A geometric shape regularity effect in the human brain

Mathias Sablé-Meyer[*1,2,3], Lucas Benjamin[2], Cassandra Potier Watkins[1], Chenxi He[2], Maxence Pajot[2], Théo Morfoisse[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

Abstract

The perception and production of regular geometric shapes is a characteristic trait of human cultures since prehistory, whose neural mechanisms are unknown. Behavioral studies suggest that humans are attuned to discrete regularities such as symmetries and parallelism, and rely on their combinations to encode regular geometric shapes in a compressed form. To identify the relevant brain systems and their dynamics, we collected functional MRI and magnetoencephalography data in both adults and six-year-olds during the perception of simple shapes such as hexagons, triangles and quadrilaterals. The results revealed that geometric shapes, relative to other visual categories, induce a hypoactivation of ventral visual areas and an overactivation of the intraparietal and inferior temporal regions also involved in mathematical processing, whose activation is modulated by geometric regularity. While convolutional neural networks captured the early visual activity evoked by geometric shapes, they failed to account for subsequent dorsal parietal and prefrontal signals, which could only be captured by discrete geometric features or by more advanced transformer models of vision. We propose that the perception of abstract geometric regularities engages an additional symbolic mode of visual perception.

Introduction

> 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, the very first detectable graphic productions of prehistoric humans were highly regular non-pictorial geometric signs such as parallel lines, zig-zags, triangular or checkered patterns (*1*, *2*). Human cultures throughout the world compose complex figures using simple geometrical regularities such as parallelism and symmetry in their drawings, decorative arts, tools, buildings, graphics and maps (*3*). Cognitive anthropological studies suggest that, even in the absence of formal western education, humans possess intuitions of foundational geometric concepts such as points and lines and how they combine to form regular shapes (*4*, *5*). The scarce data available to date suggests that other primates, including chimpanzees, may not share the same ability to perceive and produce regular geometric shapes (*6*–*10*), though unintentional-but-regular mark-marking behavior have been reported in macaques (*11*). Thus, studying the brain mechanisms that support the perception of geometric regularities may shed light on the origins of human compositionality and, ultimately, the mental language of mathematics. Here, we provide a first approach through the recording of functional MRI and magneto-encephalography signals evoked by simple classical geometric shapes such as triangles or squares. Our goal is to probe whether, over and above the pathways for processing the shapes of images such as faces, places or objects, the regularities of geometric shapes evoke additional activity.

The present brain-imaging research capitalizes on a recent series of studies of how humans perceive quadrilaterals (*8*). In this study, we created 11 tightly matched stimuli which were all simple, non-figurative, textureless four-sided shapes, yet varied in their geometric regularity. The most regular was the square, with four parallel sides of equal length and four identical right-angles. By progressively removing some of these features (parallelism, right-angles,

equality of length, equality of angles), we created a hierarchy of quadrilaterals ranging from highly regular to completely irregular (Fig. 1A). In a variety of tasks, geometric regularity had a large effect on human behavior. For instance, for equal objective amounts of deviation, human adults and children detected a deviant shape more easily among shapes of high regularity, such as squares or rectangles (<5% errors), than among irregular quadrilaterals (>40% errors). The effect appeared as a human universal, present in preschoolers, first-graders, and adults without access to formal western math education (the Himba from Namibia). However, when baboons were trained to perform the same task, they showed no such geometric regularity effect. Baboon behavior was 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 (*12*).

We therefore formulated the hypothesis that, in the domain of geometry, humans deploy an additional cognitive process, specifically attuned to geometric regularities. On top of the circuits for object recognition, which are largely homologous in human and non-human primates (*13–15*), the human code for geometric shapes would involve a distinct "language of thought", an encoding of discrete mathematical regularities and their combinations (*7*, *8*, *12*, *16–19*).

This hypothesis predicts that the most elementary geometric shapes, such as a square, are not solely processed within the ventral and dorsal visual pathways, but may also evoke a later stage of geometrical feature encoding in brain areas that were previously shown to encode arithmetic, geometric and other mathematical properties, i.e. the bilateral intraparietal, inferotemporal and dorsal prefrontal areas (*20*, *21*). We hypothesized that (1) such cognitive processes encode shapes according to their discrete geometric properties including parallelism, right angles, equal lengths and equal angles; (2) the brain compress this information when those properties are more regularly organized, and thus exhibit activity proportional to minimal description length (*7*, *22*, *23*); and (3) these computations occur downstream of other visual processes, since they rely on the initial output of visual processing pathways.

Here, we assessed these spatiotemporal predictions using two complementary neuroimaging techniques (functional MRI and magnetoencephalography). We presented the same 11 quadrilaterals as in our previous research and used representational similarity analysis (*24*) to contrast two models for their cerebral encoding, based either on classical CNN models or on exact geometric features. In the fMRI experiment, we also collected simpler images contrasting the category of geometric shapes to other classical categories such as faces, places or tools. Furthermore, to evaluate how early the brain networks for geometric shape perception arise, we collected those fMRI data in 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). If geometric shape perception involves elementary intuitions of geometric regularity common to all humans, then the corresponding brain networks should be detectable early on.

Experiment 1: Estimating the Representational Similarity of Quadrilaterals with Online Behavior

Our research focuses on the 11 quadrilaterals previous used in our behavioral research (*8*), which are tightly matched in average pairwise distance between vertices and length of the bottom side, yet differing broadly in geometric regularity (**Fig.1A)**. The goal of our first experiment was to obtain a 11x11 matrix of behavioral dissimilarities between the 11 geometric shapes shown in **Fig.1A,** in order to compare it with predictions from classical visual models, embodied by CNNs, as well as geometric feature models of shape perception. To evaluate perceptual similarity in an objective manner, in experiment 1 we assessed the difficulty of visual search for one shape among rotated and scaled versions of the other (**Fig.1B**) (*25*, *26*). Within a grid of 9 shapes, 8 are similar and 1 is different, and participant have to click on it. Intuitively, if two shapes are very dissimilar, we expect both the response time and the error rate of finding one exemplar of one shape amongst exemplars of the other shape to be low. Conversely, we expect both to be high if shapes are similar. This gives us an empirical measure of shape dissimilarity which we can compare to the distance predicted by different models.

Results

The 11x11 dissimilarity matrix, estimated by aggregating response time and errors from n=330 online participants is shown in **Fig.1D**. The distance estimated as the average success rate divided by the average response time; method section for details. To better understand its similarity structure, we performed 2-dimensional ordinal Multi-Dimensional Scaling (MDS) projection (**Fig.1C**) (*27*). The projection of the 11 shapes on the first dimension showed a strong geometric regularity, with the square and the rectangle landing at the far right, rhombus and

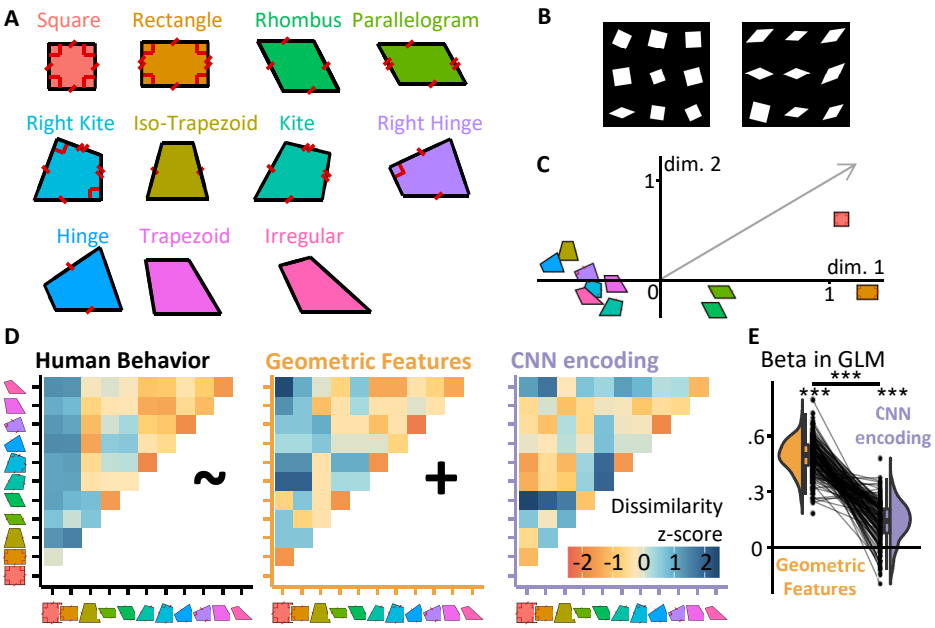

Fig. 1. Measuring and modeling the perceptual similarity of geometric shapes. (**A**) The eleven quadrilaterals used throughout the experiments (colors are consistently used in all other figures). (**B**) Sample displays for the behavioral visual search task used to estimate the 11x11 shape similarity 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; the grey arrow indicates the projection on the MDS space of the number of geometric primitives in a shape. (**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.

parallelogram in the middle, and less regular shapes at the far left. Thus, human perceptual similarity seemed primarily driven by geometric properties. To quantify this resemblance using that 2D MDS projection, we examined 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). This vector (shown in grey in Fig. 1C) had a projection that was significantly different from 0 for both principal axes (both $p<.01$).

Most diagnostically, we compared the full human dissimilarity matrix to those generated by two competing models of shape processing (*8*). The geometric feature model proposes that each shape is encoded by a feature vector of its discrete geometric regularities and predicts dissimilarity by counting the number of features not in common: this makes squares and rectangles very similar, but squares and irregular quadrilaterals very dissimilar. On the other hand, we operationalize our visual model by propagating shapes 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 and other layers of Cornet-S, with unchanged conclusions). Shape similarity was estimated as the crossnobis distance between activation vectors in late layers (*28*). Note that these two models are not significantly correlated ($r^2=.04$, $p>.05$).

Multiple regression of the human dissimilarity matrix with the predictions of those two visual and symbolic models (**Fig.1E**) showed that both contributed to explaining perceived similarity ($ps<.001$), 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 educated humans (*8*). Additional models and comparisons are presented in **Fig. S5**: in particular, we have included two distance measures based on skeletal representations (*29*, *30*), both of which perform better than chance but significantly less than the symbolic model.

Experiment 2: fMRI Geometric Shape Localizer

To understand the neural underpinning of the cognition of geometric shape using fMRI, we started with the simplest foray into geometric shape perception by including geometric shapes as an additional visual category in a standard visual localizer used in the lab (Fig. 2). Across short mini-blocks, this fMRI run 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). In 20 adults in functional MRI (9 females; 19-37 years old, mean age 24.6), we collected three localizer runs using a fast miniblock design. To maintain attention, participants were asked to detect a rare target, which could appear in any miniblock.

