Two brain systems for the perception of geometric shapes

Mathias Sablé-Meyer[*1,2,3], Lucas Benjamin[2], Cassandra Potier Watkins[1], Chenxi He[2], Fosca Al Roumi[2] and Stanislas Dehaene[1,2]

[1]Collège de France, Université Paris-Sciences-Lettres (PSL), 11 Place Marcelin Berthelot, 75005 Paris, France
[2]Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay, NeuroSpin Center, 91191 Gif/Yvette, France
[3]Sainsbury Wellcome Centre for Neural Circuits and Behaviour, University College London, London W1T 4JG, UK
[*]Corresponding author. Email: mathias.sable-meyer@ucl.ac.uk

One-Sentence Summary: In human adults and children, perceiving a geometric shape involves distinct visual and symbolic brain representations.

Abstract: Many human cultures produce and enjoy geometric signs, a uniquely human trait whose neural mechanisms are unknown. We formulate and test the hypothesis that, beyond an evolutionarily ancient ventral visual circuit for object recognition, the perception of geometry also relies on an additional system encoding discrete regularities such as symmetries and parallelism. Functional MRI and magnetoencephalography, in adults and six-year-olds, supports this hypothesis: while classical convolutional neural networks capture the early visual activity evoked by geometric shapes, subsequent signals from a dorsal parietal and prefrontal network arise from a distinct representation of discrete mathematical features. Thus, the mere perception of a regular quadrilateral suffices to engage a mathematically oriented mode of perception, inadequately captured by current neural networks models.

> In geometry, the essential is invisible to the eyes, one sees well only with the mind.
>
> Emmanuel Giroux, French blind mathematician

Long before the invention of writing, prehistoric humans drew non-pictorial geometric signs such as zigzags or triangles (*1*, *2*). Human cultures use geometrical regularities such as parallelism and symmetry in their decorative arts, tools, buildings, graphics and maps (*3*). Cognitive studies suggest that all humans, even in the absence of formal western education, entertain abstract intuitions of geometric concepts such as points, lines, and surfaces (*4*, *5*). Those capacities have not been found in non-human primates (*6*–*9*), suggesting that the study of the origins of geometry in humans may ultimately shed light on the cognitive singularity of humans (*9*).

Little is known, however, about the neural mechanisms of the perception of geometric shapes. While there is extensive brain-imaging research on all sorts of visual categories such as faces, tools or houses (*10*–*16*), geometric shapes are an exception. Some studies analyzed the brain activation evoked by outlines of objects or animals (*17*–*21*) or arbitrary shapes (*22*–*24*), but they did not ask how humans encode the geometry of figures as simple as a square.

Here, we show that geometric shape perception calls upon a distinct brain system for extracting the mathematical regularities of shapes. We capitalize on a recent psychophysical study of quadrilateral shapes of variable complexity (*7*), which found that human adults and children of Western or non-Western cultures detect an intruder more easily among shapes of high regularity, such as squares or rectangles, than among irregular shapes. When baboons were trained to perform the same task, they showed no such geometric regularity effect. Baboon behavior was well accounted for by Convolutional Neural Network (CNN) models of object recognition, but human behavior could only be explained by appealing to a representation of discrete geometric properties of parallelism, right angle and symmetry, in this and other tasks (*25*).

We therefore formulated the hypothesis of two distinct pathways for geometric shape perception: in addition to the ventral visual circuit for object recognition, which is similar in human and non-human primates, human geometry perception relies on a distinct "language of thought": an inner combinatorial language which encodes discrete mathematical regularities and their combinations (*7*, *9*, *25*). Counterintuitively, the hypothesis predicts that human geometric signs such as a square, although considered the most elementary of shapes, are not recognized quickly and solely using the ventral visual stream, but require a later stage of geometrical feature encoding in math-responsive dorsal parietal and inferotemporal areas (*26*, *27*). Here, we assessed these spatiotemporal predictions using two neuroimaging techniques (functional MRI and magnetoencephalography) and two age groups: adults, and children in 1st grade (6-years-old, this year was selected as it marks the first year French students receive formal instruction in mathematics).

Estimating the Representational Similarity of Quadrilaterals

Our study focused on 11 quadrilaterals tightly matched in average pairwise distance between vertices and length of the bottom side, yet differing broadly in geometric regularity (**Fig.1A)**. To evaluate their perceptual similarity in an objective manner, we assessed the difficulty of visual search for one shape among rotated and scaled versions of the other (**Fig.1B**) (*28*, *29*). The 11x11 dissimilarity matrix, estimated by aggregating response time and errors from n=330 participants, is shown in **Fig.1D**, together with its 2-dimensional ordinal Multi-Dimensional Scaling (MDS) projection in **Fig.1C** (*30*). As predicted,

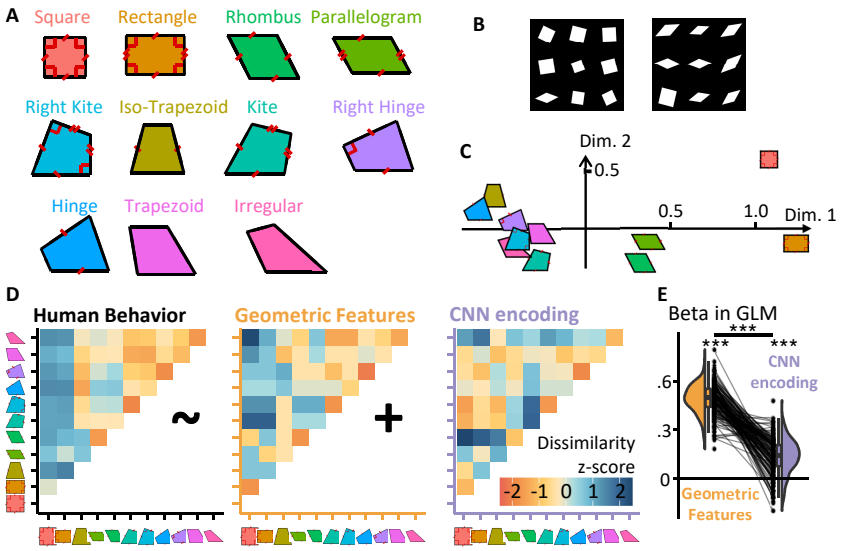

Fig.1. Measuring and modeling the perceptual similarity of geometric shapes. (**A**) The eleven quadrilaterals used throughout the experiments (colors are consistently used in all other figures). (B) Sample displays for the behavioral visual search task used to estimate the 11x11 shape similarity matrix. Participants had to locate the deviant shape. The right insert shows two trials from the behavioral visual task search task, used to estimate the 11x11 shape similarity matrix. Participants had to find the intruder within 9 shapes. (C) Multidimensional scaling of human dissimilarity judgments. (**D**) The behavioral dissimilarity matrix (left) was better captured by a symbolic coding model (middle) than by a convolutional neural network (right). The graph at right shows the GLM coefficients for each participant.

human perceptual similarity was primarily driven by geometric properties: in that 2D projection, the vector which corresponded to simply counting the number of geometric regularity (number of right angles, pairs of parallel lines, pairs of equal angles and pair of sides of equal length) was significantly different from 0 for both axes (both p<.01). This is especially visible with the projection of shapes on the first dimension with the square and the rectangle scoring at one extreme, rhombus and parallelogram in the middle, and less regular shapes at the other extreme.

We compared human similarity matrices to those generated by two competing models of shape processing (*7*). The geometric feature model proposes that each shape is encoded by a feature vector of its discrete geometric regularities and predicts similarity by counting the number of shared features: this makes squares and rectangles very similar, but squares and irregular quadrilaterals very dissimilar. The visual model proposes that images are propagated through a feedforward CNN model of the ventral visual pathway (we use Cornet-S, but see **Fig.S1** in Supplementary On-line Materials for other CNNs, with unchanged conclusions). Shape similarity is estimated as the crossnobis distance between activation vectors in late layers (*31*). The two models are uncorrelated (r^2=.04, p>.05).

Multiple regression of human behavior with symbolic model and visual model (**Fig.1E**) showed that both contributed to explaining perceived similarity (p*s*<.001), but that the weight of the geometric feature model was 3.6 times larger than that of the visual model, a significant difference (p<.001). This finding supports the prior proposal that the two strategies contribute to human behavior, but that the symbolic one dominates, especially in highly educated humans (*7*).

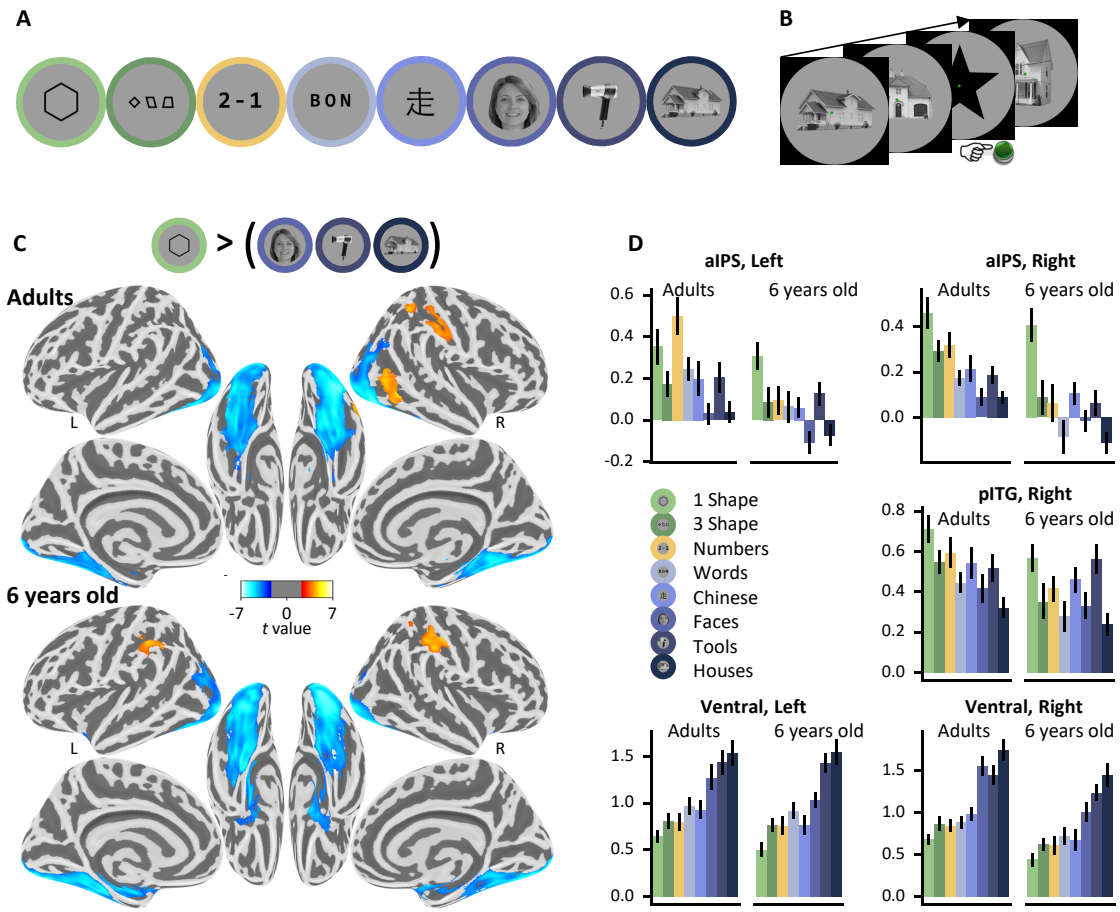

Fig.2 Localizing the brain systems involved in geometric shape perception (**A**) Visual categories used in the localizer (face displayed here is artificially generated for anonymization reason, see appendix for stimuli details). (**B**) Task: In each miniblock, among a series of 6 pictures from a given category, participants had to detect a rare target star. (**C**) Statistical map associated with the contrast "single geometric shape > faces, houses and tools", projected on an inflated brain (top: adults; bottom: children; clusters significant at cluster-corrected p<.05 with nonparametric two-tailed bootstrap test as reported in the text). (**D**) BOLD response amplitude (regression weights, arbitrary units) within each significant cluster with subject-specific localization. Geometric shapes activate the intraparietal sulcus (IPS) and posterior inferior temporal gyrus (pITG), while causing reduced activation in broad bilateral ventral areas compared to other stimuli; see Fig.S4 for analysis of subject-specific ventral subregions.

