratoire de Sciences Cognitives et Psycholinguistique, Dept d'Etudes Cognitives, ENS, PSL University, EHESS, CNRS, 75005 Paris, France

**Stanislas Dehaene**
Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France
Collège de France, Université Paris Sciences Lettres (PSL), 75005 Paris, France

## Abstract

**While impressive in many vision tasks, artificial neural networks have proved lacking in the field of symbolic shape perception. In this present work, we evaluate the ability of Convolutional Neural Networks (CNNs) and Vision Transformers, with varying sizes and training datasets, to recognize and process abstract shapes. We compare the models' internal representations to human data collected from an outlier detection task on quadrilaterals. We find that networks trained on a large amount of data achieve human-like representations of the tested shapes.**

**Keywords:** geometrical cognition; modeling; Convolutional Neural Networks; Vision Transformers

## Introduction

The ability to understand and generate abstract shapes has been considered a hallmark of human cognition (Dehaene et al., 2022; Sablé-Meyer et al., 2021). While research on non-human primates and birds has explored their capacity for visual processing, there is little evidence that these species can recognize, generate, or manipulate abstract shapes in a human-like manner (Sablé-Meyer et al., 2021; Westphal-Fitch et al., 2012). Even with extensive training, such behaviors remain absent in non-human animals, whereas human children engage with abstract figures effortlessly (Saito et al., 2014).

Building on the language of thought hypothesis introduced in *The Language of Thought (Fodor, 1975)*, Sablé-Meyer et al. (2022) proposed that abstract shapes are represented symbolically in the brain. While the symbolic model offers a compelling framework for understanding the types of computations the brain might perform when processing abstract shapes, the neural mechanisms underlying these computations remain unclear.

More recently, (Campbell et al., 2024) showed that neural networks with number of parameters in the billions seem to possess capabilities to handle abstract shapes similar to humans. Our research investigates which neural networks can simulate human abstract shape recognition by assessing the cognitive plausibility of two prominent computer vision architectures: Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). By understanding which factors allow neural networks to process abstract shapes efficiently, we hope to shed light on how humans developed this ability.

Following work done by (Sablé-Meyer et al., 2024), we investigate the representation of quadrilaterals with varying levels of complexity, comparing how these shapes are encoded by humans and neural networks.

## Methods

To compare how neural networks and humans represent quadrilaterals, we applied Representational Similarity Analysis (RSA) (Kriegeskorte et al., 2008). For the human data, we used behavioral results from (Sablé-Meyer et al., 2024), who constructed a Representational Dissimilarity Matrix (RDM) based on success rates and reaction times in an outlier detection task (figure 1A).

To define the RDM of a neural network, we measure the dissimilarity between two quadrilaterals using the Euclidean distance between their respective prototypes. Each prototype represents the average activation of that quadrilateral in the embedding space, computed across various scales and rotations. We use the embeddings at the last layer before classification, as it is where correlation with human RDM is higher for most networks.

To investigate which components are essential for neural networks to develop human-like abstract shape perception, we evaluated vision models with diverse characteristics. Specifically, we tested two architectural families—Vision Transformers (Dosovitskiy et al., 2021) and Convolutional Neural Networks (CNNs)—across three training objectives: classification, Dino (Oquab et al., 2024), and CLIP (Radford et al., 2021). These models also varied in terms of number of parameters and the size of their training datasets. Additionally, we assessed the correlation with a symbolic model proposed by (Sablé-Meyer et al., 2021). This model represents quadrilaterals using properties such as symmetry and the presence of parallel lines, and computes dissimilarity as the L1 distance between the corresponding property vectors.