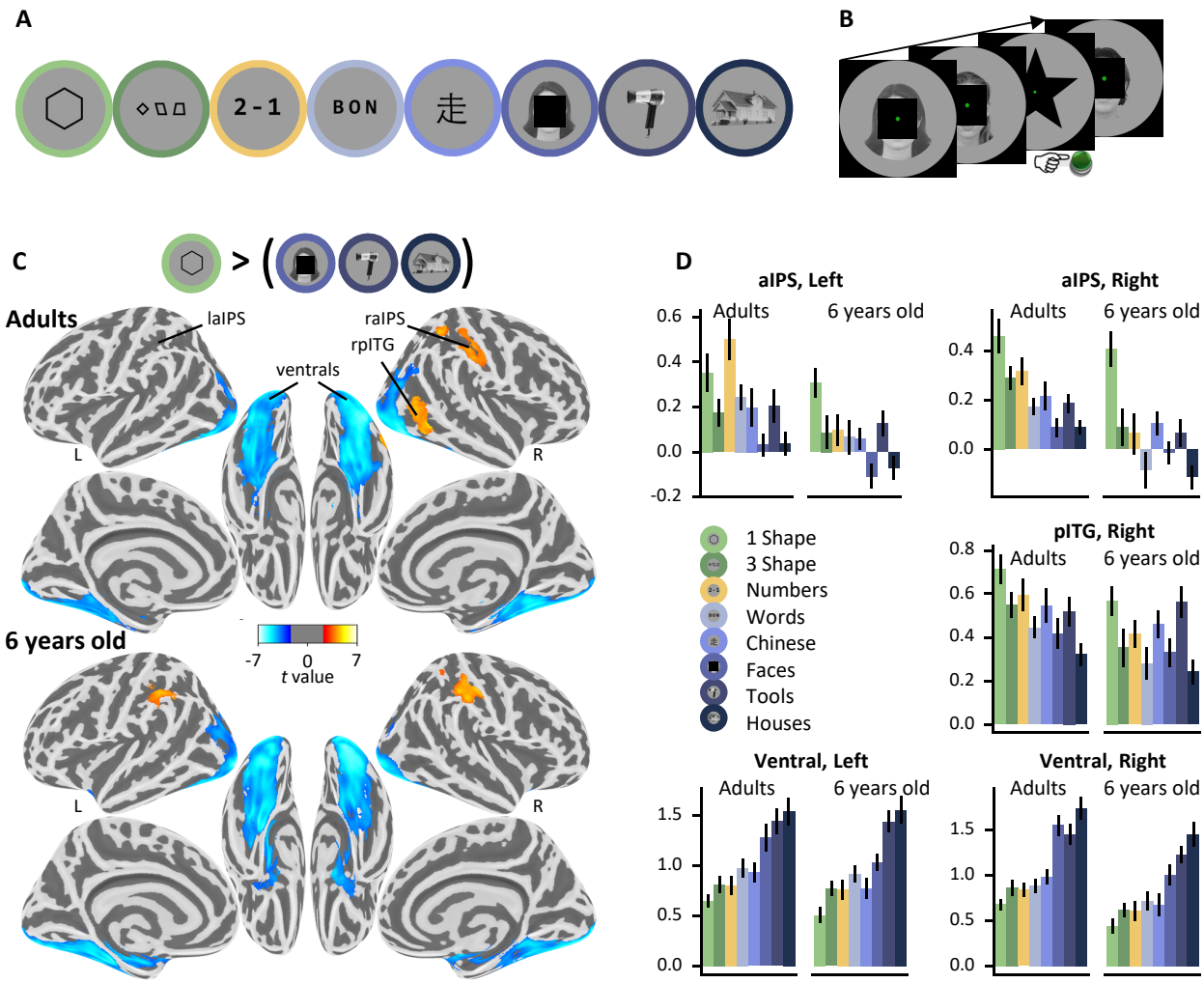

Fig. 2. Localizing the brain systems involved in geometric shape perception (**A**) Visual categories used in the localizer. (**B**) Task: In each miniblock, among a series of 6 pictures from a given category, participants had to detect a rare target star. (**C**) Statistical map associated with the 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.

Results

Reduced activity to geometric shapes in ventral visual cortex. Classical ventral visual category-specific responses, for instance to faces or words, were easily replicated (**Fig.2**; **Fig.S2)**. However, when contrasting geometric shapes to faces, houses or tools, we observed a massive under-activation of bilateral ventral occipito-temporal areas (unless otherwise stated, all statistics are at voxelwise p<.001, clusterwise p<.05 permutation-test corrected for multiple comparisons across the whole brain). All of the regions specialized for words, faces, tools or houses showed this activity reduction when presented with geometric shapes (**Fig.2C; Table S1**).

This group analysis was further supported by subject-specific analyses of ROIs specialized for various categories of images (see Appendix and **Fig.S4**). First, unsurprisingly, the FFA, which is known for its strong category-selectivity, showed the lowest responses to geometric shapes in the fusiform face area. Second, in a subject-specific ROI analysis of the visual word form area (VWFA), identified by its stronger response to alphabetical stimuli than to face, tools and houses, no activity was evoked by single shapes, or strings of three shapes, above the level of other image categories such as objects or faces. This finding eliminates the possibility that geometric shapes might have been processed similarly to letters. Similarly, one could have thought that geometric shapes would be processed together with other complex man-made objects, which are often designed to be regular and symmetrical. However, geometric shapes again yielded a low activation in individual 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 ones presented as Lego blocks (*31*). However, again, geometric shapes actually induced minimal activity in individually defined PPA voxels (see **Fig.S4** for a summary).

Increased activity to geometric shapes in intraparietal and inferior temporal cortices. While the activity of the ventral visual cortex thus seemed to be globally reduced during geometric shape perception, we observed, conversely, a superior activation to geometric shapes than to face, tools and houses in only two significant positive clusters, 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<.001 uncorrected), the symmetrical aIPS in the left hemisphere was also activated (also see SOM for additional results concerning the "3-shapes" condition and with both shape conditions together).

Those areas are similar to those active during number perception, arithmetic, geometric sequences, and the processing of high-level math concepts (*7*, *20*, *21*). To test this idea formally, we used the localizer to identify ROIs activated by numbers more than words. This contrast identified a left IPS cluster (p<.05), while the symmetrically identified cluster in right IPS did not reach significance at the whole brain level (p=.18). In both cases, however, those two ROIs were also significantly more activated by geometric shapes than other visual categories.

The observed overlap with number-related areas of the IPS is compatible with our hypothesis that geometric shapes are encoded as mental expressions that combines number, length, angle and other geometric features. However, it could also reflect an association with math knowledge acquired during education. To test whether the brain activity observed in adults reflects a basic intuition of geometry which is present early on in education, we replicated our fMRI study in 22 6-year-old first graders. When comparing the single shape condition and faces, houses and tools, we observed the same reduction in ventral visual activity. We also observed greater aIPS activity, now significant in both hemispheres. Furthermore, the right pITG voxels extracted from adults also reached significance in children for the shape versus face, houses and tools. Conversely, the left aIPS voxels extracted in children reached significance in adults. Indeed, the activation profiles were quite similar in both age groups (**Fig.2D; see Fig.S3A** and **Fig.S3B** for uncorrected statistical maps of adults and children, which are quite similar). In particular, aIPS activity was strongest to geometric shapes in children, while in adults strong responses to both geometric shapes and numbers were seen, indicating an overlap with previously observed areas involved in arithmetic (*21*, *32*). Our findings in six-year-olds suggest that geometric shapes activate these regions even before they become responsive to symbolic arithmetic (*33*) (p<.05 for "shape > numbers", two-tailed t-test, in each of the bilateral IPS ROIs identified for shapes in children; p>.05 for both ROIs in adults).

In sum, geometric shapes led to reduced activity in ventral visual cortex, where other categories of visual images show strong category-specific activity. Instead, geometric shapes activated math- and number-related areas, particular the right aIPS.

Experiment 3: fMRI of the geometric intruder task

In a second fMRI experiment, we measured the detailed fMRI patterns evoked by our quadrilaterals, with the goal to submit them to a representational similarity analysis and test our hypothesis of a double dissociation between regions

encoding visual (CNN) versus symbolic codes. Adults and children performed an intruder task similar to our previous behavioral study (*8*), with miniblocks allowing us to evaluate the activity pattern evoked by each quadrilateral shape. To render the task doable by 6-year-olds, we tested only 6 quadrilaterals, displayed as two half-circles of three items, and merely asked participants whether the intruder was on the left or the right of the screen (**Fig.3**).

Results

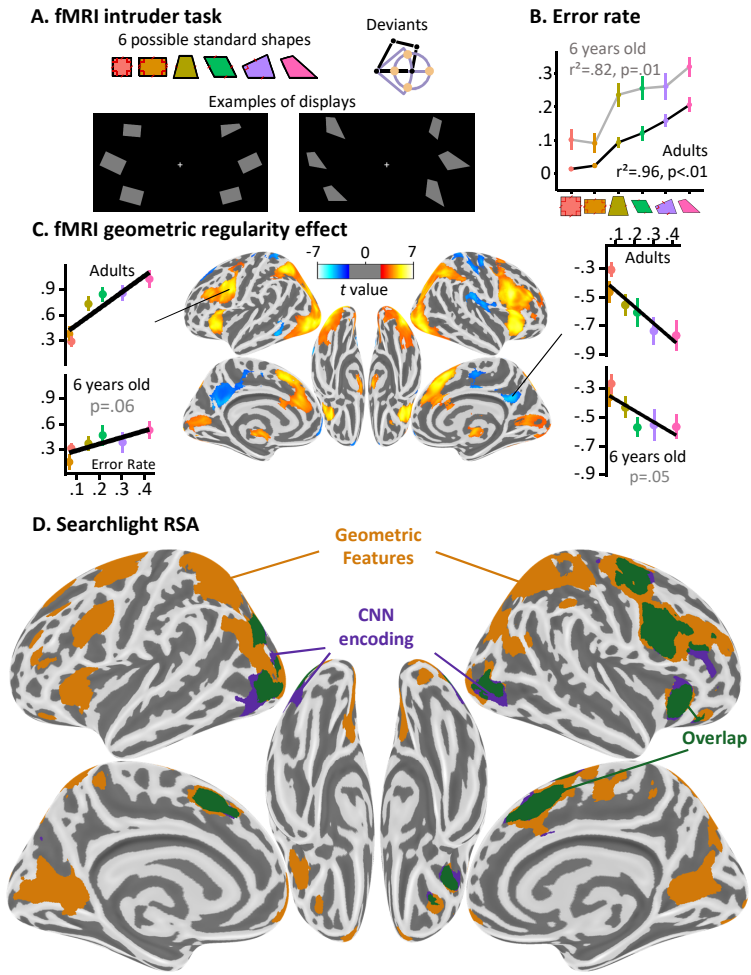

Fig. 3. Dissociating two neural pathways for the perception of geometric shape. (A) fMRI intruder task. Participants indicated the location of a deviant shape via button clicks (left or right). Deviants were generated by moving a corner by a fixed amount in four different directions. **(B)** Performance 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 (*8*). 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.