Localizing Neural Systems for Geometric Shape Perception

To test the hypothesis that the perception of geometric shapes goes beyond ventral visual cortex, we tested 20 adults in functional MRI (9 females; 19-37 years old, mean age 24.6) on a localizer task. Three localizer runs, using a fast block design, probed whole-brain responses to geometric shapes (triangles, squares, hexagons, etc.) and to a variety of matched visual categories (faces, objects, houses, arithmetic, words and Chinese characters). For the latter, we replicated the classical ventral visual category-specific responses (**Fig.2**; **Fig.S2)**. However, when contrasting geometric shapes to faces, houses or tools, we observed a massive reduction in the activation of bilateral ventral occipito-temporal areas (unless otherwise stated, all statistics are at voxelwise p<0.001, clusterwise p<0.05 corrected for multiple comparisons across the whole brain). None of the regions specialized for words, faces, tools or houses responded to geometric shapes (**Fig.2C; Table S1**). Instead, we observed superior activation to shapes in only two significant positive clusters, both outside the ventral visual pathways, in the right anterior intraparietal sulcus

(aIPS) and posterior inferior temporal gyrus (pITG) bordering on the lateral occipital sulcus. At a lower threshold (voxel p<0.001 uncorrected), the symmetrical aIPS in the left hemisphere was also activated (also see SOM for additional results concerning the "3-shapes" condition).

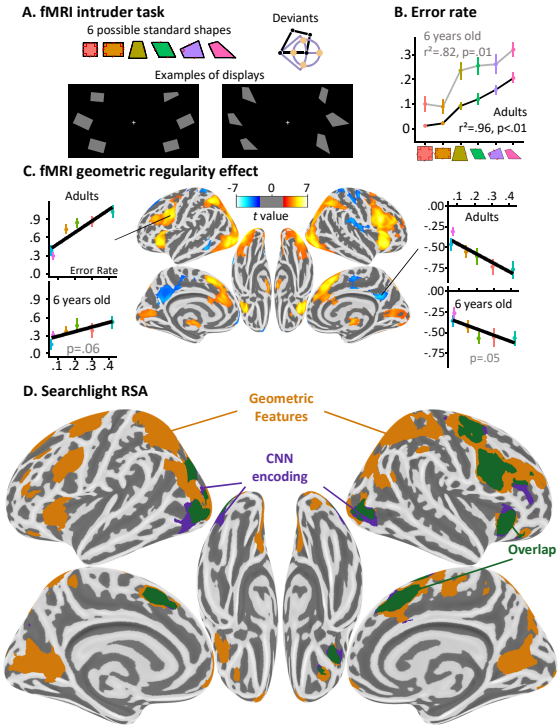

Fig.3 Dissociating two neural pathways for the perception of geometric shape. (A) fMRI intruder task. Participants indicated the location of a deviant shape via button clicks (left or right). Deviants were generated by moving a corner by a fixed amount in four different directions. **(B)** Performance inside the fMRI: both populations tested displayed an increasing error rate with geometric shape complexity, which significantly correlates with previous data collected online. **(C)** Whole brain correlation of the BOLD signal with geometric regularity in adults, as measured by the error rate in a previous online intruder detection task (*7*). Positive correlations are shown in red and negative ones in blue. Voxel threshold p<.001, cluster-corrected by permutation at p<.05. Side panels show the activation in two significant ROIs whose coordinates were identified in adults and where the correlation was also found in children (one-tailed test, corrected for the number of ROIs tested this way). **(D)** Whole-brain searchlight-based RSA analysis in adults (same statistical thresholds). Colors indicate the model which elicited the cluster: purple for CNN encoding, orange for the geometric feature model, green for their overlap.

Those areas are similar to those active during number perception, arithmetic, geometric sequences, and the processing of high-level math concepts (*9*, *26*, *27*). Their activation could therefore indicate that geometric shapes are encoded as a mental expression that combines number, length, angle and other geometric features, in agreement with our hypothesis, but it could also reflect an association with math knowledge acquired during education. To test this alternative, we replicated our fMRI study in 22 6-year-old first graders, in whom we observed the same reduction in ventral visual activity, as well as the same aIPS activity, now significant in both hemispheres. Furthermore, the right pITG voxels extracted from adults also reached significance in children, and conversely for the left aIPS. Indeed, activation profiles were quite similar in both age groups (**Fig.2D, see Fig.S3A** and **Fig.S3B** for uncorrected statistical maps where similarities are striking), with aIPS activity being selective to geometric shapes in children and shared with arithmetic in adults (*27*, *32*). Thus, geometric shapes activate these regions even before they become responsive to complex math knowledge and theorems.

fMRI data allowed us to reject several plausible alternative hypotheses on the encoding of geometric shapes (see Appendix and **Fig.S4**). First, geometric shapes might have been processed similarly to other cultural symbols learned at school, such as letters and written words. However, in a subject-specific ROI analysis of the visual word form area (VWFA), identified by its stronger response to alphabetical stimuli, no activity was evoked by single shapes, or strings of three shapes, above the level of other image categories. Second, our shapes, which differed along quantitative dimensions, could have been encoded in the fusiform face (FFA), since face-selective neurons are sensitive to such quantitative dimensions of face variation (*33*) – but the FFA showed the lowest responses to geometric shapes in the fusiform face area. Third, geometric shapes could have been processed together with other complex man-made objects, which are often designed to be regular and symmetrical. However, they failed to activate ventral visual voxels selective to tools, thus refuting this possibility. Finally, geometric shapes could have been encoded in the parahippocampal place area (PPA), which is known to encode the geometry of scenes, including abstract Lego-block ones (*34*), but they actually induced minimal activity in individually defined

PPA voxels (see **Fig.S4** for a summary). In sum, geometric shapes did not seem to be processed ventrally alongside other visual images, but to activate math- and number-related areas, particular in right aIPS.

Next, we measured the detailed fMRI patterns evoked by quadrilaterals, in order to submit them to a representational similarity analysis and test our hypothesis of a double dissociation between regions encoding visual (CNN) versus symbolic codes. After the localizer, adults and children performed an intruder task similar to our previous behavioral study (*7*), with miniblocks allowing us to evaluate the activity pattern evoked by a given standard shape. To render the task doable by 6-year-olds, we tested only 6 quadrilaterals, and merely asked participants whether the intruder was on the left or the right (**Fig.3**). Behavioral performance inside the scanner replicated the geometric regularity effect in adults and children: performance was best for squares and rectangles and decreased monotonically for figures with fewer geometric regularities (**Fig.3B**), significantly correlating with an empirical measure of geometric regularity measured with an online intruder detection task (*7*), and used for the following analyses when we refer to geometric regularity. The neural bases of this effect were studied at the whole-brain level by searching for areas whose activation varied monotonically with geometrical regularity. In adults, this contrast identified a broad bilateral dorsal network including occipito-parietal, middle frontal, anterior insula and anterior cingulate cortices, with accompanying deactivation in posterior cingulate (**Fig.3C**). Testing the ROIs from the adults in children, after correcting for multiple comparisons, only the posterior cingulate was significant, though two clusters were very close to significance, both with p=.06: one positive in the left precentral gyrus (shown in **Fig.3C, Fig.S3C**), and one negative in the dorsolateral prefrontal cortex. Testing the ROIs identified in the visual localizer showed that they all exhibited a positive geometric difficulty effect in adults (all p<.05), but did not reach significance in children, possibly due to excessive noise (as indicated by the much higher error rate in the task inside the scanner).

This broad network, however, probably encompassed both regions involved in geometric shape encoding and those involved in the decision about the intruder task. Indeed, a mere increase in activation level could be driven by an overall increase in difficulty, effort and attention when geometric regularity decreased. To isolate the regions specifically involved in shape coding, we performed searchlight RSA analyses, which focus on the pattern rather than the level of activation. Within spheres spanning the entire cortex, we asked in which regions the similarity matrix between the activation patterns evoked by the six shapes was predicted by the CNN encoding model, the geometric feature model, or both (**Fig.3D**; **Table S2)**. In adults, the CNN encoding model predicted neural similarity in bilateral lateral occipital clusters, while the geometric feature model yielded a much large set of clusters in occipito-parietal, superior prefrontal and right dorsolateral prefrontal cortex. Calcarine cortex was also engaged, possibly due to top-down feedback (*35*). Both CNN encoding and geometric feature models had overlapping voxels in anterior cingulate and right premotor cortex, possibly reflecting the pooling of both visual codes in decision-related areas.

In children, no cluster associated to the geometric feature model reached significance in a searchlight analysis at the whole-brain level (**Fig.S3D**). However, a right lateral occipital cluster was significantly captured by the CNN encoding model in children, and its symmetrical counterpart was close to the significance threshold (**Fig.S3D**). This result might indicate that geometric features are not well differentiated prior to schooling. It could also reflect that children weight the symbolic strategy less, as was found in (*7*); combined with difficulty of obtaining precise subject-specific activation patterns in young children, this could make the symbolic strategy harder to localize.

Timing the Neural Responses to Visual and Geometric Features

The fMRI experiment suffered from several limitations: we solely studied six quadrilaterals, used an explicit intruder detection which required the simultaneous presentation of multiple shapes and blended the activations due to shape identification and to task decision (which could even be incorrect); and finally, the poor temporal resolution of fMRI prevented us from tracking the dynamic of mental representations over time. To overcome those issues, we replicated the experiment in adult magnetoencephalography (MEG) with several changes: all 11 quadrilaterals were studied; shapes were presented serially, one at a time, in miniblocks comprising rare intruders; and participants were merely asked to fixate and attend to every shape, without performing any explicit task.

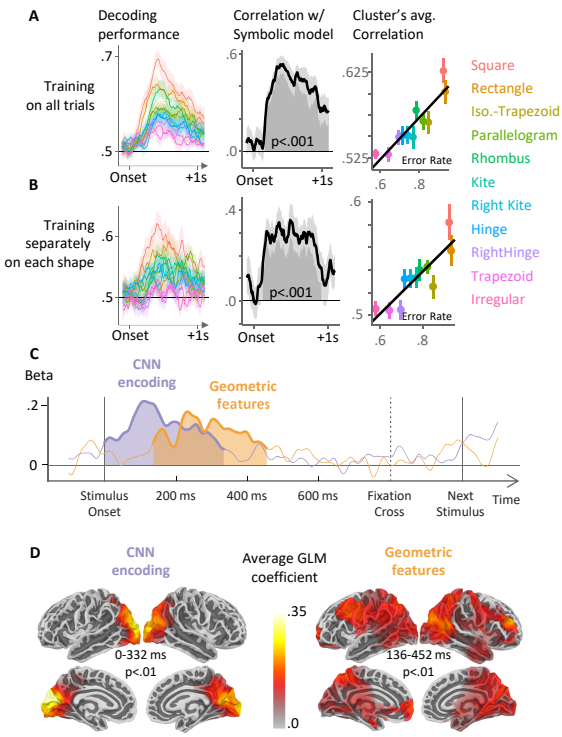

Fig.4 Using MEG to time the two successive neural codes for geometric shapes (**A,B**) Performance of a classifier using MEG signals to predict whether the stimulus is a regular shape or an oddball. Left: performance for each shape; middle: correlation with geometrical regularity (same x axis as in figure 3C); right: visualization of the average decoding performance over the cluster. In A, training of the classifier was performed on MEG signals from all 11 shapes; In B, eleven different classifiers were trained separately, one for each shape. (**C**) Sensor-level temporal RSA analysis. At each time point, the 11x11 dissimilarity matrix of MEG signals was modeled by the two model RDMs in **Fig. 1D**, and the graph shows the time course of the corresponding GLM beta weights. (**D**) Source-level temporal-spatial searchlight RSA. Same analysis as in C, but now after reconstruction of cortical source activity.