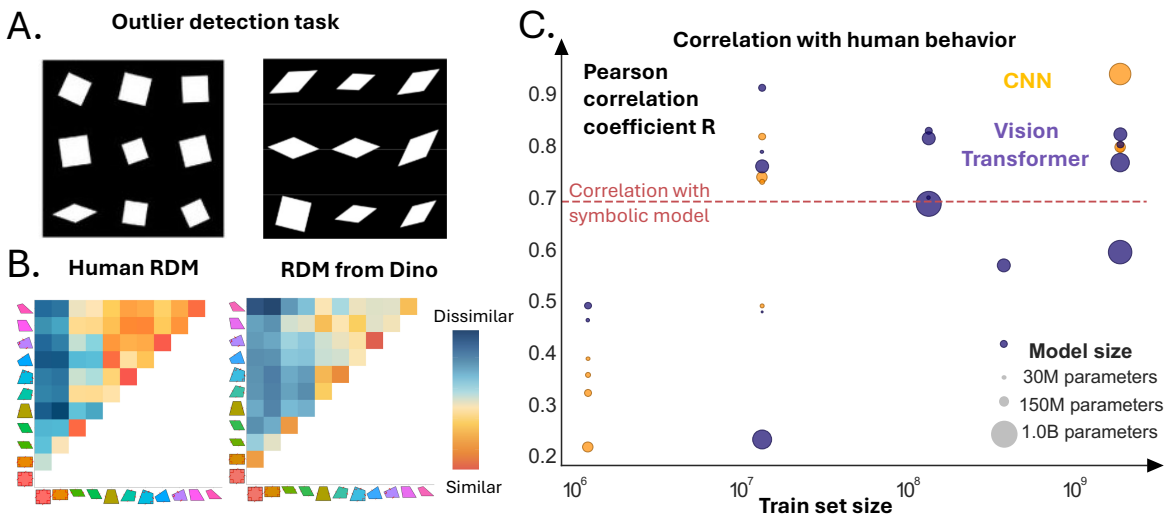


Figure 1: **(A)** Example trials in the outlier detection task. **(B)** Representational Dissimilarity Matrix (RDM) extracted from human behavioral data (left), and RDM from Dino's embeddings (right). **(C)** Correlations between different model RDMs and human behavioral RDM, plotted as a function of the size of the training dataset on a log-scale. Size of the dot is proportional to the size of the model, and color indicates the architecture.

## Results

As reported by (Campbell et al., 2024), we also find that large neural networks, such as Dino, exhibit representations of quadrilaterals that closely resemble those of humans (figure 1B). However, it is unclear if there exists a single parameter that allows for such human-like representations. As shown in figure 1C, all models trained on the smallest dataset, Imagenet-1k, display low correlation with human behavior. Conversely, all but one of the networks trained on LAION-2B—the largest dataset—show higher correlations than the symbolic model. Still, the size of a model is a big confounding factor, as only big models are trained on very large datasets, and inversely those big models are not trained only on smaller datasets.

Interestingly, we find that even relatively small models, when trained on moderately large datasets, can develop representations that align closely with human behavior. Meanwhile, some larger models trained on substantially bigger datasets fail to do so.

Overall, model architecture appears to have minimal influence on the similarity between neural network and human RDMs. Both CNNs and Vision Transformers benefit from increases in model size and the number of training images, suggesting that scaling is a more critical factor than architectural differences.

## Discussion

In this study, we demonstrated that both Convolutional Neural Networks and Vision Transformers are capable of modeling human-like symbolic shape perception. Among the factors considered, the size of the training dataset emerged as the most influential in determining how closely a network's representations align with that of humans.

We could have expected ViTs to fare better than CNNs, given prior findings that Vision Transformers rely more on shape cues than texture, aligning more closely with certain characteristics of human visual perception. (Naseer et al., 2021; Tuli et al., 2021). Overall, these findings align with theories suggesting that human intelligence emerged from increased information capacity (Cantlon & Piantadosi, 2024), rather than from the presence of specific architectural biases.

## References


Campbell, D., Kumar, S., Giallanza, T., Griffiths, T. L., & Cohen, J. D. (2024). *Human-Like Geometric Abstraction in Large Pre-trained Neural Networks* (No. arXiv:2402.04203). arXiv. https://doi.org/10.48550/arXiv.2402.04203

Cantlon, J. F., & Piantadosi, S. T. (2024). Uniquely human intelligence arose from expanded information capacity. *Nature Reviews Psychology*, *3*(4), 275–293. https://doi.org/10.1038/s44159-024-00283-3

Dehaene, S., Roumi, F. A., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: A hypothesis about human singularity. *Trends in Cognitive Sciences*, *26*(9), 751–766. https://doi.org/10.1016/j.tics.2022.06.010

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (No. arXiv:2010.11929). arXiv. http://arxiv.org/abs/2010.11929

Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Naseer, M., Ranasinghe, K., Khan, S., Hayat, M., Khan, F. S., & Yang, M.-H. (2021). *Intriguing Properties of Vision Transformers* (No. arXiv:2105.10497). arXiv. https://doi.org/10.48550/arXiv.2105.10497

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., … Bojanowski, P. (2024). *DINOv2: Learning Robust Visual Features without Supervision* (No. arXiv:2304.07193). arXiv. http://arxiv.org/abs/2304.07193

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision* (No. arXiv:2103.00020). arXiv. http://arxiv.org/abs/2103.00020

Sablé-Meyer, M., Benjamin, L., Watkins, C. P., He, C., Roumi, F. A., & Dehaene, S. (2024). *Two brain systems for the perception of geometric shapes*. https://doi.org/10.1101/2024.03.13.584141

Sablé-Meyer, M., Fagot, J., Caparos, S., van Kerkoerle, T., Amalric, M., & Dehaene, S. (2021). Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity. *Proceedings of the National Academy of Sciences*, *118*(16), e2023123118. https://doi.org/10.1073/pnas.2023123118

Saito, A., Hayashi, M., Takeshita, H., & Matsuzawa, T. (2014). The Origin of Representational Drawing: A Comparison of Human Children and Chimpanzees. *Child Development*, *85*(6), 2232–2246. https://doi.org/10.1111/cdev.12319

Tuli, S., Dasgupta, I., Grant, E., & Griffiths, T. L. (2021). *Are Convolutional Neural Networks or Transformers more like human vision?* (No. arXiv:2105.07197). arXiv. http://arxiv.org/abs/2105.07197

Westphal-Fitch, G., Huber, L., Gómez, J. C., & Fitch, W. T. (2012). Production and perception rules underlying visual patterns: Effects of symmetry and hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2007–2022. https://doi.org/10.1098/rstb.2012.0098