Behavioral performance inside the scanner replicated the geometric regularity effect in adults and children: performance was best for squares and rectangles and decreased linearly for figures with fewer geometric regularities (**Fig.3B**). Behavior in the fMRI scanner was significantly correlated with an empirical measure of geometric regularity measured with an online intruder detection task (*8*), which is used in all following analyses whenever 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**). This broad network, however, probably encompassed both regions involved in geometric shape encoding and those involved in the decision about the intruder task, whose difficulty increased when geometric regularity decreased. To isolate the regions specifically involved in shape coding, we performed searchlight RSA analyses, which focused 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 (*34*). 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 for decision-making.

In children, possibly because the task difficulty was too high, few results were obtained. The correlation of brain activity with regularity at the whole-brain level yielded a single significant cluster (see Figure S3C for an uncorrected statistical map (voxelwise p<.001), with the single significant whole-brain, cluster corrected significant cluster (p=.012) in the ventromedial prefrontal cortex). When testing the ROIs from the adults in children, after correcting for multiple comparisons across the 14 ROIs, only the posterior cingulate was significant (p=.049), though two clusters were very close to significance, both with p=.06: one positive in the left precentral gyrus (shown in **Fig.3C**), 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). In the searchlight analysis, no cluster associated to the geometric feature model reached significance at the whole-brain level (**Fig.S3D**). However, a right lateral occipital cluster was significantly captured by the CNN encoding model in children (p=.019) and its symmetrical counterpart was close to the significance threshold (p=.062) (**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 previous work (*8*); combined with difficulty of obtaining precise subject-specific activation patterns in young children, this could make the symbolic strategy harder to localize.

Experiment 4: oddball paradigm of geometric shapes in MEG

The temporal resolution of fMRI does not allow to track the dynamic of mental representations over time. Furthermore, the previous fMRI experiment suffered from two limitations: we studied six quadrilaterals only, and we used an explicit intruder detection which blended the activations due to shape identification and to task decision. To overcome those issues, we replicated the experiment in adult magnetoencephalography (MEG) with three important changes; (1) all 11 quadrilaterals were studied; (2) participants were simply asked to fixate and attend to every shape, without performing any explicit task; (3) shapes were presented serially, one at a time, in miniblocks with a fixed quadrilateral shape and with rare intruders with the bottom right corner shifted by a fixed amount (*8*). This design allowed us to study the neural mechanisms of the geometric regularity effect: would the shapes be automatically encoded according to their geometric regularities, even in a passive task? And would the intruders be detected more easily among geometric regular shapes than among irregular ones, as previously found behaviorally in an active task (*8*)?

Results

In spite of the passive design, MEG signals revealed an automatic detection of intruders, 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**). Overall, the decoder performed above chance level, reaching a peak at 428ms, indicating the presence of brain responses specific to intruders. Crucially, although trained on all shapes together, the decoder performed better with geometrically regular shapes than with irregular shapes, indicating that oddball shapes were more easily detected within blocks of regular shapes, as previously found behaviorally (*8*). Indeed, a regression of decoding performance on geometrical regularity yielded a significant spatio-temporal cluster of activity, which first became significant around ~160ms and peaked at 432ms (here and after, temporal clusters are purposefully reported with approximate bounds following (*35*)). A geometrical regularity effect was also seen in the latency of the outlier response (See **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=-1.83, p=.041) indicating that oddballs yielded both a larger and a faster response when the shapes were geometrically more regular. 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 (*8*) and suggest that the computation of a geometric code and its deviations occurs under passive instructions and starts with a latency of about ~150ms.

We next used temporal RSA to further probe the dynamics of the perception of the reference, non-intruder shapes. For each time point, we estimated the 11x11 neural dissimilarity matrix across all pairs of reference shapes using sensor data with crossnobis distances (*28*), 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 (observed cluster extent at approximately 0 to 300ms, p<.001) showed a significant correlation of the CNN encoding model on brain similarity. Its early onset in this analysis, 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, and indeed performing the same analysis with no smoothing yields a very similar significant cluster from 60ms to 320ms. It was followed by a significant cluster associated with the geometric feature model (~150-450ms, p<.001). Thus, those results are compatible with our hypothesis of two distinct stages in geometric shape perception, and suggest that a symbolic encoding arises around ~200ms.

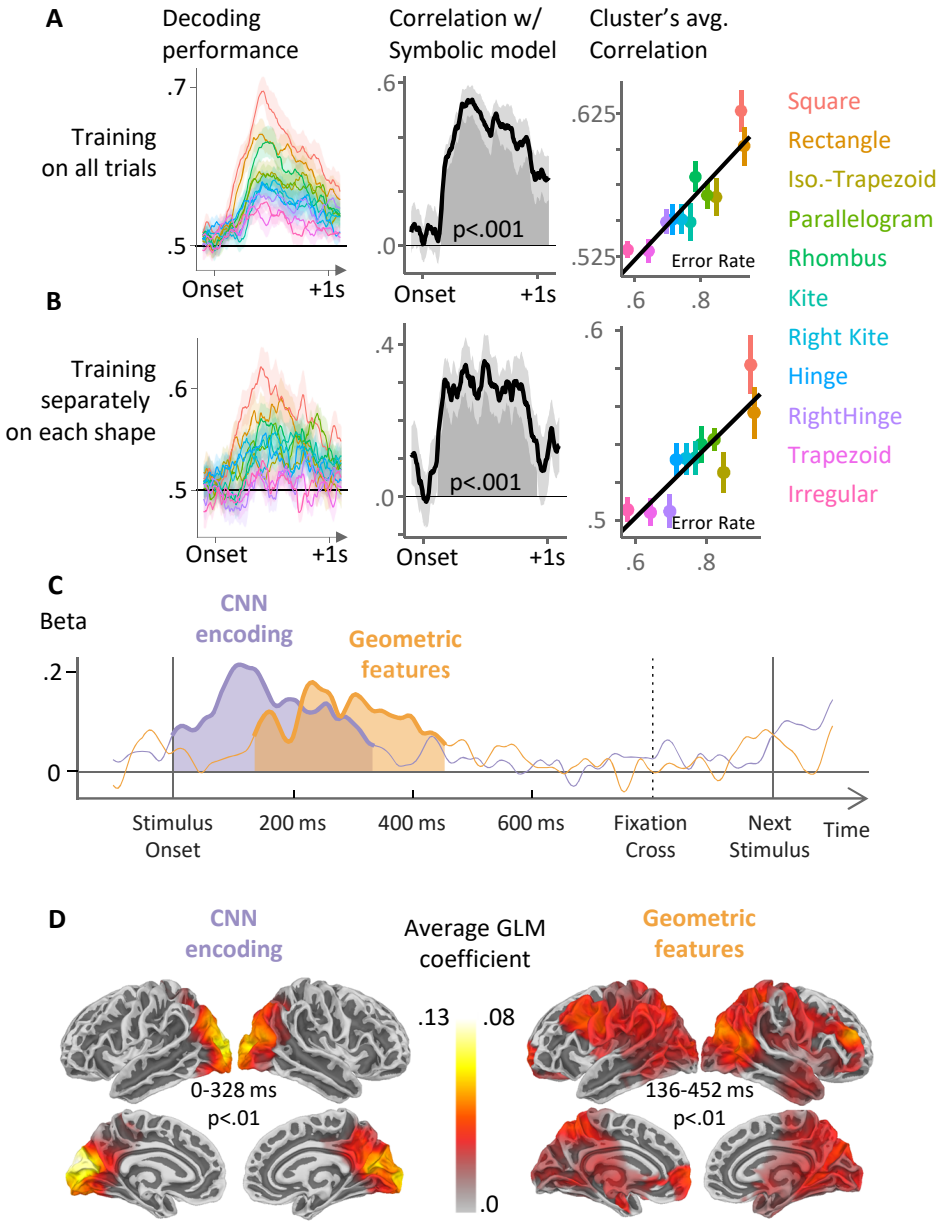

Fig. 4. Using MEG to time the two successive neural codes for geometric shapes (**A,B**) Performance of a classifier using MEG signals to predict whether the stimulus is a regular shape or an oddball. Left: performance for each shape; middle: correlation with geometrical regularity (same x axis as in 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.

To understand which brain areas generated these distinct patterns of activations, and probe whether they fit with our previous 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 a 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.

Modelling of the results with alternative models of visual perception

In the above analyses, we contrasted two models for our data: a CNN versus a list of geometric properties. However, there are many other models of vision. In this section we model our data with two additional classes of models, based on the shape skeleton and principal axis theory, or vision transformers neural networks.

Skeleton models

Skeletal representations (*36*) offer biologically plausible and computationally tractable models of object representation in human cognition. Several methods to derive skeletal representation from object contour have been put forward, including Bayesian estimation that trade off accuracy for conciseness (*37*). More recently, distance estimators between shapes based on their skeletal structure have been used to model human behavior (*29*, *30*, *38*–*40*), and hierarchical object decomposition have been rigorously described and tested (*41*). Remarkably, even when simply asked to tap a geometric shape on a touchscreen anywhere they want, human taps cluster along the shape skeleton (*42*).

To model our data, we needed a distance metric between two shape skeletons. We explored two metrics: one that measures the shortest distance between two skeletons after optimal alignment and rotation (*29*, *43*), and one that measures the difference in angle and length of matched sub-parts of the shapes' skeleton (*30*). Our results are summarized in Fig. S5A; the first observation is that over our 11 quadrilaterals, these two skeletal metrics were not significantly correlated, suggesting that they capture very different properties when applied to minimal geometric shapes and are possibly not best suited to characterize them. Additionally, neither of these metrics predicted the behavioral data better than the CNN models.

When tested on fMRI data, the first method based on optimal alignment and rotation elicited did not elicit any significant cluster in either population. In adults, but not in children, the second method based on the structure of the skeletons significantly correlated with bilateral clusters in the visual cortex, as well as a significant cluster in the right premotor cortex (see Fig.S6). These areas constitute a subset of the areas identified with the CNN encoding model.

When tested on MEG data, the first method similarly did not elicit any significant temporal cluster at the $p<.05$ level. The second method yielded two significant temporal clusters, the first from ~75ms to ~260ms and the second from ~285ms to ~370ms. When looking at the reconstructed sources, the second temporal cluster did not yield and significant spatial cluster, and the first cluster elicited a bilateral cluster encompassing both early occipital area and the early dorsal visual stream, with no significant localization in the frontal cortex.

Overall, for our dataset, these models appear partially similar to the CNN model insofar as they capture subsets of the timings and localizations of the neural data captured by the CNN model, but they do not strongly predict the behavior data as the geometric feature model does, nor capture the broad cortical networks associated with the geometric feature model.