In spite of those passive instructions, MEG signals revealed an automatic detection of intruders, again driven by geometric regularity. We trained logistic regression decoders to classify the MEG signals at each time point following a shape as arising from a reference shape or from an intruder (see **SOM** and **Fig.4A**). Although trained on all shapes together, the decoder performed better with geometrically regular shapes than with irregular shapes. Indeed, a regression of decoding performance on geometrical regularity yielded a significant cluster, which first became significant from 176 ms onwards. A geometrical regularity effect was also seen in the latency of the outlier response (**Fig.4A;** correlation between geometric regularity and the latency when average decoding performance first exceeded 57% correct; one-tailed t-test of each participant's regression slope against 0: t=-2.88, p=.005). The same effect was found when training separate outlier decoders for each shape, thus refuting an alternative hypothesis according to which all outliers evoke identical amounts of surprise, but in different directions of neural space (**Fig.4B**). Overall, the results fully replicate prior behavioral observations of the geometric regularity effect (*7*) and suggest that the computation of a geometric code occurs under passive instructions and starts at a latency of ~170 ms.

We next used temporal RSA to further probe those dynamics. For each time point, we estimated the 11x11 neural dissimilarity matrix across all pairs of reference shapes using sensor data with crossnobis distances (*31*), and entered them in a multiple regression with those predicted by CNN encoding and geometric feature models. The coefficients associated to each predictor are shown in **Fig. 4C**. An early cluster (~0-300ms, p<.001) showed a significant correlation of the CNN encoding model on brain similarity. Its early onset, at ~0 ms, is likely due to the fact that (i) a given shape was repeatedly presented

within a block at a constant rate, thus permitting it to be predicted; and (ii) to increase the signal to noise ratio, the data was smoothed with a window of 100ms. It was followed by a significant cluster associated with the geometric feature model (~150-450ms, p<.001). Thus, this analysis supported our hypothesis of two distinct stages in geometric shape perception, and indicated that a symbolic encoding arises around ~200 ms.

To understand which brain areas generated these distinct patterns of activations and replicate the fMRI results, we performed a source reconstruction of our data, then replicated the RSA analysis with searchlights sweeping across cortical patches of 2 cm geodesic radius (**Fig. 4D**). The CNN encoding model captured brain activity in an early bilateral occipital and posterior parietal cluster, while the geometric model accounted for subsequent activity starting at ~200 ms and spanning over broader dorsal occipito-parietal and intraparietal, prefrontal, anterior and posterior cingulate cortices. This double dissociation closely paralleled fMRI results (compare **Fig. 4D** and **Fig. 3D**), with greater spatial spread due to the unavoidable imprecision of MEG source reconstruction.

Discussion

Previous behavioral work suggested that the human perception of geometric shapes cannot be solely explained by current CNN encoding models, and that humans also encode them using nested combinations of discrete geometric regularities (*7*, *9*, *25*). Here, we provide direct human brain imaging evidence for the existence of two distinct cortical networks with distinct representational schemes, timing, and localization: an early occipitotemporal network well modeled by a CNN, and a later dorsal parietal and inferotemporal network sensitive to geometric regularity. The latter areas overlap with those active during the comprehension of elementary and advanced mathematical concepts (*9*, *26*, *27*), in agreement with their proposed involvement in an abstract, modality-independent symbolic encoding. Further support for this idea comes from the fact that these regions can be equally activated in sighted and in blind individuals, whether they perform mathematics (*36*, *37*) or just process the shapes of manmade objects (*38*). Thus, shape representation in these regions is amodal and more abstract than a hierarchy of visual filters.

While CNNs are predictive of early ventral visual activity (*39*), the present work adds to a growing list of their limits as full models of visual perception (*40*, *41*). Beyond the present failure to capture dorsal parietal-inferotemporal processing of geometric shapes, CNNs also fail to model many visual illusions (*40*, *41*), geometrically impossible objects (*42*), global shape, part-whole relations and other Gestalt properties (*40*). These limitations suggest the need to supplement current models not only by increasing their size (*39*) but also by changing the architecture in ways that incorporate findings from psychology and neuroscience (*43*, *44*). A key challenge of the present work lies in its demonstration that human geometry, even in the simplest case of perceiving a square, involves discrete, abstract, symbolic features such parallelism and right angle and their composition into compact expressions with minimal description length (*9*, *45*). The neurophysiological implementation of such symbolic composition is still unknown. In the future, the present paradigm is simple enough to contribute to the search for such a neural code and helps predict where it might be found. Comparing neural responses to quadrilaterals in human and non-human primates could test the hypothesis that CNN-like representations are present in both, while symbolic representations may be unique to humans. The methodology can also be extended to the perception of a broader set of geometric patterns (circles, spirals, crosses, zigzags, plaids, etc) which recur since prehistory and in the drawings of children and adults of various cultures, thus testing whether they too originate in a minimal and universal “language of thought” (*25*).

Acknowledgments:

We are grateful to Ghislaine Dehaene-Lambertz, Leila Azizi, and the NeuroSpin support teams for help in data acquisition, and Lorenzo Ciccione, Christophe Pallier, Minye Zhan, Alexandre Gramfort and the MNE team for support in data processing.

Funding: FYSSEN, INSERM, CEA, Collège de France, ERC MathBrain

Author contributions:

Conceptualization: MSM, SD

Methodology: MSM, SD

Investigation: MSM, LB, CPW, CH, FAR

Visualization: MSM, LB, FAR, SD

Funding acquisition: SD

Project administration: MSM, SD

Supervision: SD

Writing – original draft: MSM, SD

Writing – review & editing: all authors

Competing interests: Authors declare that they have no competing interests

Data and materials availability: scripts for all of the analyses are available at https://github.com/mathias-sm/TwoBrainSystemsGeometricShapes; behavioral data and scripts to generate the models are also available at this url. Neuroimaging data provided upon request.

References

References

1. C. S. Henshilwood *et al.*, *Nature* **562**, 115–118, ISSN: 0028-0836, 1476-4687 (Oct. 2018).
2. B. L. Van der Waerden, *Geometry and Algebra in Ancient Civilizations* (Springer Science & Business Media, 2012).
3. B. Tversky, *Topics in Cognitive Science* **3**, 499–535, ISSN: 1756-8757, 1756-8765 (July 2011).
4. S. Dehaene, V. Izard, P. Pica, E. Spelke, **311**, 5 (2006).
5. V. Izard, P. Pica, E. S. Spelke, S. Dehaene, *Proceedings of the National Academy of Sciences* **108**, 9782–9787 (2011).
6. J. Close, J. Call, *Animal Cognition* **18**, 437–449, ISSN: 1435-9456 (Mar. 1, 2015).
7. M. Sablé-Meyer *et al.*, *Proceedings of the National Academy of Sciences* **118**, ISSN: 0027-8424, 1091-6490, (2021; https://www.pnas.org/content/118/16/e2023123118) (Apr. 20, 2021).
8. A. Saito, M. Hayashi, H. Takeshita, T. Matsuzawa, *Child Development* **85**, 2232–2246, ISSN: 1467-8624 (Dec. 2014).
9. S. Dehaene, F. Al Roumi, Y. Lakretz, S. Planton, M. Sablé-Meyer, *Trends in Cognitive Sciences* **26**, 751–766, ISSN: 1364-6613 (Sept. 1, 2022).

10. M. J. Arcaro, P. F. Schade, J. L. Vincent, C. R. Ponce, M. S. Livingstone, *Nature Neuroscience* **20**, 1404–1412, ISSN: 1097-6256, 1546-1726 (Oct. 2017).

11. E. Margalit *et al.*, *The Journal of Neuroscience* **40**, 3008–3024, ISSN: 0270-6474, 1529-2401 (Apr. 8, 2020).

12. A. M. Rauschecker, R. F. Bowen, J. Parvizi, B. A. Wandell, *Proceedings of the National Academy of Sciences* **109**, ISSN: 0027-8424, 1091-6490, (2023; https://pnas.org/doi/full/10.1073/pnas.1121304109) (June 12, 2012).

13. D. Y. Tsao, M. S. Livingstone, *Annual Review of Neuroscience* **31**, 411–437, ISSN: 0147-006X, 1545-4126 (July 1, 2008).

14. M. Khosla, N. A. R. Murty, N. Kanwisher, *Journal of Vision* **22**, 4184–4184 (2022).

15. E. J. Allen *et al.*, *Nature Neuroscience*, Bandiera_abtest: a Cg_type: Nature Research Journals Primary_atype: Research publisher: Nature Publishing Group Subject_term: Cortex;Neural encoding;Object vision;Perception Subject_term_id: cortex;neural-encoding;object-vision;perception, 1–11, ISSN: 1546-1726 (Dec. 16, 2021).

16. N. Kriegeskorte *et al.*, *Neuron* **60**, 1126–1141, ISSN: 0896-6273 (Dec. 26, 2008).

17. S. Bracci, H. Op De Beeck, *The Journal of Neuroscience* **36**, 432–444, ISSN: 0270-6474, 1529-2401 (Jan. 13, 2016).

18. E. Freud, J. C. Culham, D. C. Plaut, M. Behrmann, *eLife* **6**, e27576, ISSN: 2050-084X (Oct. 5, 2017).

19. E. Freud, M. Behrmann, *Cortex* **129**, 423–435, ISSN: 00109452 (Aug. 2020).

20. P. Papale *et al.*, *Scientific Reports* **9**, 7601, ISSN: 2045-2322 (May 20, 2019).

21. P. Papale *et al.*, *Journal of Neurophysiology* **124**, 1560–1570, ISSN: 0022-3077, 1522-1598 (Dec. 1, 2020).

22. C.-C. Hung, E. T. Carlson, C. E. Connor, *Neuron* **74**, 1099–1113, ISSN: 0896-6273 (June 21, 2012).

23. M. Stylianou-Korsnes, M. Reiner, S. J. Magnussen, M. W. Feldman, *Brain Structure and Function* **214**, 355–359, ISSN: 1863-2653, 1863-2661 (May 2010).

24. C. S. Konen, S. Kastner, *Nature Neuroscience* **11**, 224–231, ISSN: 10976256 (Feb. 2008).

25. M. Sablé-Meyer, K. Ellis, J. Tenenbaum, S. Dehaene, *Cognitive Psychology* **139**, 101527, ISSN: 0010-0285 (Dec. 1, 2022).

26. M. Amalric, S. Dehaene, *Proceedings of the National Academy of Sciences*, 201603205, ISSN: 0027-8424, 1091-6490 (Apr. 11, 2016).

27. M. Amalric, S. Dehaene, *NeuroImage* **189**, 19–31 (2019).

28. A. Agrawal, K. Hari, S. Arun, *Psychological Science* **30**, 1707–1723 (2019).

29. A. Agrawal, K. Hari, S. Arun, *eLife* **9**, e54846, ISSN: 2050-084X (May 5, 2020).

30. J. De Leeuw, P. Mair, *Journal of statistical software* **31**, 1–30 (2009).

31. A. Walther *et al.*, *NeuroImage* **137**, 188–200, ISSN: 10538119 (Aug. 2016).

32. P. Pinheiro-Chagas, A. Daitch, J. Parvizi, S. Dehaene, *Journal of cognitive neuroscience* **30**, 1757–1772 (2018).

33. N. Kanwisher, G. Yovel, *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 2109–2128 (2006).

34. R. Epstein, A. Harris, D. Stanley, N. Kanwisher, *Neuron* (1999).

35. M. A. Williams *et al.*, *Nat Neurosci* **11**, 1439–1445, ISSN: 1546-1726 (Electronic) (2008).

36. M. Amalric, I. Denghien, S. Dehaene, *Developmental Cognitive Neuroscience* **30**, 314–323, ISSN: 1878-9293 (Apr. 1, 2018).