Vision Transformers

While CNN models have well-known limits in capturing several aspects of human perception (*43*, *44*), connectionist models based on the transformer architecture may provide a closer match, especially when it comes to symbolic or mathematical regularities (*45*). To test this possibility, following a similar method as with CNNs, we modeled our empirical RDMs with distance measured between pair of shapes in the late layers of a visual transformer, DINOv2 (*46*). In Fig. S5B, we report the empirical RDMs, the DINO RDM, as well as the CNN RDM, together with a plot of each subject's correlation with the CNN and DINO matrices in a GLM. The DINO predictor yielded a dissimilarity matrix quite similar to the human behavioral one; indeed, the distribution of coefficients for the DINO predictor was significantly greater than 0 (95% confidence interval [0.73, 0.75]; $p<.001$) and significantly different from the CNN

predictor (p<.001). Note that in this GLM, the distribution of CNN predictors is significantly lower than zero (p<.001), which was not the case when contrasted with the symbolic model. This suggests that the DINO predictor may involve a mixture of symbolic-like features and perceptual features. Indeed, its correlation with the symbolic model is .78 and the correlation with the cornet is .56. This may explain why human are best predicted with a mixture of these two models, but with different weights: to achieve the best fit, a model with both DINO and CNN puts a high weight on DINO and then negatively corrects the excess perceptual features by putting a negative weight on the CNN model.

This analysis was confirmed by a fit of the fMRI data (Fig.S6) and MEG data (Fig.S5C): in both cases, the RSA analysis with DINO last layer yielded results that were very similar to a mixture of the CNN and the symbolic model. In fMRI in adults, a distributed set of cortical areas were significantly correlated with the DINO model, and these areas overlapped with the union of the areas predicted by both of our original models. In children, DINOv2 elicited a significant cluster in the right visual cortex, in areas overlapping with the areas activated by the CNN model in children.

In MEG, a large and significant cluster was predicted by the DINO model. Its timing (significant cluster from 0ms to 456ms) overlapped with the spatiotemporal clusters found with both the CNN model and the symbolic model. Source analysis indicated that a correlation with DINO originated from widespread sources that, again, overlapped with both those found with the CNN model (early occipital areas) and the symbolic model (widespread regions included dorsal, anterior temporal, and frontal sources).

Taken together, these results suggest that the last layer of DINO is driven by both early visual and higher-level geometric features, and that both are needed to fit human data. A systematic investigation of the impact of network architecture, dataset size and dataset content is beyond the scope of the present paper, but the present data could serve as a reference dataset with which to understand which of these factors ultimately cause a neural network to display symbolic properties close to those found in the human brain.

Discussion

Our 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*, *8*, *12*). 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 dorso-frontal network sensitive to geometric regularity.

Behavioral signatures of geometric regularity

At the behavioral level, the new large-scale data that we report here (n=330 participants) fully replicated our prior finding that human perception of geometric shape dissimilarity, as evaluated by a visual search task, is best predicted by a model that relies on exact geometric features, rather by a CNN. These exact geometric feature representations can be thought of as a more abstract and compressed representations of shapes: they replace continuous variations along certain dimensions (such as angles or directions) with categorical features (right angle or not, parallel or not). Such a representation, which remains invariant over changes in size and planar rotation, turns a rich visual stimulus into a compressed internal representation.

Even in educated human adults, a small proportion of the variance in behavioral dissimilarity remained predicted by the CNN model, so that both CNN and geometry predictors were significant in a multiple regression of human judgments. Previously, we only found such a significant mixture of predictors in uneducated humans (whether French preschoolers or adults from the Himba community) (*8*). It is likely that the greater power afforded by the present experiment yielded a greater sensitivity. This finding comforts the hypothesis that two strategies for geometric shape perception are available to humans: one based on hierarchical visual processing (captured by the CNN), and the other based on an analysis of discrete geometric features – the latter dominating strongly in educated adults.

fMRI results

Using fMRI, we found that the mere perception of geometric shapes, compared to various other visual categories, yields a reduced bilateral activation of the entire ventral visual pathway, together with a localized increased activation in the anterior IPS and in the posterior ITG. Furthermore, geometric regularity modulated fMRI activation in most of the bilateral occipito-parietal pathway, middle frontal gyrus and anterior insula. Finally, the most diagnostic result

we found is that the representational similarity matrix in these regions was predicted by geometric similarity rather than by the CNN (figure 3).

The IPS areas activated by geometric shapes overlap with those active during the comprehension of elementary as well as advanced mathematical concepts (*7*, *20*, *21*). This finding agrees 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 when they perform mathematics (*47*, *48*) or just evaluate the shapes of manmade objects (*49*). Thus, the representation of shapes computed in these regions is more abstract and amodal than the hierarchy of visual filters that is thought to capture ventral visual shape identity processing.

MEG results

Using MEG, we found that even during the passive perception of simple quadrilateral shapes, participants are sensitive to their geometric regularity: occasional oddball shapes elicited greater violation-of-expectation signals within blocks of regular shapes than within blocks of irregular shapes. Additionally, using RSA and source reconstruction, we observed a spatial and temporal dissociation in the processing of quadrilateral shapes: an early occipital response, well predicted by CNN models of object recognition, was followed by broadly distributed cortical activity that correlated with a model of exact geometric features. Despite the limited accuracy afforded by source reconstruction in MEG, there was considerable overlap between our RSA analysis in fMRI and in MEG.

This early response of occipital areas, followed by a later activation of a broad dorso-frontal network, may seem at odd with some results showing that the dorsal pathway responds faster than the ventral pathway (*50*, *51*). However, in this work, the dorsal network that we probed does not solely compute elementary features of location or motor affordance, but high-level geometrical and mathematical information of an abstract categorical nature. It seems logical that such high-level features rely on prior information processing within the occipito-temporal visual pathway, and therefore only activate thereafter.

Development of geometry representations

Interestingly, the IPS and posterior ITG activations to geometric shapes were consistently observed at very similar locations in adult and 6-year-olds. Furthermore, the overlap with a math-responsive network was present in both age groups, with ROIs that respond to number more than words also responding to geometric shapes more than other visual categories. This finding fits with previous evidence of an early responsivity of the IPS to numbers and arithmetic in young children prior to formal schooling (*52*–*54*). In agreement with the existence of a behavioral geometric regularity effect in preschoolers (*8*), we hypothesize that an intuitive encoding of number and geometric features precede schooling and possibly guide the subsequent acquisition of formal mathematics (*55*).

On the other hand, the fMRI results from the intruder task were much less stable in children, and neither the geometric regularity effect nor the geometry-based RSA analysis yielded strong results at the whole-brain level (only one ventromedial cluster was associated with the regularity effect, see Figure S3). In previous work, we have shown that besides an overall decrease in performance between adults and first graders, the profile of behavior across different shapes is partially distinct in preschoolers: in them, both the CNN and the symbolic model are on par in predicting behavior, suggesting that unlike adults, children frequently mix the two strategies for identifying geometric shapes. Such a mixture of strategies could explain why the correlation with our empirical estimation of complexity, based on adult data, yielded a weaker and more noisy correlation in children. The RSA analysis is compatible with this: in the group analysis, while the CNN predictor was significant in the visual cortex of children, the symbolic geometry model did not reach significance anywhere. Future research with a larger number of participants could attempt to sort out children as a function of whether their behavior is dominated by geometric features or by the CNN model, and then examine how their brain activity profiles differ.

Reduced activation of the ventral visual pathway

Strikingly, our fMRI data also evidenced a hypo-activation of the ventral visual pathways to geometric shapes relative to other pictures. A large bilateral cluster of reduced activity was found when comparing shapes versus other visual categories, and a similar reduced activity was found in each of four visual areas specialized for various visual categories (faces, words, tools, and places). This reduction is compatible with our hypothesis that geometric shapes

are processed differently from other pictures, and with our empirical finding that CNNs, which usually provide a relatively good fit to inferotemporal cortex activity (*56–59*), did not provide a good model of geometric shape perception. Because our shapes consisted of a few straight lines, it is likely that the computations performed by the ventral pathways were minimally solicited, giving rise to a lower activation when compared to more complex visual stimuli such as pictures of faces or houses.

Two Visual Pathways

Brain-imaging studies of the ventral visual pathway indicate that it is subdivided into patches responding to different visual categories such as faces, tools or houses, with a partial correspondence between human and non-human primates (*14*, *60–65*). Converging evidence from behavior (*66*) and neuroimaging (*67–70*) has underlined the central role of the shape of objects, contour and texture in object recognition. A basic complementarity has been identified between the ventral visual pathway, crucial for visual identification, and the dorsal occipito-parietal route, extracting visuo-spatial information about orientation and motor affordances.

Recent work, however, challenges the idea that the global shape of objects is computed in the ventral pathway, and instead suggests that the dorsal pathway, and in particular the IPS, may compute global shape information and then propagate it to the ventral pathway for further processing (*71*). Our results are compatible with this new look at dorsal/ventral interactions: in our fMRI localizer, geometric shapes elicited reduced ventral and increased dorsal activation when compared to other visual categories. Geometric shapes could be considered as conveying very pure information about global shape, entirely driven by contour geometry, and fully devoid of other information such as texture, shading, or curvature. More research is needed to understand the dynamic collaboration between the ventral and dorsal streams during shape recognition, but the present results, as well as the arguments put forward in (*71*) concur to suggest that shape identification does not solely rely on the ventral visual pathway.

Lateral Occipital Cortex

Of special relevance to this work is the role of the Lateral Occipital Cortex (LOC) in shape perception (*72*). The posterior ITG activation that we observed lied just anterior to the LOC and possibly overlapped with it. The LOC has been repeatedly associated with shape processing (*73–76*), and our work converges to suggest that this region and the cortex just anterior to it play a key role in the perception of simple shapes . However, unlike ours, previous work focused primarily on irregular potatoe-like shapes or objects and their contours. Furthermore, object representations in LOC were shown to be stable across depth changes while other visual properties that can be traced across depth do not rely on the LOC (*72*), and the LOC has been shown to be invariant to 2D and 3D rotations (*77*).

Object selectivity in the LOC is already present in early infancy, with a sensitivity to shape and not texture already observed at 6-months of age (*78*). At age 5-10, size invariance, but not viewpoint invariance, has been established (*79*). However, to our knowledge, no study has targeted geometrically regular shapes, and in particular how changes in regularity impacts LOC activity. While some of our quadrilateral stimuli can be constructed as 3D projections of one another (either by perspective or by orthogonal projection), this should actually make them more similar within LOC given its object viewpoint invariance. Further work should use within-subject LOC localizers such as objects minus scrambled-objects to establish the exact cortical relation between the present geometric regularity effect and the classical LO region. We speculate that the more anterior part of lateral occipito-temporal cortex may extract more abstract geometric descriptors of shapes, as also suggested by its sensitivity to mathematical training (*20*).

Artificial Neural Networks: limits and promising directions

While CNNs are predictive of early ventral visual activity, the present work adds to a growing list of their limits as full models of visual perception (*43*, *44*). CNNs have been shown to fail to model many visual illusions (*43*, *44*), geometrically impossible objects (*80*), global shape, part-whole relations and other Gestalt properties (*44*). These limitations suggest the need to supplement current models not only by increasing their size (*56*) but also by changing their architecture in ways that better incorporate findings from psychology and neuroscience (*81*, *82*). Indeed, connectionist model based on the transformer architecture may begin to capture human geometric perception (*45*), and we found that the results generated by our symbolic geometry model could also be captured by the late layers of visual transformer DINOv2 (*46*) as shown in Fig. S5. And Fig. S6. This finding offers an exciting avenue to analyze a mechanistic implementation of symbol-like representations in connectionist architectures, in a paradigm which is simple enough to be tested in a wide array of models and species.