37. S. Kanjlia, C. Lane, L. Feigenson, M. Bedny, *Proceedings of the National Academy of Sciences*, 201524982, ISSN: 0027-8424, 1091-6490 (Sept. 16, 2016).

38. Y. Xu *et al.*, *PLOS Biology* **21**, e3001930, ISSN: 1545-7885 (July 25, 2023).

39. C. Conwell, J. S. Prince, G. A. Alvarez, T. Konkle, presented at the SVRHM 2021 Workshop@ NeurIPS.

40. J. S. Bowers *et al.*, *Behavioral and Brain Sciences*, 1–74, ISSN: 0140-525X, 1469-1825 (Dec. 1, 2022).

41. G. Jacob, R. T. Pramod, H. Katti, S. P. Arun, *Nature Communications* **12**, 1872, ISSN: 2041-1723 (Mar. 25, 2021).

42. D. Heinke, P. Wachman, W. van Zoest, E. C. Leek, *Vision Research* **189**, 81–92, ISSN: 0042-6989 (Dec. 1, 2021).

43. J. A. Thompson, H. Sheahan, C. Summerfield, presented at the Proceedings of The 1st Gaze Meets ML workshop, ed. by I. Lourentzou *et al.*, vol. 210, pp. 199–218, (https://proceedings.mlr.press/v210/thompson23a.html).

44. I. Biederman, *Psychological review* **94**, 115 (1987).

45. J. Feldman, *Current Directions in Psychological Science (Wiley-Blackwell)* **12**, 227–232, ISSN: 09637214 (Dec. 2003).

46. J. Kubilius *et al.*, "CORnet: Modeling the Neural Mechanisms of Core Object Recognition", (2020; http://biorxiv.org/lookup/doi/10.1101/408385).

47. K. He, X. Zhang, S. Ren, J. Sun, presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

48. G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, *arXiv:1608.06993 [cs]*, (2020; http://arxiv.org/abs/1608.06993) (Jan. 28, 2018).

49. M. Schrimpf *et al.*, "Brain-Score: Which Artificial Neural Network for Object Recognition Is Most Brain-Like?", (2020; http://biorxiv.org/lookup/doi/10.1101/407007).

50. M. Zhan, R. Goebel, B. de Gelder, *eneuro* **5**, ENEURO.0285–17.2017, ISSN: 2373-2822 (Jan. 2018).

51. D. Lundqvist, A. Flykt, A. Öhman, *Cognition and Emotion* (1998).

52. Q. Cai, M. Brysbaert, *PloS one* **5**, e10729 (2010).

53. F. Vinckier *et al.*, *Neuron* **55**, 143–156, ISSN: 08966273 (July 2007).

54. G. Dehaene-Lambertz, K. Monzalvo, S. Dehaene, *PLOS Biology* **16**, ed. by K. Grill-Spector, e2004103, ISSN: 1545-7885 (Mar. 6, 2018).

55. S. Dehaene, V. Izard, P. Pica, E. Spelke, *Science* **311**, 381–384 (2006).

56. O. Esteban *et al.*, *Software* (2018).

57. O. Esteban *et al.*, *Nature Methods* (2018).

58. K. Gorgolewski *et al.*, *Frontiers in Neuroinformatics* **5**, 13 (2011).

59. K. J. Gorgolewski *et al.*, *Software* (2018).

60. N. J. Tustison *et al.*, *IEEE Transactions on Medical Imaging* **29**, 1310–1320, ISSN: 0278-0062 (2010).

61. B. Avants, C. Epstein, M. Grossman, J. Gee, *Medical Image Analysis* **12**, 26–41, ISSN: 1361-8415 (2008).

62. Y. Zhang, M. Brady, S. Smith, *IEEE Transactions on Medical Imaging* **20**, 45–57, ISSN: 0278-0062 (2001).

63. A. M. Dale, B. Fischl, M. I. Sereno, *NeuroImage* **9**, 179–194, ISSN: 1053-8119 (1999).

64. A. Klein *et al.*, *PLOS Computational Biology* **13**, e1005350, ISSN: 1553-7358 (2017).

65. A. Evans, A. Janke, D. Collins, S. Baillet, *NeuroImage* **62**, 911–922 (2012).

66. V. Fonov, A. Evans, R. McKinstry, C. Almli, D. Collins, *NeuroImage* **47, Supplement 1**, S102 (2009).

67. D. N. Greve, B. Fischl, *NeuroImage* **48**, 63–72, ISSN: 1095-9572 (2009).

68. M. Jenkinson, P. Bannister, M. Brady, S. Smith, *NeuroImage* **17**, 825–841, ISSN: 1053-8119 (2002).

69. R. W. Cox, J. S. Hyde, *NMR in Biomedicine* **10**, 171–178 (1997).

70. J. D. Power *et al.*, *NeuroImage* **84**, 320–341, ISSN: 1053-8119 (2014).

71. Y. Behzadi, K. Restom, J. Liau, T. T. Liu, *NeuroImage* **37**, 90–101, ISSN: 1053-8119 (2007).

72. T. D. Satterthwaite *et al.*, *NeuroImage* **64**, 240–256, ISSN: 10538119 (2013).

73. C. Lanczos, *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis* **1**, 76–85, ISSN: 0887-459X (1964).

74. A. Abraham *et al.*, *Frontiers in Neuroinformatics* **8**, ISSN: 1662-5196, (https://www.frontiersin.org/articles/10.3389/fninf.2014.00014/full) (2014).

75. J. Diedrichsen *et al.*, *Comparing Representational Geometries Using Whitened Unbiased-Distance-Matrix Similarity*, Aug. 5, 2021, (2022; http://arxiv.org/abs/2007.02789).

76. F. Al Roumi, S. Planton, L. Wang, S. Dehaene, *eLife* **in press**, 2022.10.15.512361, (2022; https://www.biorxiv.org/content/10.1101/2022.10.15.512361v1) (2023).

77. L. Benjamin, M. Sable-Meyer, A. Flo, G. Dehaene-Lambertz, F. Al Roumi, *Journal of Neuroscience* (2024).

78. M. Jas *et al.*, *Frontiers in Neuroscience* **12**, 530, ISSN: 1662-453X (Aug. 6, 2018).

79. G. Niso *et al.*, *Scientific Data* **5**, 180110, ISSN: 2052-4463 (Dec. 2018).

80. C. R. Pernet *et al.*, *Scientific Data* **6**, 103, ISSN: 2052-4463 (Dec. 2019).

81. S. Taulu, M. Kajola, *Journal of Applied Physics* **97**, 124905 (2005).

82. M. A. Uusitalo, R. J. Ilmoniemi, *Medical and biological engineering and computing* **35**, 135–140 (1997).

83. M. Jas, D. A. Engemann, Y. Bekhti, F. Raimondo, A. Gramfort, *NeuroImage* **159**, 417–429, ISSN: 10538119 (Oct. 2017).

84. M. A. Jatoi, N. Kamel, A. S. Malik, I. Faye, *Australasian physical & engineering sciences in medicine* **37**, 713–721 (2014).

85. R. D. Pascual-Marqui *et al.*, *BioRxiv*, 269753 (2018).

86. R. M. Cichy, A. Oliva, *Neuron* **0**, ISSN: 0896-6273, (2020; https://www.cell.com/neuron/abstract/S0896-6273(20)30518-3) (July 27, 2020).

87. M. Sablé-Meyer, K. Ellis, J. Tenenbaum, S. Dehaene, *Cognitive Psychology* **139**, 101527, ISSN: 0010-0285 (Dec. 1, 2022).

Appendix

A. Materials and Methods

A.1. Ethics

All studies were conducted in accordance with the Declaration of Helsinki and French bioethics laws. On-line collection of behavioral data was approved by the Paris-Saclay University Committee for Ethical Research (CER CER-Paris-Saclay-2019-063). Behavioral and brain-imaging studies in adults and 6-year-old children (MEG and fMRI) were approved by the CEA ethics boards and by a nationally approved ethics committee (MEG: CPP_100049; fMRI in children: CPP 100027 and CPP 100053; in adults CPP 100050)

A.2. Derivation of the model representational dissimilarity matrices (RDMs)

To derive the RDMs used to model our data, we proceeded as follows. For the CNN encoding model, we downloaded from GitHub the weights for several neural networks (CORnet (*46*), ResNet (*47*) and DenseNet (*48*) ; all high scoring on brain-score (*49*)), all pre-trained with ImageNet and not specifically trained for our task. We extracted the activation vectors in each hidden layer associated to each shape with the 6x6=36 different orientations and scaling used in the experiment. For each shape, we separated their 36 exemplars randomly into two group to have independent estimation of the representation vectors from the network, and used the cross-validated Mahalanobis distance between these two splits to estimate the distance between each pair of shapes. This provides us with an RDM that captures how dissimilar shapes are according to their internal representations in a CNN of object recognition. Unless specified otherwise (see supplementary text), we report correlation with CORnet layer IT following (*7*).

For the geometric feature model, we estimate a feature vector of geometric features for each shape. This feature vector includes information about (i) right angles (one for each angle, 4 features); (ii) angle equality (one for each pair of angles, 6 features); (iii) side length equality (one for each pair of sides, 6 features); and (iv) side parallelism (one for each pair of sides, 6 features); all of this was done up to a tolerance level (e.g. an angle slightly off a right angle still counts as a right angle), using the tolerance value of 12,5% which was fitted previously to independent behavioral data (*7*). The dissimilarity between each pair of shapes was the difference between the number of features that each shape possesses: because both the square and the rectangle share many features, they are similar. Two very irregular shapes also end up similar as well. Conversely, the square and an irregular shape end up very dissimilar.

A.3. On-line Behavioral Study

Participants: 330 participants (142 Females, 177 Males, 11 Others; mean age 51.1 years, SD=16.8) were recruited via a link provided on a New York Time article news available at https://www.nytimes.com/2022/03/22/science/geometry-math-brain-primates.html. The article reported previous research from the lab and featured a link to our new online experiment. When participants clicked the link, they landed on a page with our usual procedure for online experiments, including informed consent and demographic questions. No personally identifying information was collected.

Materials and Stimuli: The goal of the experiment was to obtain a 11x11 matrix of similarities between the 11 geometric shapes shown in **Fig.1A**. To this aim, on each trial, participants were shown a 3x3 square

grid of shapes, eight of which were copies of the same shape up to rotation and scaling, and one of which was a different shape. Participants were asked to detect the intruder shape by clicking on it. Auditory feedback was provided in the form of ascending or descending going tunes, as well as coloring of the shapes (the intruder was always colored in green indicating what the right answer was, and in case of an erroneous choice, the chosen shape was colored in red indicating a wrong answer). Shapes were presented in pure white on pure black. Each shape was differently scaled and rotated by sampling nine values without replacement in the following scaling factors [0.85, 0.88, 0.92, 0.95, 0.98, 1.02, 1.05, 1.08, 1.12, 1.15] and rotation angles [-25°, -19.4°, -13.8°, -8.3°, -2.7°, 2.7°, 8.3°, 13.8°, 19.4°, 25°]. Note that the rotation angles were centered on 0° but excluded this value so that the sides of the shapes were never strictly vertical or horizontal. Participants did 110 trials (11x10), one for each pair of different reference and intruder shapes. The order of trials was randomized, subject to the constraint that no two identical reference shapes were used on consecutive trials, and that the outlier of a trial was always different from the reference shape of the previous trial. Two examples of trials are shown in **Fig.1B**.

Procedure: The experimental procedure started with instructions, followed by a series of questions: device used (mouse or touchscreen), country of origin, gender, age, highest degree obtained. Participants then provided subjective self-evaluation assessments, with answers on a Likert scale from 1 to 10, for the following items: current skills in mathematics; current skills in first language. Finally, participants performed the task. The instructions text was the following: "The game is very simple: you will see sets of shapes on your screen. Apart from small rotation and scaling differences, they will be identical, except for one intruder. Your task is to respond as fast and accurately as you can about the location of the intruder by clicking on it. The difficulty will vary, but you always have to answer."

A.4. Functional MRI

Participants

Twenty healthy French adults (9 females; 19-37 years old, mean age = 24.6 years old, SD: 5.2 years old) and 25 French first graders (13 females; all 6 years old) participated in the study. Three children quit the experiment before any task began because they did not like the scanning noise or lying in the confined space. One child completed the localizer task but not the other tasks. One child was missing a single run from the intruder task. All participants had normal hearing, normal or corrected-to-normal vision, and no neurological deficit. All adults and guardians of children provided informed consent, and adult participants were compensated for their participation.

Category Localizer

Task. In three localizer runs, children and adult participants were exposed to eight different images categories, such that single geometric shapes could be compared to matched displays of faces, houses and tools, and rows of three geometric shapes could be compared to matched displays of numbers, French words, and Chinese characters. To maintain attention, participants were asked to keep fixating on a green central fixation dot (radius=8pixels, RGB color=26, 167, 19, always shown on the screen), and to press a button with their right-hand whenever a star symbol was presented. The star spanned roughly the same visual angle as the stimuli from the eight categories, and appeared randomly once in one of the two blocks per category (8 target stars total), between the 3rd to the 6th stimuli within that block. As feedback, a 300 ms 650 Hz beep sound was provided after each button press.

Stimuli. In each miniblock, participants saw a series of 6 grayscale images, one per second, belonging to one of eight different categories: faces, houses, tools, numbers, French words, Chinese characters, single geometric shapes, and rows of three geometric shapes. Each category comprised 20 exemplars. All faces,

16 houses, and 18 tools had been used in previous localizer experiments (*50*). For face stimuli, front-view neutral faces (20 identities, 10 males) were selected from the Karolinska Directed Emotional Faces database (*51*). The stimuli were aligned by the eyes and the iris distances. A circular mask was applied to exclude the hair and clothing below the neck. House and tool stimuli were royalty-free images obtained from the internet. House stimuli were photos of 2 to 3-story residence houses. Tool stimuli were photos of daily hand-held tools: half of the images were horizontally flipped, so that there were 10 images in a position graspable for the left and right hand respectively. For French word stimuli, 3-letter French words were selected which were known to first graders and had high occurrence frequencies (range=7-2146 occurrences per million, mean=302, SD=505, based on Lexique, http://www.lexique.org/). Chinese characters were selected from the school textbook of Chinese first graders. Chinese word frequency (range=11-1945 occurrences per million, mean=326, SD=451 (*52*)) was not significantly different from French words used here (t(38)=0.2, p=0.87). Single-digit formula stimuli were 3-character simple operations in the form of "x+y" or "x-y" with x greater than y, x ranging from 2 to 5, and y from 1 to 4. Single shapes consisted in a single, centered outline of a geometrical shape (diamond, hexagon, rhombus, parallelogram, rectangle, square, trapezoid, isosceles triangle, equilateral triangle and right triangle), and were matched in luminance, contrast, and visual angle to the faces/houses/tools/words/Chinese characters which also displayed single objects. A row of shapes consisted of three different shapes side by side, whose total width, size, and line width were matched with 3-letter French words and 3-character single-digit operations. To match the appearance of the monospaced font in (*53*), the monospaced font Consolas was used for the French words and numbers, with identical font weight 900. The font for Chinese characters was Heiti, which looks similar to Consolas. Random font size (uniform in 35-55px font size) were repeatedly sampled until text stimuli achieved similar variability as with the other categories.

The stimuli were embedded in a gray circle (RGB color=157, 157, 157, radius=155 pixels), on the screen with a black background. Within the gray circle, the mean luminance and contrast of the 8 stimuli categories did not differ significantly (luminance: F(7,152)=0.6, p=0.749; contrast: F(7,152)=1.2, p=0.317), see **Fig.S2**

Block structure. The eight categories were presented in distinct blocks of 6s each, fully randomized for block presentation order, with the restriction that there were no consecutive blocks from the same category. Each miniblock comprised 6 stimuli, in random order. Each stimulus was presented for 1s, with no interval in between (*54*). The inter-block interval duration was jittered (4, 6, or 8s; mean = 6s). Each of the eight-block types appeared twice within each run. A 6s fixation period was included at both the beginning and the end of the run. Each run lasted for 3min 24s, and participants performed three such runs during the fMRI session.

Geometry Task

Task. We adapted the geometric intruder detection task (*7*, *55*) to the fMRI scanner. On each trial, participants (children and adults) saw an array of 6 shapes around fixation (3 on the right, and 3 on the left; see **Fig.3A**). Five shapes were identical except for a small amount of random rotation and scaling, while one was a deviant shape. Because the pointing task used in (*7*) was not possible in the limited space of the fMRI scanner, participants were merely asked to click a button with their left or right hand, thereby indicating on which side they thought the deviant was. Note that participants responded on every trial, the side of the correct response was counterbalanced within each shape, and we verified that the average motor response side was unconfounded with geometric shape or complexity. After each answer, auditory feedback was provided with a tone of high, increasing pitch when the answer was correct, and a low-pitch tone otherwise.

Stimuli. Geometric shapes were generated following the procedure described in (*7*). To fit the experiment within the time constraints of children fMRI, a subset of shapes was used, comprising square, rect-

angle, isosceles trapezoid, rhombus, right hinge and irregular shapes to span the range of complexity found in (*5*). Following (*7*), deviants were generated by displacing the bottom right corner by a constant distance in four possible positions (see **Fig.3A**). That distance was a fraction of the average distance between all pairs of points, which was standardized across shapes (45% change). On each trial, six gray-on-black shapes were shown (shape color rgb values: 127,127,127). Shapes were displayed along two semicircles: the positions were determined by positioning the three leftmost (resp. rightmost) shapes on the left side (resp. right side) of a circle of radius 120px, at angles 0, pi/2 and pi, and then shifting them 100px to the left (resp. right). The rotation and scaling of each shape were randomized so that no two shapes had the same scaling or rotation factor, and values were sampled in [0.875, 0.925, 0.975, 1.025, 1.075, 1.125] for scaling and [-25°, -15°, -5°, 5°, 15°, 25°] for rotations, avoiding 0° to prevent alignment of specific shapes with screen borders. One of the shapes was an outlier, whose position was sampled uniformly in all six possible positions such that no two consecutive trials featured outliers in the same position. Outliers were sampled uniformly from the four possible types of outliers, so that all outlier types occurred as often, but no two consecutive trials featured identical outlier types.

Procedure. The six shapes were presented in miniblocks, in randomized order, with no two consecutive blocks with the same type of shape. Each block comprised 5 consecutive trials with an identical base shape, each with 2s of stimulus presentation and 2s of fixation. There was a 4s, 6s or 8s delay between blocks. A central green fixation cross was always on display, and it turned bold 600 ms before a block would start. Each run of the outlier detection task lasted 3m40s.

MRI Acquisition Parameters MRI acquisition was performed on a 3T scanner (Siemens, Tim Trio), equipped with a 64-channel head coil. Exactly 113 functional scans covering the whole brain were acquired for each localizer run, as well as on 179 functional scans covering the whole brain for each run of the geometry task. All functional scans were using a T2*-weighted gradient echo-planar imaging sequence (69 interleaved slices, TR = 1.81 s, TE = 30.4 ms, voxel size = 2x2x2mm, multiband factor = 3, flip angle = 71 degrees, phase encoding direction: posterior to anterior). To reconstruct accurate anatomical details, a 3D T1-weighted structural image was also acquired (TR = 2.30 s, TE = 2.98 ms, voxel size = 1x1x1mm, flip angle = 9 degrees). To estimate distortions, two spin-echo field maps with opposite phase encoding directions were acquired: one volume in the anterior-to-posterior direction (AP) and one volume in the other direction (PA). Each fMRI session lasted for around 50min for children including (by order) 3 runs of a task not discussed here, 3 Category localizer runs, T1 collection, and 2 Geometry runs. For adults, the session lasted for around 1h 20min because they took the same runs as for children as well as an additional harder version of the geometry task. This version, which involved smaller deviant distances and a stimulus presentation duration of only 200 ms, turned out to be too difficult. While the overall performance still shows a correlation with complexity (r^2=.63, p<.02), it was entirely driven by two shapes, the square and the rectangle: other shapes were equally hard, and although they were better than chance they did not correlate with complexity (r^2=.35,p=.22) while the correlations remained for the simpler condition.

Data analysis. Preprocessing was performed with the standard pipeline fMRIPrep. Results included in this manuscript come from preprocessing performed using fMRIPrep 20.0.5 (*56*, *57*), which is based on Nipype 1.4.2 (*58*, *59*), and generated the following detailed method description.

Anatomical Data. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (*60*), distributed with ANTs 2.2.0 (*61*), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (*62*). Brain surfaces were reconstructed using recon-all (*63*), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and

FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (*64*). Volume-based spatial normalization to two standard spaces (MNI152NLin6Asym, MNI152Nlin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. the following templates were selected for spatial normalization: FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model (*65*) and ICBM 152 Nonlinear Asymmetrical template version 2009c (*66*).

Functional Data. For each of the 10 BOLD EPI runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Susceptibility distortion correction (SDC) was omitted. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (*67*). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (*68*). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207 (*69*). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD time-series were resampled into several standard spaces, correspondingly generating the following spatially-normalized, preprocessed BOLD runs: MNI152Nlin6Asym, MNI152Nlin2009cAsym. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by (*70*)). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*71*). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask was obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (*72*). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (*73*). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.6.2 (*74*), mostly within the functional processing workflow.

fMRI GLM Models. fMRI first-level models (general linear models or GLM) were computed by convolving the experimental design matrix (specific to each task, see below) with SPM's HRF model as implemented in Nilearn, with the following parameters: spatial smoothing using a full width at half maximum window (fwhm) of 4mm; a second-order autoregressive noise model; and signal standardized to percent signal change relative to whole-brain mean. The following confound regressors were added: polynomial drift models from constant to 5th order (6 regressors); estimated head translation and rotation on three axes (6 regressors) as well as the following confound regressors given by fmriprep: average cerebro-spinal fluid signal (1 regressor), average white matter signal (1 regressor), and the first five high-variance confounds estimates (*71*) (5 regressors).

In the visual category localizer, the design matrix contained distinct events for each visual stimulus, grouped by category, each with a duration of 1s, including a specific one for the target star, leading to 9 regressors (Chinese, face, house, tools, numbers, words, single shape, three shapes, and star). In the geometry task runs, each reference shape was associated to a regressor with trial duration set to 2s, thus leading to 6 regressors (square, rectangle, rhombus, isotrapezoid, hinge and random). In both cases, button presses were not modeled as they were fully correlated with predictors of interest (either the star for the localizer, or every single trial for the geometry task).

Second-level models were estimated after additional smoothing with a full width at half maximum window of 8mm. Statistical significance of clusters was estimated with a bootstrap procedure as follows: given an uncorrected p-value of .001, clusters were identified by contiguity of voxels that had p-values below this threshold. Then, the p-value of a cluster was derived by comparing its statistical mass (the sum of its t-values) to the distribution of the maximum statistical mass obtained by performing the same contrast after randomly swapping the sign of each participant's statistical map. We performed this swapping 10,000 times for each contrast we estimated, and computed the corrected p-value accordingly; for instance, a cluster whose mass was only outperformed by 3 random swaps out of the 10,000 was assigned a p-value of .0003.