Shape Skeleton and Principal Axis

Skeletal representations have been proposed as human-like representations of visual shapes, particularly appropriate for biological shapes such as animals or plants (*36*). Even for simple geometric shapes, skeletal representation may be automatically and unconsciously computed (*42*). However, in the present work, the two skeletal shape representations we tested did not model human data well (see Fig. S5 and Fig. S6). For our quadrilaterals, which are visually similar to one another, all but the square have the same skeleton topological graph, with only variations in the length and angles of the skeletal segments. Thus, a skeletal representation is probably not the source of the large geometric regularity effect that we observed here and in past work (*8*).

As previously argued (Sablé-Meyer et al., 2022), geometric and medial axis theories need not be seen as incompatible. Rather, we conceive of them as complementary codes, best suited for distinct shape domains. Even after extraction of a figure's skeleton, further compression could be achieved by encoding its geometric regularity, and this may be what humans do when they draw a snake, for instance, as a geometrically regular zigzag.

Neurophysiological implementation of geometry

How could the postulated geometric code be implemented neurophysiologically? There are several challenges. First, some neurons should encode individual features such as lines and curves, or features formed by their relationships (e.g. parallelism, right-angle). Second, such codes should be discrete and categorical, forming sharp boundaries between, say, parallel versus non-parallel lines. Third, they should enter into compositional expressions describing how individual features combine into the whole shape (e.g. "a shape with 4 equal sides and 4 equal right-angles").

The first point has been studied in both humans and monkeys in the context of research on the non-accidental properties (NAPs) of objects. NAPs are qualitative features of object shapes, such as straight versus curved, which remain invariant when an object is rotated. Metric properties (MPs), on the other hand, are quantitative properties such as amount of curvature that do not exhibit such invariance. Behavioral research has demonstrated that, equivalent amounts of pixel change in the image, changes in NAPs are more discriminable than changes in MPs, in educated and uneducated human adults, toddlers and even in infants (*83*–*85*). Furthermore, the firing of neurons in monkey infero-temporal cortex is more sensitive to changes in NAPs than in MPs, whether they are conveyed by 3D shapes (*86*) or 2D shapes (*87*, *88*). These findings occurred even in the absence of a task other than passive fixation, and without particular training for the stimuli. Further work with 2D shapes, including triangles and quadrilaterals partially overlapping with the present stimuli, showed that IT cortex neurons could also be sensitive to more global shape properties such as axis curvature or "taper" (the difference between a rectangle seen upfront or at a slanted axis) (*87*). Multidimensional scaling of neuronal population responses organized the tested shapes in a systematic "shape space" where, for instance, the rectangle occupies a extreme corner, thus making it distinct from its curved or tapered variants (*87*, figure 5).

Thus, monkey IT neurons may provide some of the basic features such as straight line or parallelism that are needed to account for the human encoding of geometric shapes. However, human shape similarity judgments are more discrete and categorical than predicted by either monkey IT neural populations or CNNs (*88*, *89*). The tuning of most IT neurons does not exhibit discrete changes, but is monotonic with continuous changes in shape space, with the strongest firing rates associated to either extreme of the continuum. Monotonic coding of shape variations is also true of other domains such as face perception (e.g. *90*). For instance, Suzuki and Tanaka (*91*) trained macaques for fine-grained object discrimination along small changes in continuous image dimensions (morphed animal pictures), and found neurons in IT that responded monotonically along these shape dimensions, with maximal firing rates at one extreme or the other. With categorical training, neurons in primate prefrontal, temporal and parietal cortices do become sharply tuned to categorical boundaries within a continuum of stimuli (*92*–*94*). Thus, prefrontal and parietal neurons may play a top-down role in imposing a task-dependent categorization to a continuum of shapes (*95*).

While these previous findings may provide a neurophysiological basis for the fMRI and MEG responses observed here, note that humans do not need to be extensively trained to understand the differences between geometric shapes (*55*, *96*), and spontaneously impose geometric categories such as "parallel" or "right angle" to a continuum of angles between lines (*97*). Several behavioral findings suggest that a distinct neural code for geometry may exist in humans. Non-human primates perform poorly on a broad variety of perceptual and production tasks with geometric shapes: baboons do not exhibit a human-like geometric complexity effect (*7*, *8*); chimpanzees do not transfer learning of visual categories from concrete pictures to geometric line drawings (*6*); and chimpanzees behave very differently from children in free-drawing experiments where they have to complete partial line-drawings of faces (*9*). These findings suggests that an understanding of the mechanisms that underlie the human coding of geometric shapes may ultimately shed light on the cognitive and neural singularity of the human brain (*7*).

In the future, replicating the present experiments with monkey fMRI and electrophysiology is therefore an important goal. The methodology could 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" (*12*).

Method

Experiment 1

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 (available at https://www.nytimes.com/2022/03/22/science/geometry-math-brain-primates.html) which 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 personal identity information was collected.

Task

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 tones of ascending or descending pitch, 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).

Stimuli

Shapes design followed previous work exactly (*8*): eleven quadrilaterals with a varying number of geometric features, matched for average pairwise distance of all vertices and length of the bottom side. Shapes were presented in pure screen white on pure screen 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 performed 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 touch-screen), 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."

Estimation of empirical representational dissimilarity

To estimate the representational dissimilarity across shapes, first we estimated the dissimilarity between two shapes as the average success rate divided by the average response time. Indeed, if two shapes are very dissimilar, we expect participants to make few mistakes (high success rate) and find the intruder fast (low response time), yielding a high value, and vice versa. Because we didn't have predictions using the asymmetry in visual search (e.g. finding a square within rectangles versus a rectangle versus squares), we then averaged over these paired conditions, thereby turning the square dissimilarity matrix into a triangular dissimilarity matrix. Finally, we z-scored these dissimilarity estimates.

As participants performed a single trial per pair of shapes, the estimation of the dissimilarity is noisy at the single participant level (either 0 or 1/RT depending on whether they answered correctly). We kept this estimate for analyses at the single participant level or mixed-effect analysis; however, for analyses that required a single RDM estimate

across participants, we pooled the data from participants to estimate the average success rates and response times, hence before estimating the empirical RDMS, rather than estimate one RDM per participant and then averaging.

Comparison with Model RDMs

To obtain theoretical representational dissimilarity matrices (RDMs) with which to compare the present behavioral data, as well as subsequent brain-imaging data, we proceeded as follows. For the CNN encoding model, we downloaded from GitHub the weights for several neural networks (CORnet (*98*), ResNet (*99*) and DenseNet (*100*) ; all high scoring on brain-score (*58*)), 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 (*8*).

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 (*8*). 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.

Experiment 2

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 MRI noise or lying in the confined space for the scanner. 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 known neurological deficit. All adults and guardians of children provided informed consent, and adult participants were compensated for their participation.

Task

In three localizer runs, children and adult participants were exposed to eight different image 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 (*101*). For face stimuli, front-view neutral faces (20 identities, 10 males) were selected from the Karolinska Directed Emotional Faces database (*102*). 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 (*103*)) was not significantly different from French words used here (t(38)=.2, p=.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 previous work (*104*), 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)=.6, p=.749; contrast: F(7,152)=1.2, p=.317), see **Fig.S2**

Procedure

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 (*105*). 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.

Experiment 3

Participants

Identical to experiment 2

Procedure and Stimuli

Task

We adapted the geometric intruder detection task (*8*, *106*) 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 (*8*) 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 previous work (*8*): to fit the experiment within the time constraints of children fMRI, a subset of shapes was used, comprising square, rectangle, isosceles trapezoid, rhombus, right hinge and irregular shapes to span the range of complexity found in (*8*). Following previous work (*8*), 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.

fMRI methods common to experiments 2 and 3

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 (*107*, *108*), which is based on Nipype 1.4.2 (*109*, *110*), and generated the following detailed method description.

Anatomical Data

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (*111*), distributed with ANTs 2.2.0 (*112*), 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 (*113*). Brain surfaces were reconstructed using recon-all (*114*), 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 (*115*). 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 (*116*) and ICBM 152 Nonlinear Asymmetrical template version 2009c (*117*).

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 (*118*). 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 (*119*). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207 (*120*). 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 (*121*)). 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 (*122*). 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(*123*). 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(*124*). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.6.2 (*125*), 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 (*122*) (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, iso-trapezoid, 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 (*126*).

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.

The bar plot for ROIs in figure S4 reflect a 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.

Experiment 4

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. *127*, *128*), 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 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 previous work(*8*) 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.

MEG Acquisition Parameters

Participants were instructed to look at a screen while sitting inside an electromagnetically shielded room. The magnetic component of their brain activity was recorded with a 306-channel, whole-head MEG by Elekta Neuromag© (Helsinki, Finland). The MEG helmet is composed of 102 triplets of sensors, each comprising one magnetometer and two orthogonal planar gradiometers. The brain signals were acquired at a sampling rate of 1000 Hz with a hardware high-pass filter at 0.03 Hz.

Eye movements and heartbeats were monitored with vertical and horizontal electro-oculograms (EOGs) and electrocardiograms (ECGs). Head shape was digitized using various points on the scalp as well as the nasion, left and right pre-auricular points (FASTTRACK, Polhemus). Subjects' head position inside the helmet was measured at the beginning of each run with an isotrack Polhemus Inc system from the location of four coils placed over frontal and mastoïdian skull areas.

Preprocessing of MEG signals

The preprocessing of the data was performed using `MNE-BIDS-Pipeline`, a streamlined implementation of the core ideas presented in the literature (*129*) and leveraging BIDS specifications (*130*, *131*). The pipeline performed automatic bad channel detection (both noisy and flat), then applied Maxwell filtering and Signal Space Separation on the raw data (*132*). 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 (*133*), and thresholds derived with "autoreject global" (*134*). 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 the literature (*135*, *136*)) 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 (*111*), distributed with ANTs 2.3.3 (*112*). 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 (*113*). Brain surfaces were reconstructed using recon-all (*114*) 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 (*115*).

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 since plots using the smoothed version make the separation of the different shapes easier to see. The same holds true for the next analysis.

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.

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 one significant cluster 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, 328]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, 328]ms time window, located in bilateral occipital areas. Additionally, 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.

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/AGeometricShapeRegularityEffectHumanBrain; behavioral data and scripts to generate the models are also available at this url. Neuroimaging data provided upon request.