Searchlight RSA analyses. First, we estimated the Representational Dissimilarity Matrix (RDM) within spheres centered on each voxel. For this we performed a searchlight sweep across the whole brain. We extracted the GLM coefficients of the geometric shapes from each voxel and all the neighboring voxels in a 3-voxel radius (=6mm), for a total of 93 voxels per sphere. We discarded voxels where more than 50% of this sphere fell outside the participant's brain. We extracted the betas of each shape and used a cross-validated Mahalanobis distance across runs (crossnobis, implemented in rsatoolbox) to estimate the dissimilarity between each pair of shapes. We attributed this distance to the center of the searchlight, thereby estimating an empirical RDM at each location.

Then we compared this empirical RDM with the two RDMs derived from our two models separately, using a whitened correlation metric. Both choices of metrics (crossnobis and whitened correlation) follow the recommendations from a previous methodological publication (*75*).

Finally, we computed group-level statistics by smoothing the resulting correlation map (fwhm=8mm) and performing statistical maps, cluster identification, and statistical inference at the cluster level as we did for the second-order level analysis, but with a one-tailed comparison only as we did not consider negative correlations.

When we report bar-plot for ROIs, we performed subject-specific voxel localization within ROIs. Within each ROI identified we find, for each subject, the 10% most responsive subject-specific voxels in the same contrast used to identify the cluster. To avoid double-dipping, we selected these voxels using the contrast from one run, then collected the fMRI responses (beta coefficients) from the other runs; we perform this procedure across all runs and average the responses. Error bars indicate the standard error of the mean across participants.

A.5. Magnetoencephalography (MEG)

Participants. Twenty healthy French adults (13 females; 21-42 years old, mean: 24.9 years old, SD: 8.1 years old) participated in the MEG study. All participants had normal hearing, normal or corrected-to-normal vision, and no neurological deficit. All adults provided informed consent and were compensated for their participation. For all but one participant, we had access to anatomical recordings in 3T MRI, either from prior, unrelated experiments in the lab, or because the MEG session was immediately followed by a recording. Because of one participant missing an anatomical recording, analyses that required source reconstruction were performed on nineteen subjects.

Task. During MEG, adult participants were merely exposed to shapes while maintaining fixation and attention. As in previous work (e.g. *76*, *77*), the goal was to examine the spontaneous encoding of stimuli and the presence or absence of a novelty response to occasional deviants.

Stimuli. All 11 geometric shapes in **Fig.1A** were presented in miniblocks of 30 shapes. Geometric shapes were presented centered on the screen, one shape every second, with shapes remaining onscreen for 800ms and a centered fixation cross present between shapes for 200ms. To make the shapes more attractive (and since the same shape were also used in an infant experiment, not reported here), during their 800ms presentation, the shapes slowly increased in size: in total, a scale factor of 1.2 was applied over the course of 800ms, with linear interpolation of the shape size during the duration of the presentation. Shapes were presented in miniblock, following an oddball paradigm: within a miniblock, all shapes were identical up to scaling (randomly sampled in [0.875, 0.925, 0.975, 1.025, 1.075, 1.125]) and rotation (sampled in [-25°, -15°, -5°, 5°, 15°, 25°]), except for occasional oddballs which were deviant version of the reference shape. Each miniblock comprised 30 shapes, 4 of which were oddballs that could replace any shape after the first six. Two oddballs never appeared in a row. There was no specific interval between miniblocks beyond the usual duration between shapes. A run was made of 11 miniblocks, one per shape in random order, and participants attended a 8 runs except two participants who are missing a single run due to experimenters' mistakes when setting up the MEG acquisition.

Procedure. After inclusion by the lab's recruiting team, participants were prepared for the MEG with electrocardiogram (ECG) and electrooculogram (EOG) sensors, as well as four Head Position Indicator (HPI) coils, which were digitalized to track the head position throughout the experiment. Then we explained participants that the task was a replication of an experiment with infants, and therefore was purely passive: they would be shown shapes and were instructed to pay attention to each shape, while trying to avoid blinks as well as body and eye movements. They sat in the MEG and we checked the head position, ECG/EOG and MEG signal. From that point onward, we never opened the MEG door again to avoid resetting MEG signals and allow for optimal cross-run decoding and generalization. Participants took typically 8 runs consecutively, with small breaks with no stimuli between runs to rest their eyes. At the end of the experiment, participants took the intruder test from (*7*) on a laptop computer outside of the MEG, and finally we spent some time debriefing with participants the goal of the experiment.

To ensure high accuracy of the timing in the MEG, each trial's first frame contained a white square on the bottom of the screen, which was hidden from participants but recorded with a photodiode. The same area was black during the rest of the experiment. The ramping up of the photodiode was therefore synchronized with the screen update and the appearance of the stimulus, ensuring robust timing for analyses. Then each "screen update" event was linked to a recorded list of presented shapes.

Preprocessing. The preprocessing of the data was performed using `MNE-BIDS-Pipeline`, a streamlined implementation of the core ideas presented in (*78*) and leveraging BIDS specifications (*79*, *80*). The pipeline performed automatic bad channel detection (both noisy and flat), then applied Maxwell filtering and Signal Space Separation on the raw data (*81*). The data was then filtered between .1 Hz and 40Hz, and

resampled to 250 Hz. Extraction of epochs was performed for each shape, starting 150ms before stimulus onset and stopping 1150ms after, and the relevant metadata (event type, run, trial index, etc.) for each epoch was recovered from the stimulation procedure at this step. Artifacts in the data (e.g. blinks, and heartbeats) were repaired with signal-space projection (*82*), and thresholds derived with "autoreject global" (*83*). For source reconstruction, some preprocessing steps were performed by fmriprep (see below). Then, sources were positioned using the "oct5" spacing with 1026 sources per hemisphere, and we used the e(xact)LORETA method (following recommendations from (*84*, *85*)) using empty-room recordings performed right before or right after the experiment to estimate the noise covariance matrix. Additionally, for source reconstruction anatomical MRI were preprocessed with fmriprep. T1-weighted (T1w) images were corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (*60*), distributed with ANTs 2.3.3 (*61*). The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow. Brain tissue segmentation of cerebrospinal fluid, white-matter and gray-matter was performed on the brain-extracted T1w using fast (*62*). Brain surfaces were reconstructed using recon-all (*63*) and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (*64*).

Statistical Analyses

Decoding. After epoching the data, for each timepoint within an epoch and each participant, we trained a logistic regression decoder to classify epochs as reference or oddball using samples from all shapes. For this analysis we discarded the 6 first trials at the beginning of each block, since (i) those could not be oddballs ever and (ii) there was no warning of transitions between blocks and so the first trials were also "oddballs" with respect to the previous block's shape. Each epoch was normalized before training the classifier. To avoid decoders being biased by the overall signal's autocorrelation across timescales, we used six folds of cross validation over runs with the following folds: even runs versus odd runs; first half versus second half; and runs 1, 2, 5, 6 versus 3, 4, 7 and 8. Folds were used in both directions, e.g. even for training and odd for testing; as well as odd for training and even for testing. The decoders were trained on data from all of the shapes conjointly. When testing its accuracy, we tested it separately on data from each shape (e.g. detecting oddballs within squares only, within rectangles only, etc.) –using runs independent from the training data. In order to estimate accuracy without being biased by the imbalanced number of epochs in the different classes (there are 4 oddballs for every 20 references), we report the ROC Area Under the Curve of the Receiver Operating Characteristic (ROC AUC) for each shape in **Fig.4A**, left subfigure. Then at each timepoint, we correlated the decoding performance with the number of geometric features present in each shape: the r correlation coefficient at each timepoint is plotted in the central subfigure, together with shading for a significant cluster identified with permutation tests across participants (implemented by mne, one tailed, 2^{13} permutations, cluster forming threshold <.05). Finally, we average the decoding performance in the identified cluster and plot each shape's average decoding performance against online behavioral data: this is effectively visualizing the same data as the central column's figure, and therefore no statistical test is reported. The analyses were performed both without any smoothing, and with a uniform averaging over a sliding window of 100ms; the results are identical but we chose the latter as plots using the smoothed version make the separation of the different shapes easier to see. The same holds true for the next analysis as well.

In **Fig.4B**, we display a similar sequence of plots, but now instead of training a single classifier to identify epochs as reference or oddball conjointly on all shapes, we train eleven separate such classifier, one for each shape. This produces very similar results from the previous analysis.

ERP. Different participants can in principle end up with very different decoder weights, since the decoders are trained independently for each participant. Subsequent analyses are based on the decoders' performance only, and therefore the decoder's weights associated to each sensor are not considered.

To stay closer to the MEG data, we replicated the previous analysis directly from evoked potential data in the gradiometers. Using spatiotemporal permutation testing, we identified a set of sensors ant timepoints across participants where the reference and the oddball epochs were significantly different. We identified three significant clusters: [268, 844]ms, [440, 896]ms and [492, 896]ms.

Then we computed the average difference between the reference and the oddball trials across these sensors separately for each shape. Finally, we correlated the differences with the number of geometric features in the reference shape. The first cluster, from 268 ms to 844 ms, did not elicit any significant correlation with the geometric regularity; however, the two others yielded significant clusters at the $p<.05$ level, although with later timing than the decoding analysis, at around ~600ms.

RSA analysis. For RSA analyses, data from the oddballs was discarded and we used data from the magnetometers only. The goal of this analysis was to pinpoint when and where the mental representation of shapes followed distances that matched either a neural network model or a geometric feature model. We used the same model RDMs as the one we used to analyze behavior and fMRI data, and provide below the details of how we derived the empirical RDMs for our analyses.

In sensor space. We estimated, at each timepoint, the dissimilarity between each pair of shape across sensors: we relied on rsatoolbox to compute the Mahalanobis distance, crossvalidated with a leave-one-out scheme over runs. This provided us with one empirical RDMs for each timepoint, with no spatial information as this was performed across all sensors. Then we compared this RDM with our two models: since our model RDMs are effectively orthogonal, we performed the comparisons with the empirical RDM separately. We used a whitened correlation metric to compare RDMs. This gave use one timeseries for each participant and each model, and we then performed permutation testing in the [0, 800]ms window to identify significant temporal clusters for each model separately. As shown in **Fig.4C**, this yields one significant cluster associated to the CNN encoding model, and two significant clusters associated to the geometric feature model. As in the decoding analysis, we performed the analyses with both no temporal smoothing, and with sliding averaging (savgol filter, 2nd order, 100ms window); the results are virtually identical but we chose to report the smoothed ones in Fig.4C.

In source space. In order to understand not only the temporal dynamic of the mental representations, but also get an estimate of the localization of the various representations, we turned to RSA analysis in source space. We performed source reconstruction of each reference shape trial using eLORETA and empty-room recordings either just before or just after the scanning session to estimate the noise covariance. Then we averaged the data over the three identified temporal clusters from the sensor-space RSA analysis: [0, 332]ms and [136, 452]ms. We then performed RSA analysis at each source's location using its neighboring sources (geodesic distance of 2cm on the reconstructed cortical surface). Finally, we compared the resulting RDMs to either the CNN encoding model or the geometric feature model. These steps were performed independently for each subject. Next, we projected the whitened correlation distance between empirical RDMs and model RDM onto a common cortical space, fsaverage. Finally, we performed a permutation spatial cluster test across participants, using adjacency matrices between sources (sources are adjacent if they were immediate neighbors on the reconstructed surface's mesh). This resulted in two significant clusters associated to the CNN encoding model during the [0, 332]ms time window, located in bilateral occipital areas. Additionaly, this resulted in two significant clusters associated to the geometric feature model during the [136, 452]ms, located in very broad bilateral networks encompassing dorsal and frontal areas.

Joint MEG-fMRI RSA. Representational similarity analysis also offers a way to directly compared similarity matrices measured in MEG and fMRI, thus allowing for fusion of those two modalities and tentatively assigning a "time stamp" to distinct MRI clusters (*86*). However, we did not attempt such an analysis here for several reasons. First, distinct tasks and block structures were used in MEG and fMRI. Second,

a smaller list of shapes was used in fMRI, as imposed by the slower modality of acquisition. Third, our study was designed as an attempt to sort out between two models of geometric shape recognition. We therefore focused all analyses on this goal, which could not have been achieved by direct MEG-fMRI fusion, but required correlation with independently obtained model predictions.

B. Supplementary text

B.1. Behavioral intruder task and generation of behavioral similarity matrix

Additional Results in the Behavior task: comparison with other CNNs

In **Fig.S1** we report additional results from correlating human behavior with various models (top; DenseNet and ResNet) and various layers of a single model (bottom; many layers of CORnet). In all cases the CNN encoding model is a much worse predictor of the human behavior than the geometric feature model (all p<.001). The late layers of CORnet, DenseNet and ResNet all capture some of the variance of participants behaviors to comparable; while ResNet is the highest scoring in this analysis, we can see in **Fig.S1A** that they are all highly correlated, and in order to stay close to previous work we used CORnet's layer IT throughout the article.

Additionally, the fit of the successive layers of CORnet (V1, V2, V4: small effect sizes, only significant for V1; IT, flatten: much higher effect sizes, increasing between IT and flatten) indicates that the feed-forward processing of the visual input by the CNN encoding yields internal representations that are increasingly closer to humans'. This suggest that even the visual model goes beyond local properties that V1 would capture – but in all cases, the fit is much worse than the geometric feature model.

B.2. Intruder task during fMRI

Intruder task: Overall, both age groups performed better than chance (all p<.001) at detecting the intruder, with an average response time of 1.03s in adults, 1.36s in children, for the easy condition, and 0.79s in adults in the hard condition. Both error rates and response times are modulated by the reference shape (all p <.001), and both are correlated with error rate data from outside the scanner (all p<.05).

B.3. Intruder task in MEG Participants

Participants recruited for the MEG studied performed no behavioral task inside the MEG, as we relied on a passive presentation oddball paradigm. After the scanner session, participants took one run of the intruder detection task previously used online and described in (*7*) – we presented them with the task after the scanning session to avoid biasing participants toward actively looking for intruders in the oddball paradigm. At the group level, the data fully replicated the geometric regularity effect (*7*) (correlation of error rate with that of previous participants, regression over 11 shapes: r^2=.90, p<.001). A mixed effect model correlating the error rates of the two groups, with both the slope and the intercept as separate random effects, yielded an intercept not significantly different from 0 (p=.42), and a slope significantly different from 0 (p<1e-10) and not significantly different from 1 (p=.19) suggesting that the data was similar in the two groups. We also correlated each participant's average error rate per shape to the group data from our previously dataset (*7*). A one-tailed test for a positive slope indicated that 19 out of 20 participants displayed a significant geometric regularity effect. We still included the data from the

participant that did not display a significant effect, as we had not decided on such a rejection criterion beforehand.

B.4. fMRI contrasts for geometric shape in the visual localizer

In order to isolate the brain responses to geometric shapes, we focused on the simplest possible contrast, i.e. greater activation to the presentation of a single geometric shape versus all of its single-image controls (face, house, tool). This contrast is presented in Figure 2C. Note that we excluded Chinese characters from this comparison because they often include geometric features (e.g. parallel lines), but including them gave virtually identical results (compare **Fig.S3A** and **Fig.S3B**, identical list of cluster-level corrected clusters at the p<.05 level in both age groups). We also included in the design rows of three distinct geometric shapes (e.g. square, triangle, circle). Our logic here was that this condition, although somewhat artificial from the geometric viewpoint, could be very tightly matched with two other conditions, namely a string of three letters ("words" condition, e.g. BON) or a small 3-symbol mathematical operation ("numbers" condition, e.g. 3+1). However, the corresponding contrast (3 geometric shapes > words and numbers) did not give any positive activation: no positive cluster was significant at the whole brain cluster-level corrected p<.05 level in either age group; additionally, none of the ROIs identified in the single shape contrast was significant for this contrast: aIPS left, p=.975 for adults and p=.389 for children; aIPS right, p=.09 in both adults and children; pITG right, p=.13 in adults and p=.362 in children). We reasoned, however, that numbers should be excluded from this contrast because, by our very hypothesis, geometric shapes should activate a number-based program-like representations (e.g. square = repeat(4){line, turn}) (*87*). When restricting the contrast to "3 geometric shapes > words", at the whole brain level, no cluster reached significance at the p<.05 level (cluster-level correction). However, in this case, testing the ROIs identified in the single shape condition revealed that while the left aIPS still did not reach significance at the p<.05 level (p=.71 in adults, p=.30 in children), both the right aIPS and the right pITG reached significance (aIPS right: p<.001 in adults and p=.014 in chidren; pITG, p=.008 in adults and p=.046 in children).

B.5. Additional ROI analysis of the ventral pathway in fMRI (Fig.S4)

We used individually defined regions of interest (ROIs) to probe the fMRI response to geometric shapes in four cortical regions defined by their preferential responses to four other visual categories known to be selectively processed in the ventral visual pathway: faces, houses, tools and words. To this aim, we first identified, at the group level, clusters of voxels associated to each visual category. Such clusters were identified using a non-parametric permutation test across participants at the whole-brain level using a contrast for the target category against the mean of the other three (voxel p<.001 then clusters p<.05 except for ffa in children, p=.09, and the absence of a well-identified vwfa in children). Significant clusters which intersect the MNI coordinate z=-14 are shown in **Fig.S4**; in the case of the fusiform face area (FFA) and the visual word form area (VWFA), we restricted ourselves to clusters at MNI coordinates typically found in the literature, respectively (-45, -57, -14) and (40, -55, -14).

Then, within each such ROI identified in adults, we identified for each subject, including children, the 10% most responsive subject-specific voxels in the same contrast used to identify the cluster. To avoid double-dipping, we selected the voxels using the contrast from a single run, then collected the fMRI responses (beta coefficients) to all categories from the other runs, and then replicated this procedure across all runs while averaging the responses to a given category. The average coefficients within each such individually-defined cortical ROI are shown in **Fig.S4**, separately for children and adults. Several observations are in order, and detailed below for each visual category.

In the left-hemispheric VWFA, we can see that voxels are indeed responsive to written words in the participants' language (French), more than to an unknown language (Chinese), in both adults and children (paired t-tests, p<.001 in adults, p=0.003 in children). VWFA voxels also responded to the symbolic display of numbers entering into small computations (e.g. 3+1) in adults, but this response did not appear to be developed yet in children. VWFA voxels also showed a response to tools, particularly in young children, as already been reported (*54*). In the opposite direction, they were particularly under-activated by houses in adults. Finally and most crucially, geometric shapes, whether presented alone or in a string of 3, did not elicit a strong response, indeed no stronger than non-preferred categories such as Chinese characters or faces (p=.37 in adults, p=.057 in children on a one-tailed paired t-test).

In the right-hemispheric FFA, we only saw a purely selective response to faces in both adults and children. All the other visual categories yielded an equally low level of activity in this area. In particular, the responses to geometric shapes were not significantly different from those to other visual categories.

In the bilateral ROIs responsive to tools, a very similar result was found: apart from their strong response to tools and their slightly increased activation to houses and reduced activations to written words and numbers in the right hemisphere, these voxels elicit activations that were essentially equal across all non-preferred visual categories, with geometric shapes being no exception.

Finally, bilateral house-responsive ROIs corresponding to the parahippocampal place area (PPA) were also mildly responsive to tools in both populations and both hemispheres. However, geometric shapes did not activate these clusters more than other visual categories such as faces (both hemispheres, both age groups have p>.05 in one-tailed paired t-tests).

C. Tables

C.1. Table S1. Coordinates and characteristics of significant fMRI clusters responding to geometric shapes in localizer runs

Age Group	X	Y	Z	Peak t-value	Volume in cm^2	Cluster corrected p-value
Adults	51.5	-54.5	-8.5	5.12	4.4	p<.01
	35.5	-48.5	55.5	4.93	7.2	p<.01
	39.5	-76.5	-12.5	-6.33	51.4	p<.01
	-34.5	-70.5	-10.5	-5.92	42	p<.01
Children	-12.5	-76.5	-48.5	5.51	3.9	p=.01
	55.5	-24.5	45.5	5	8.3	p<.01
	21.5	-80.5	-46.5	4.9	4.4	p=.01
	-42.5	-38.5	35.5	4.9	6.4	p<.01
	21.5	-94.5	-8.5	-6.34	40.1	p<.01
	-24.5	-98.5	-8.5	-6.17	55.3	p<.01

For each age group, each line gives the peak coordinates, volume and statistics of a cluster with p<.05 for the contrast "single shape > other single visual categories". The sign of the peak t-value and the shading indicate whether the contrast was positive (white background) or negative (grey background). Coordinates are given in MNI space.

C.2. Table S2. Coordinates and characteristics of significant fMRI clusters in the RSA analysis

Age group	Model	X	Y	Z	Peak t-value	Volume in cm^2	Cluster corrected p-value
Adults	Geometric Features	-0.5	13.5	49.5	5.4	64.4	p<.01
		-22.5	-54.5	51.5	5.38	13.6	p<.01
		-14.5	-66.5	5.5	5.28	6.1	p<.01
		31.5	31.5	-8.5	4.86	2.2	p=.03
		-26.5	-2.5	49.5	4.79	5.3	p<.01
		-8.5	-72.5	-38.5	4.62	7	p<.01
		-34.5	23.5	5.5	4.15	3.2	p<.01
		-10.5	71.5	3.5	4.03	1.9	p=.03
		-46.5	-0.5	33.5	3.88	1.6	p=.04
		21.5	-98.5	-6.5	3.72	2.3	p=.02
	CNN Encoding	23.5	-14.5	57.5	5.4	2.4	p=.03
		-48.5	-82.5	-0.5	4.96	4.4	p<.01
		-30.5	-80.5	25.5	4.55	2.9	p=.02
		1.5	21.5	45.5	4.54	3.8	p<.01
		27.5	33.5	5.5	4.51	3.7	p<.01
		45.5	-80.5	-2.5	4.38	2.2	p=.03
		53.5	13.5	31.5	4.05	3.9	p<.01
Children	CNN Encoding	-22.5	-84.5	11.5	4.59	1.4	p=.06
		43.5	-82.5	11.5	4.29	2.2	p=.02

Same organization as Table S1.