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 (2015).
7. 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).
8. 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).
9. A. Saito, M. Hayashi, H. Takeshita, T. Matsuzawa, *Child Development*, n/a–n/a, ISSN: 00093920 (Nov. 2014).
10. M. Tanaka, *Animal cognition* **10**, 169–179 (2007).
11. C. Sueur, *Primates*, 1–6 (2025).
12. M. Sablé-Meyer, K. Ellis, J. Tenenbaum, S. Dehaene, *Cognitive Psychology* **139**, 101527, ISSN: 0010-0285 (Dec. 1, 2022).
13. P. Bao, L. She, M. McGill, D. Y. Tsao, *Nature*, 1–6, ISSN: 1476-4687 (June 3, 2020).
14. N. Kriegeskorte *et al.*, *Neuron* **60**, 1126–1141, ISSN: 0896-6273 (Dec. 26, 2008).
15. D. Y. Tsao, S. Moeller, W. A. Freiwald, *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19514–19519, ISSN: 1091-6490 (Dec. 9, 2008).
16. P. Cavanagh, *Perception* **50**, 195–215, ISSN: 0301-0066, 1468-4233 (Mar. 2021).
17. J. A. Fodor, *The Language of Thought* (Harvard University Press, 1975), vol. 5.
18. E. L. J. Leeuwenberg, *The American Journal of Psychology* **84**, 307, ISSN: 00029556 (Sept. 1971).
19. J. Quilty-Dunn, N. Porot, E. Mandelbaum, *Behavioral and Brain Sciences*, 1–55, ISSN: 0140-525X, 1469-1825 (Dec. 6, 2022).
20. M. Amalric, S. Dehaene, *Proceedings of the National Academy of Sciences*, 201603205, ISSN: 0027-8424, 1091-6490 (Apr. 11, 2016).
21. M. Amalric, S. Dehaene, *NeuroImage* **189**, 19–31 (2019).
22. N. Chater, P. Vitányi, *Trends in Cognitive Sciences* **7**, 19–22, ISSN: 1879-307X (Jan. 2003).
23. J. Feldman, *Current Directions in Psychological Science (Wiley-Blackwell)* **12**, 227–232, ISSN: 09637214 (Dec. 2003).
24. N. Kriegeskorte, M. Mur, P. A. Bandettini, *Frontiers in systems neuroscience* **2**, 4 (2008).
25. A. Agrawal, K. Hari, S. Arun, *Psychological Science* **30**, 1707–1723 (2019).
26. A. Agrawal, K. Hari, S. Arun, *eLife* **9**, e54846, ISSN: 2050-084X (May 5, 2020).
27. J. De Leeuw, P. Mair, *Journal of statistical software* **31**, 1–30 (2009).
28. A. Walther *et al.*, *NeuroImage* **137**, 188–200, ISSN: 10538119 (Aug. 2016).
29. V. Ayzenberg, S. F. Lourenco, *Scientific Reports* **9**, 9359, ISSN: 2045-2322 (June 27, 2019).
30. T. Morfoisse, V. Izard, *Journal of Vision* **21**, 2621–2621 (2021).
31. R. Epstein, A. Harris, D. Stanley, N. Kanwisher, *Neuron* (1999).
32. P. Pinheiro-Chagas, A. Daitch, J. Parvizi, S. Dehaene, *Journal of cognitive neuroscience* **30**, 1757–1772 (2018).
33. D. Ansari, *Nature reviews neuroscience* **9**, 278–291 (2008).
34. M. A. Williams *et al.*, *Nat Neurosci* **11**, 1439–1445, ISSN: 1546-1726 (Electronic) (2008).
35. J. Sassenhagen, D. Draschkow, *Psychophysiology* **56**, e13335 (2019).
36. H. Blum, *Journal of Theoretical Biology* **38**, 205–287, ISSN: 00225193 (Feb. 1973).
37. J. Feldman, M. Singh, *Proceedings of the National Academy of Sciences* **103**, 18014–18019 (2006).
38. V. Ayzenberg, F. S. Kamps, D. D. Dilks, S. F. Lourenco, *Neuropsychologia* **164**, 108092 (2022).
39. K. Denisova, J. Feldman, X. Su, M. Singh, *Vision research* **126**, 347–361 (2016).
40. A. S. Lowet, C. Firestone, B. J. Scholl, *Attention, Perception, & Psychophysics* **80**, 1278–1289 (2018).
41. V. Froyen, J. Feldman, M. Singh, *Psychological Review* **122**, 575–597, ISSN: 1939-1471, 0033-295X (Oct. 2015).
42. C. Firestone, B. J. Scholl, *Psychological science* **25**, 377–386 (2014).
43. G. Jacob, R. T. Pramod, H. Katti, S. P. Arun, *Nature Communications* **12**, 1872, ISSN: 2041-1723 (Mar. 25, 2021).
44. J. S. Bowers *et al.*, *Behavioral and Brain Sciences*, 1–74, ISSN: 0140-525X, 1469-1825 (Dec. 1, 2022).
45. D. I. Campbell, S. Kumar, T. Giallanza, J. Cohen, T. Griffiths, presented at the Proceedings of the Annual Meeting of the Cognitive Science Society, vol. 46.
46. M. Oquab *et al.*, *arXiv preprint arXiv:2304.07193* (2023).
47. M. Amalric, I. Denghien, S. Dehaene, *Developmental Cognitive Neuroscience* **30**, 314–323, ISSN: 1878-9293 (Apr. 1, 2018).
48. S. Kanjlia, C. Lane, L. Feigenson, M. Bedny, *Proceedings of the National Academy of Sciences*, 201524982, ISSN: 0027-8424, 1091-6490 (Sept. 16, 2016).
49. Y. Xu *et al.*, *PLOS Biology* **21**, e3001930, ISSN: 1545-7885 (July 25, 2023).

50. V. Ayzenberg, C. Simmons, M. Behrmann, *Cerebral Cortex Communications* **4**, tgad003 (2023).

51. E. Collins, E. Freud, J. M. Kainerstorfer, J. Cao, M. Behrmann, *Journal of Cognitive Neuroscience* **31**, 821–836 (2019).

52. D. Ansari, *Nat Rev Neurosci* **9**, 278–91, ISSN: 1471-0048 (Electronic) (2008).

53. J. F. Cantlon, R. Li, *PLoS biology* **11**, e1001462, ISSN: 1545-7885 (2013).

54. V. Izard, G. Dehaene-Lambertz, S. Dehaene, *PLoS Biol* **6**, 275–285 (2008).

55. V. Izard, P. Pica, E. S. Spelke, *Cognitive Psychology* **136**, 101494, ISSN: 0010-0285 (Aug. 1, 2022).

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

57. J. Kubilius *et al.*, presented at the Advances in Neural Information Processing Systems, vol. 32, (2024; https://proceedings.neurips.cc/paper/2019/hash/7813d1590d28a7dd372ad54b5d29d033-Abstract.html).

58. 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).

59. D. L. Yamins *et al.*, *Proceedings of the National Academy of Sciences* **111**, 8619–8624 (2014).

60. 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).

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

62. 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).

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

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

65. 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).

66. I. Biederman, G. Ju, *Cognitive Psychology* **20**, 38–64, ISSN: 00100285 (Jan. 1988).

67. V. Ayzenberg, F. S. Kamps, D. D. Dilks, S. F. Lourenco, *Neuropsychologia* **164**, 108092 (2022).

68. M. D. Lescroart, I. Biederman, *Cerebral cortex* **23**, 629–637 (2013).

69. P. Papale *et al.*, *Scientific Reports* **9**, 7601 (2019).

70. P. Papale *et al.*, *Journal of Neurophysiology* **124**, 1560–1570 (2020).

71. V. Ayzenberg, M. Behrmann, *Trends in Cognitive Sciences* **26**, 1119–1132, ISSN: 13646613 (Dec. 2022).

72. K. Grill-Spector, Z. Kourtzi, N. Kanwisher, *Vision research* **41**, 1409–1422 (2001).

73. K. Grill-Spector *et al.*, *Human brain mapping* **6**, 316–328 (1998).

74. N. Kanwisher, M. M. Chun, J. McDermott, P. J. Ledden, *Cognitive Brain Research* **5**, 55–67 (1996).

75. Z. Kourtzi, N. Kanwisher, *Journal of Neuroscience* **20**, 3310–3318 (2000).

76. R. Malach *et al.*, *Proceedings of the National Academy of Sciences* **92**, 8135–8139 (1995).

77. Z. Kourtzi, M. Erb, W. Grodd, H. H. Bülthoff, *Cerebral cortex* **13**, 911–920 (2003).

78. L. L. Emberson, S. L. Crosswhite, J. E. Richards, R. N. Aslin, *Journal of neuroscience* **37**, 3698–3703 (2017).

79. M. Nishimura, K. S. Scherf, V. Zachariou, M. J. Tarr, M. Behrmann, *Journal of cognitive neuroscience* **27**, 474–491 (2015).

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

81. 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).

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

83. O. Amir, I. Biederman, K. J. Hayworth, *Vision Research* **62**, 35–43, ISSN: 0042-6989 (June 1, 2012).

84. I. Biederman, X. Yue, J. Davidoff, *Psychological Science* **20**, 1437–1442, ISSN: 0956-7976, 1467-9280 (Dec. 2009).

85. G. Kayaert, J. Wagemans, *i-Perception* **1**, 149–158, ISSN: 2041-6695, 2041-6695 (Dec. 2010).

86. G. Kayaert, I. Biederman, R. Vogels, *Journal of Neuroscience* **23**, 3016–3027 (2003).

87. G. Kayaert, I. Biederman, H. P. Op de Beeck, R. Vogels, *European Journal of Neuroscience* **22**, 212–224 (2005).

88. G. Kayaert, I. Biederman, R. Vogels, *Cerebral Cortex* **15**, 1308–1321 (2005).

89. I. Kalfas, K. Vinken, R. Vogels, *PLOS Computational Biology* **14**, e1006557, ISSN: 1553-7358 (Oct. 26, 2018).

90. L. Chang, D. Y. Tsao, *Cell* **169**, 1013–1028.e14, ISSN: 1097-4172 (June 1, 2017).

91. W. Suzuki, K. Tanaka, *European Journal of Neuroscience* **33**, 748–757 (2011).

92. D. J. Freedman, M. Riesenhuber, T. Poggio, E. K. Miller, *Science* **291**, 312–6 (2001).

93. D. J. Freedman, J. A. Assad, *Nature* **443**, 85–8 (2006).

94. D. J. Freedman, E. K. Miller, *Neuroscience & Biobehavioral Reviews*, The Cognitive Neuroscience of Category Learning **32**, 311–329, ISSN: 0149-7634 (Jan. 1, 2008).

95. J. L. McKee, M. Riesenhuber, E. K. Miller, D. J. Freedman, *The Journal of Neuroscience* **34**, 16065–16075, ISSN: 0270-6474, 1529-2401 (Nov. 26, 2014).

96. V. Izard, E. S. Spelke, *Human evolution* **23**, 213–248, ISSN: 0393-9375 (Jan. 1, 2009).

97. M. R. Dillon, M. Duyck, S. Dehaene, M. Amalric, V. Izard, *Journal of Experimental Psychology: Human Perception and Performance* **45**, 1236–1247, ISSN: 1939-1277 (2019).

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

99. 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.

100. 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).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

127. 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).

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

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

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

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

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

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

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

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

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

137. 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).

Appendix

Materials and Methods

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)

Supplementary text

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.

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).

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 (*8*) – 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 (*8*) (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 (*8*). 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.

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}) (*12*). 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).

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=.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 (*105*). 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).

Additional analysis: Evoked Reponse Potentials (ERP)

Different participants can in principle end up with very different decoder weights, since the decoders are trained independently for each participant. Several of our 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.

MEG: 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 (*137*). 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.

Supplementary Figures

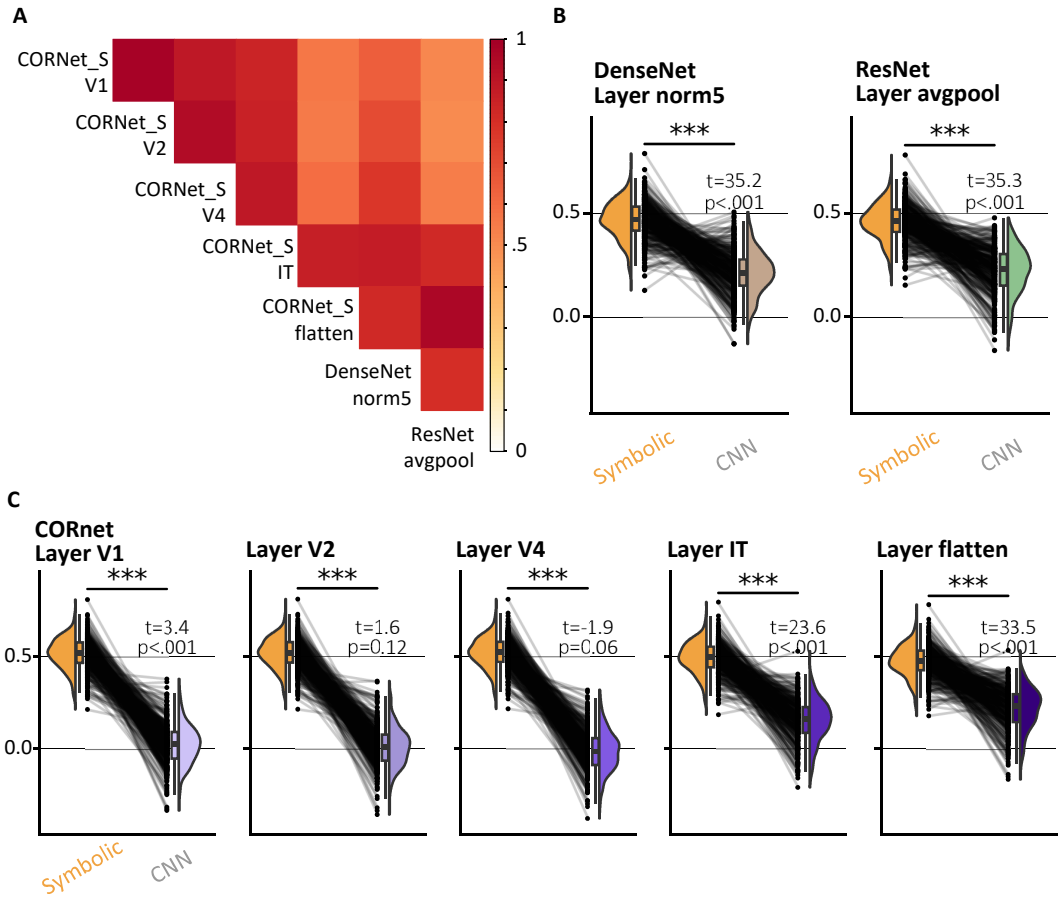

Fig. S1. Additional CNN encoding models of human behavior. **A**, Correlation matrix between all pairs of RDMs generated by each CNN and layer considered. **B,** Replication of **Fig.1D** with different CNNs. Stars indicate a significant difference between the geometric feature model and the respective 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.

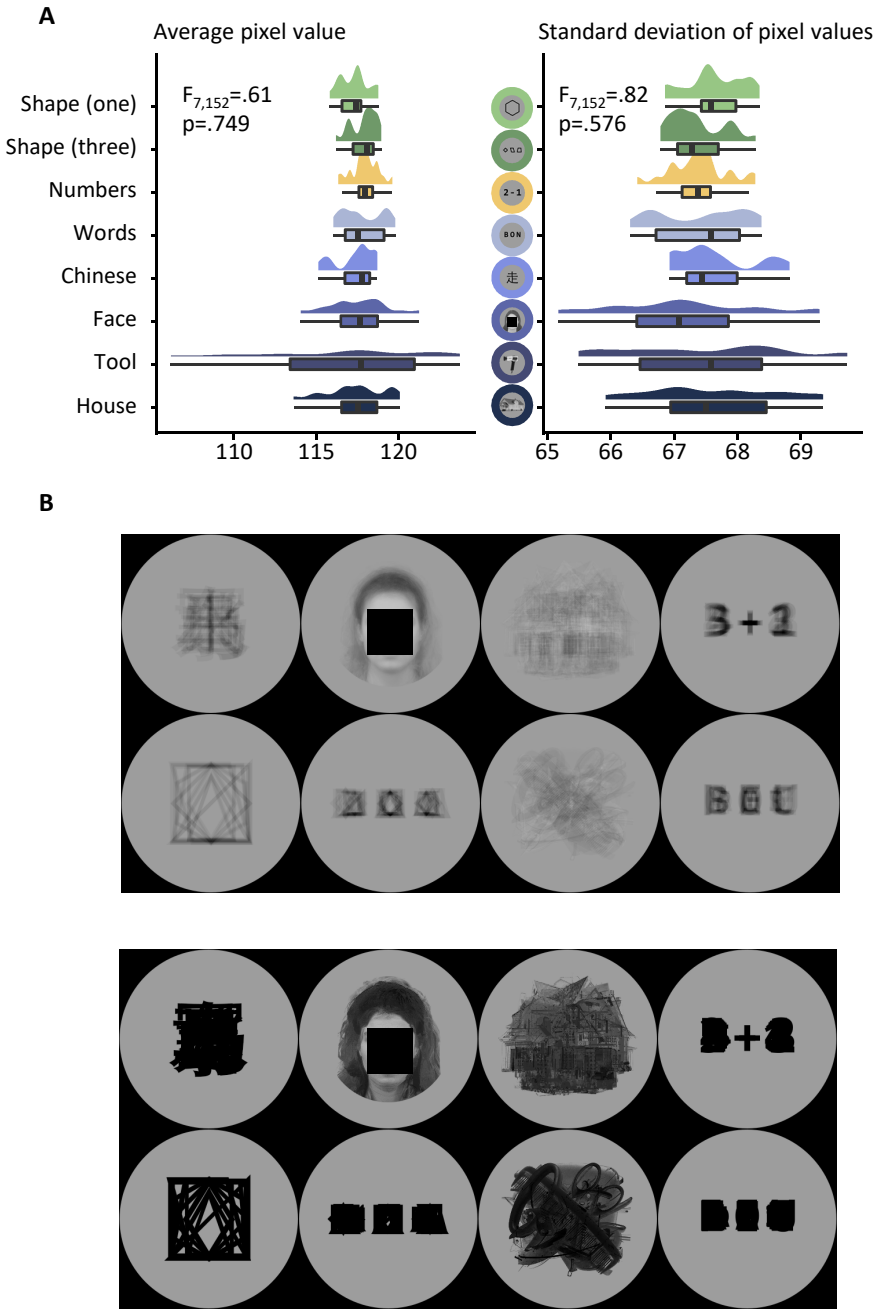

Fig. S2. Overview of the stimuli used for the category localizer**.** A, Average pixel value (left) and average standard deviation across pixels (right) for stimuli within each category (y axis). An ANOVA indicated no significant effect of the stimulus category on either the average or the standard deviation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

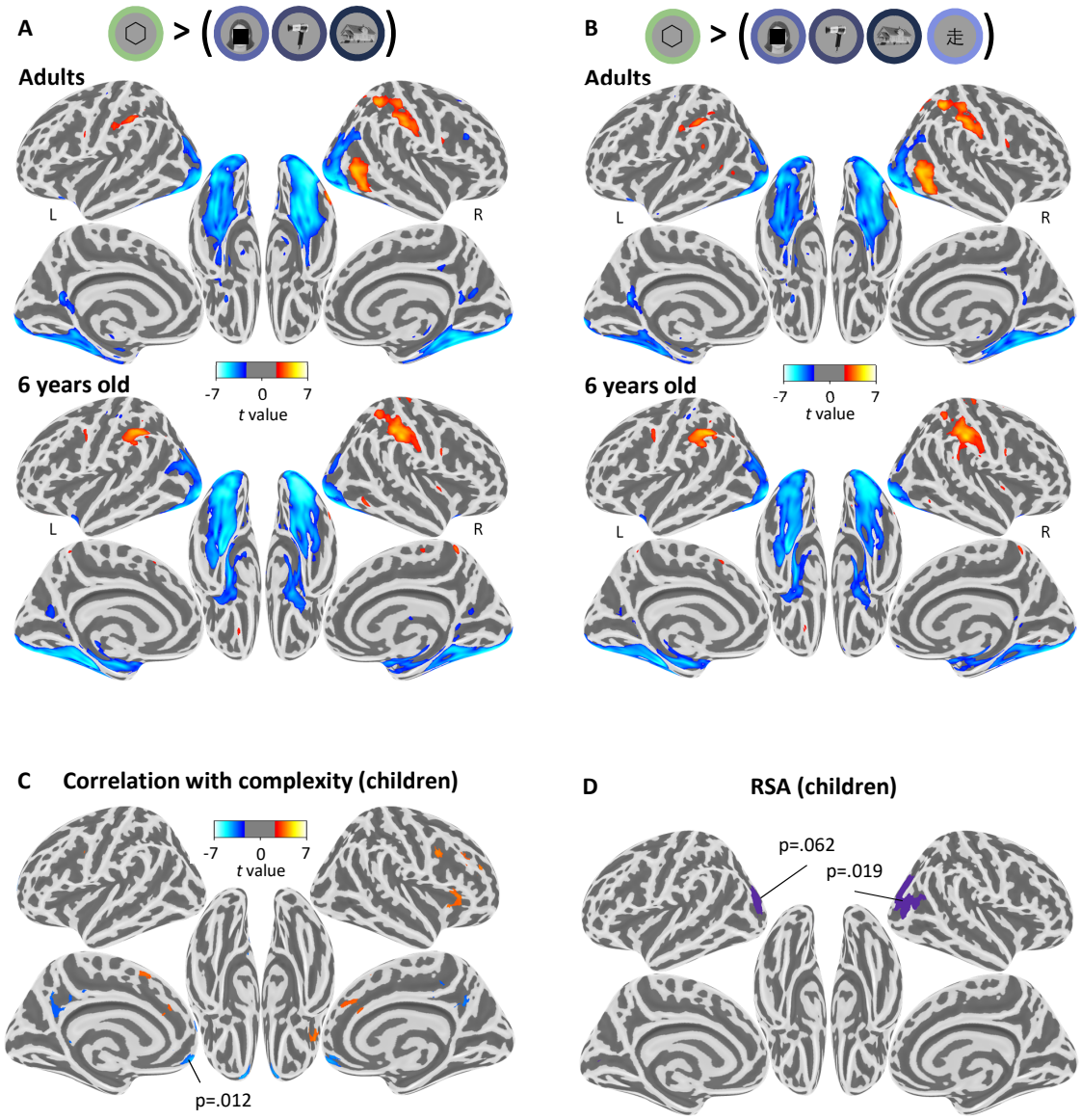

Fig. S3. Details of the fMRI results in children (complement to Fig.2 and Fig.3 in the main text) **A:** Statistical map associated with the contrast "single geometric shape > faces, houses and tools", projected on an inflated brain (top: adults; bottom: children; for illustration purpose we display 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 (*8*). 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)