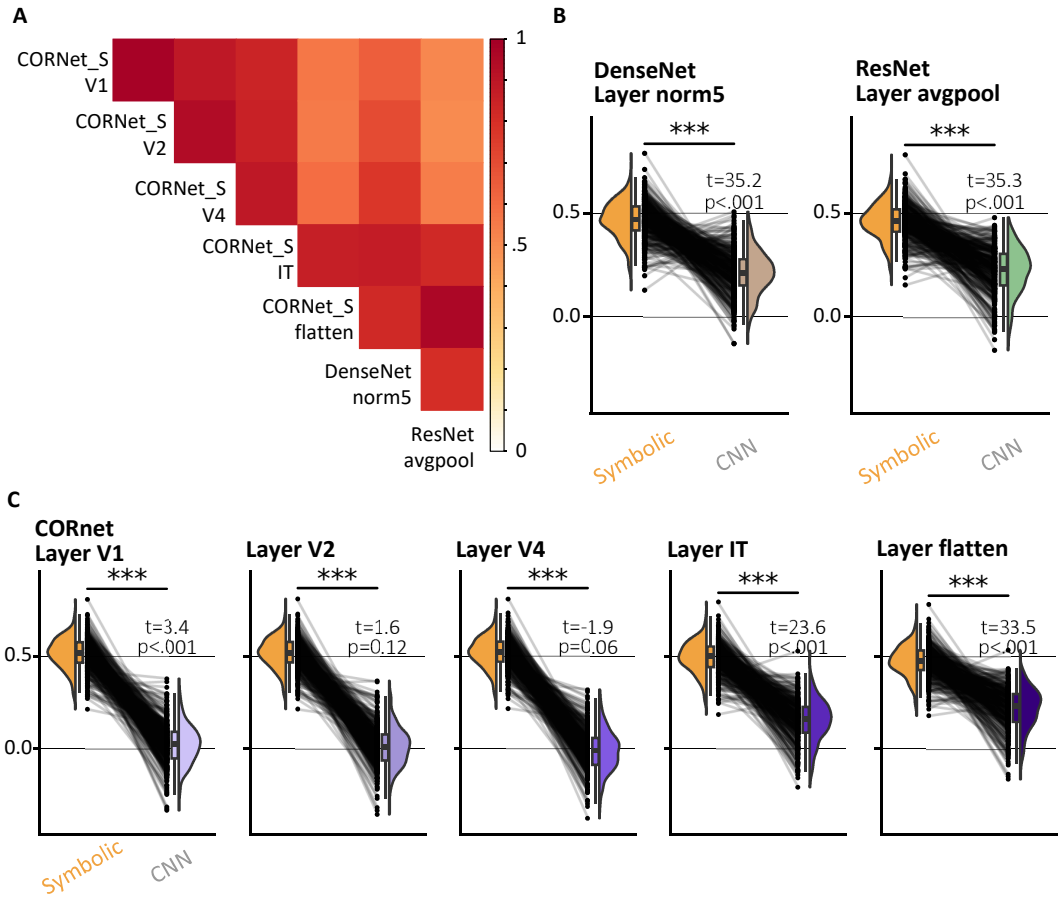

Fig.S1 Additional CNN encoding models of human behavior. **A**, Correlation matrix between all pairs of RDMs generated by each CNN and layer considered.
B, Replication of **Fig.1D** with different CNNs. Stars indicate a significant difference between the geometric feature model and the respective CNN encoding model (p<0.001). t and p values also indicate whether the CNN encoding model is a significant predictor of participant's behavior (the geometric feature model is always highly significant). **C** Replication of **B** using different layers of CORnet, organized from early to late layers, from left to right. Note that the late layers are much more significant predictors of human behavior than the early ones – although still far inferior to the geometric feature model.

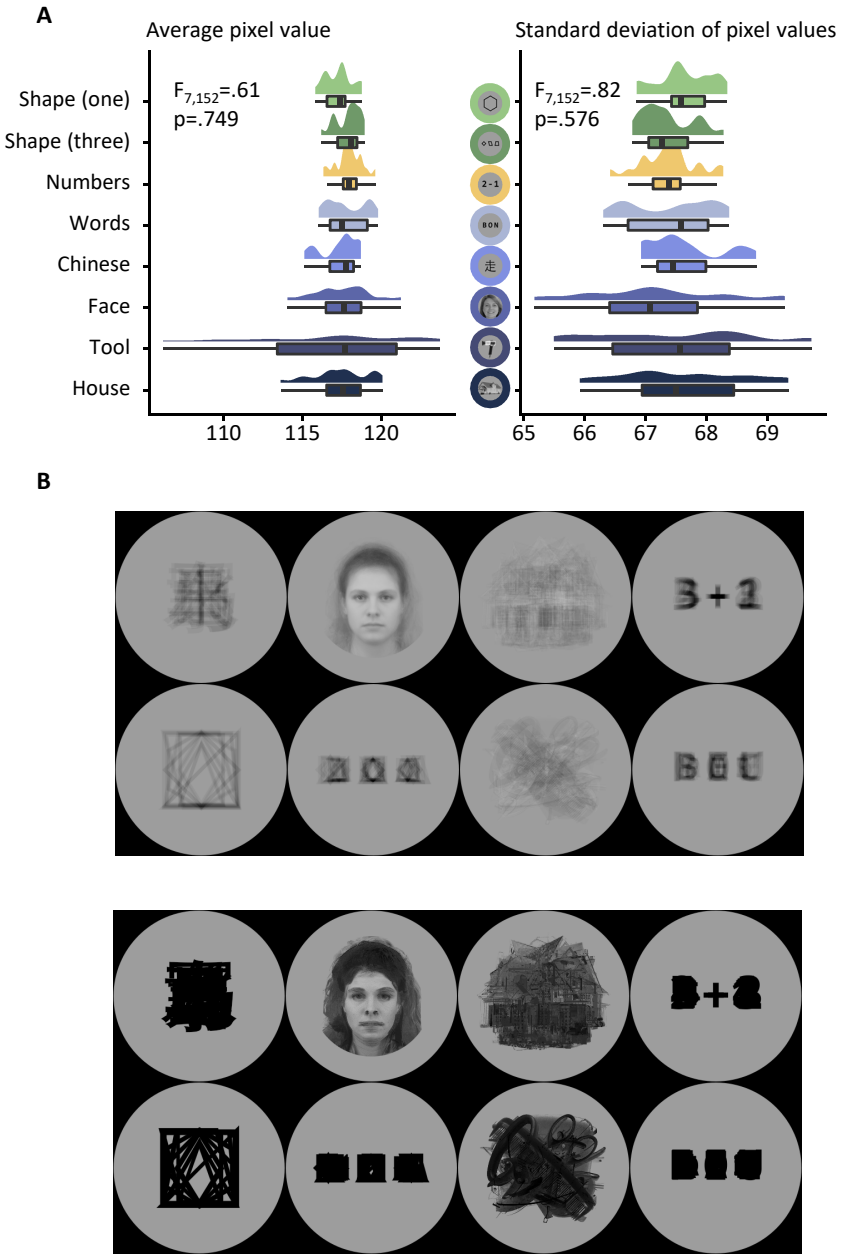

Fig.S2 Overview of the stimuli used for the category localizer. A, Average pixel value (left) and average standard deviation across pixels (right) for stimuli within each category (y axis). An ANOVA indicated no significant effect of the stimulus category on either the average or the standard deviation across pixels. B, Average (top) and max (bottom) pixel value at each location across the eight possible visual categories used in the localizer.

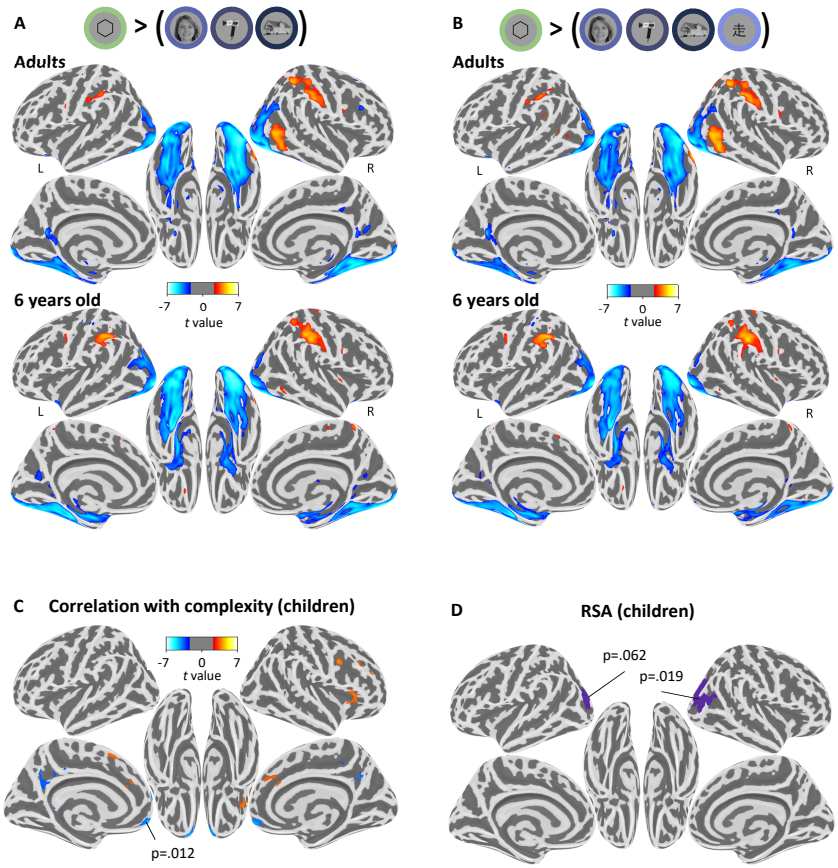

Fig.S3 Details of the fMRI results in children (complement to Fig.2 and Fig.3 in the main text) A: Statistical map associated with the contrast "single geometric shape > faces, houses and tools", projected on an inflated brain (top: adults; bottom: children; for illustration purpose we display the uncorrected statistical map at the p<.01 level). Notice how similar the activations are in both age groups. **B:** Same as A, but for the contrast "single geometric shape > all single-object visual categories (face, house, tools, Chinese characters)". The activation maps are very similar to the previous contrast, and very similar across age groups. **C**: Whole brain correlation of the BOLD signal with geometric regularity in children, as measured by the error rate in a previous online intruder detection task (*7*). Positive correlations are shown in red and negative ones in blue. Voxel threshold p<.001, no correction for multiple comparisons, but the p-value indicates the only cluster that was significant at the cluster-level corrected p<.05 threshold. **D**, Results of RSA analysis in children. No cluster was significant at the p<.05 level for the geometric feature models; one right-lateralized occipital cluster reached significance for the CNN encoding model (cluster-level corrected p=.019), and its symmetrical counterpart was close to the significance threshold (cluster-level corrected p=.062)

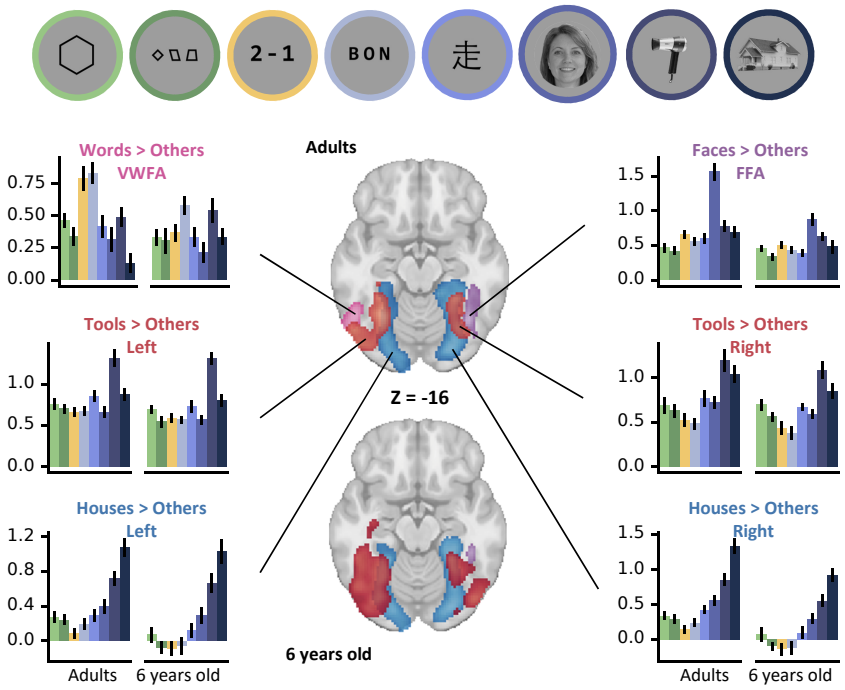

Fig.S4 fMRI response of subject-specific voxels in the ventral visual pathway to geometric shapes and other visual stimuli The brain slices show the group-level clusters associated to various contrasts known to elicit a selective response in the ventral visual pathway, in both age groups: VWFA (words > others; green), FFA (faces > others; purple), tool-selective ROIs (tools > others; red) and PPA (houses > others; light blue). Within each ROI, plots show the mean beta coefficients for the BOLD effect within a subject-specific selection of the 10% best voxels, using independent runs for selection and plotting to avoid circularity and "double-dipping".