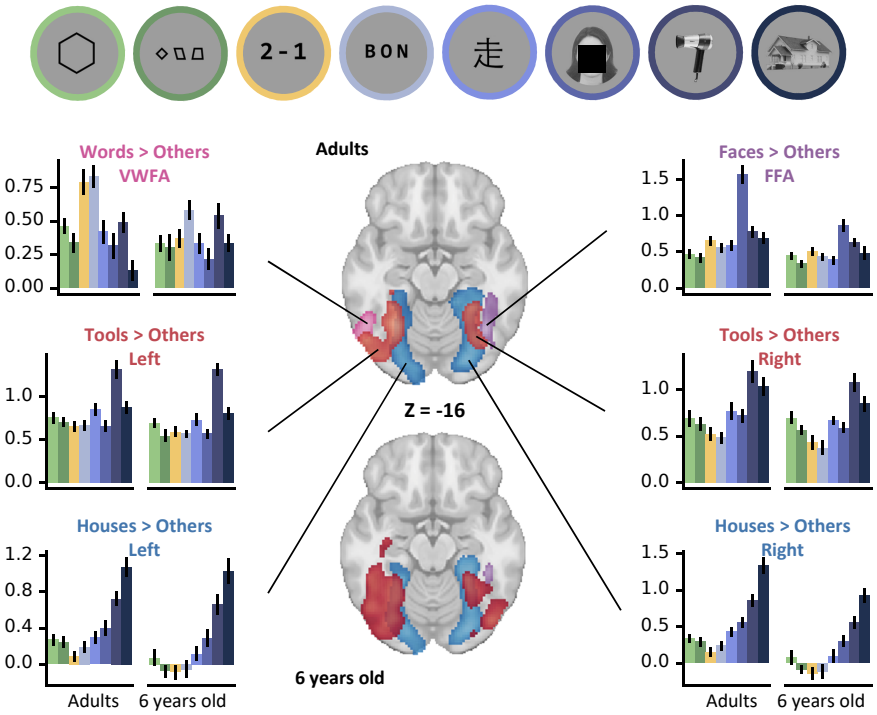

Fig. S4. fMRI response of subject-specific voxels in the ventral visual pathway to geometric shapes and other visual stimuli. The brain slices show the group-level clusters associated to various contrasts known to elicit a selective response in the ventral visual pathway, in both age groups: VWFA (words > 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".

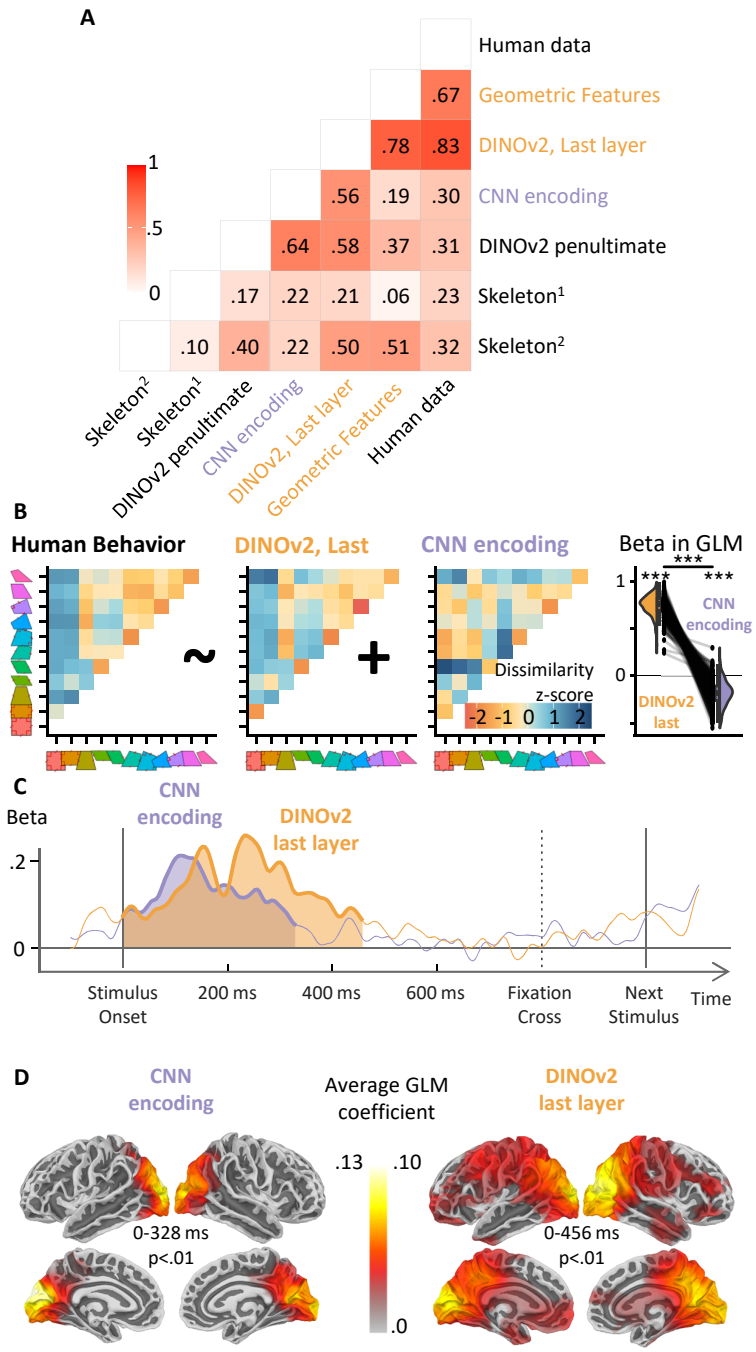

Fig. S5. Additional Models: Behavior and MEG. **A.** Correlation between several models and the average RDM across participants. In particular, we have added the last two layers of DINOv2 (*46*) as well as two different implementation of distances in skeletal spaces (*29, 30*) **B.** Fig. 1D with the symbolic model replaced with the empirical RDM obtained from the last layer of DINOv2 using Euclidean distance. **C.** Fig. 4C with the same replacement of the symbolic model with the last layer of DINOv2. **D.** Fig. 4D with the same replacement of the symbolic model with the last layer of DINOv2.

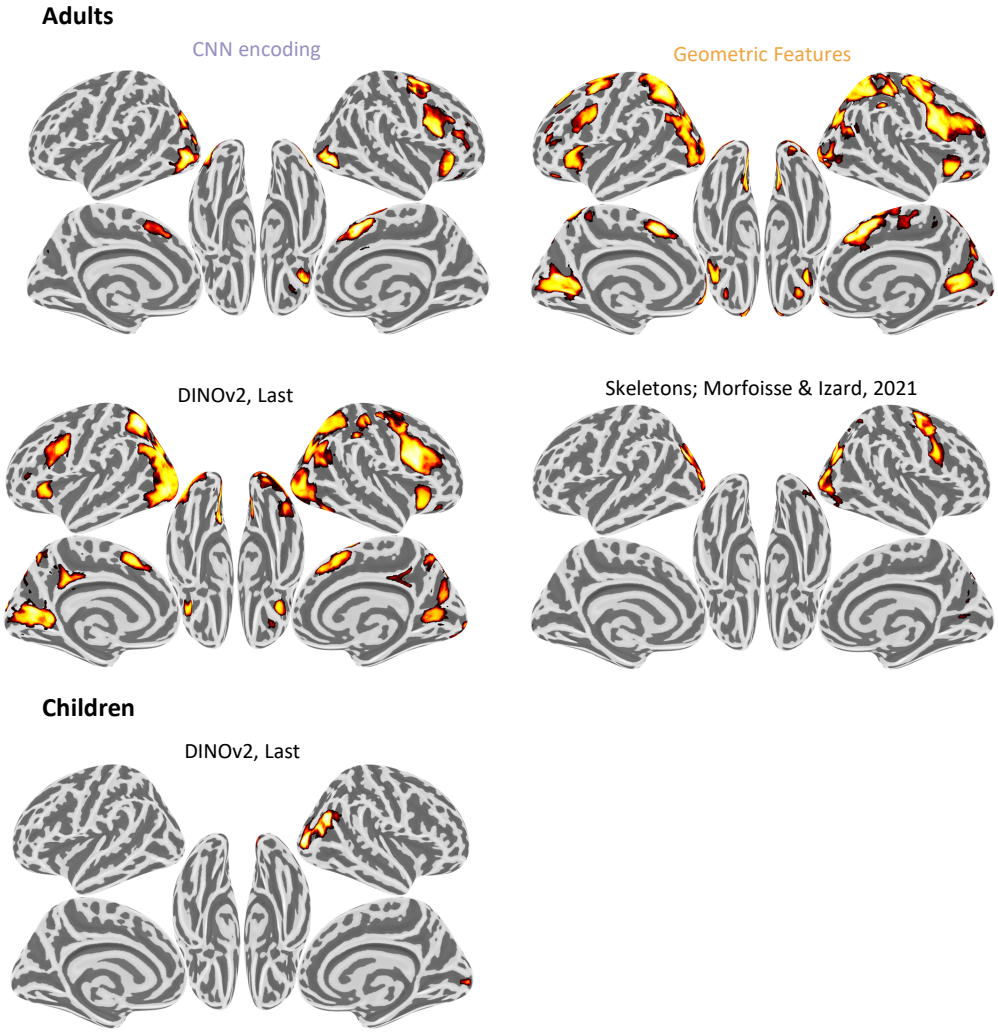

Fig. S6. Additional Models: fMRI. t-values inside significant clusters at the p<.05 level for four models: geometric features (top left), CNN encoding (top right), DINOv2 last layer (bottom left) and skeletal representations from (*30*) (bottom right). Skeletal representations from (*29*) did not yield any significant clusters in adults. In children (bottom), only the DINOv2 elicited any significant cluster.

Tables

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
	-12.5	-76.5	-48.5	5.51	3.9	p=.01
Children	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 (whole brain, permutation test) 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.

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.