Supplementary figures and legends

Figure S1

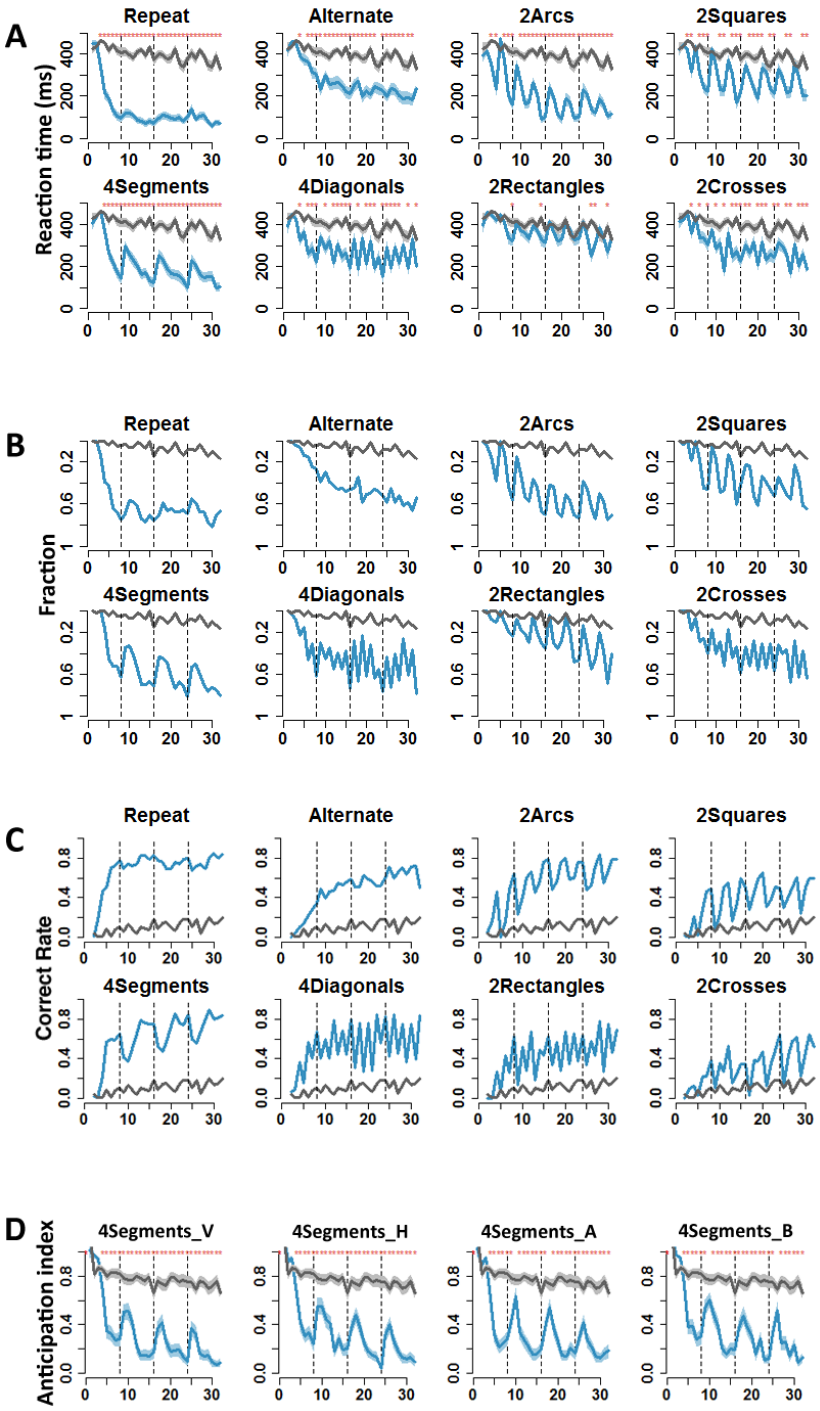

Figure S1. Additional behavioral measures of saccade anticipation (A) Evolution of response time across four trials in each regular sequence. Response time in milliseconds was calculated as the time between the onset of the target and the time point when eye position fell into the target area, i.e. within a distance of 80 pixels from target center. (**B**) Evolution of another measure of saccadic anticipation, the fraction of time that eye position fell in the corresponding target areas at the onset of the target. **(C)** Evolution of correct rate across four trials in each regular sequence. The correct trial was defined as the eye position was near the target areas at the 200 ms after the target onset. **(D)** The four different types of

"4segments" sequences showed a similar evolution of the saccadic anticipation index described in the main text (AI), justifying their averaging together in the main text. Same format as **Figure 1** and **Figure 2A**.

Figure S2

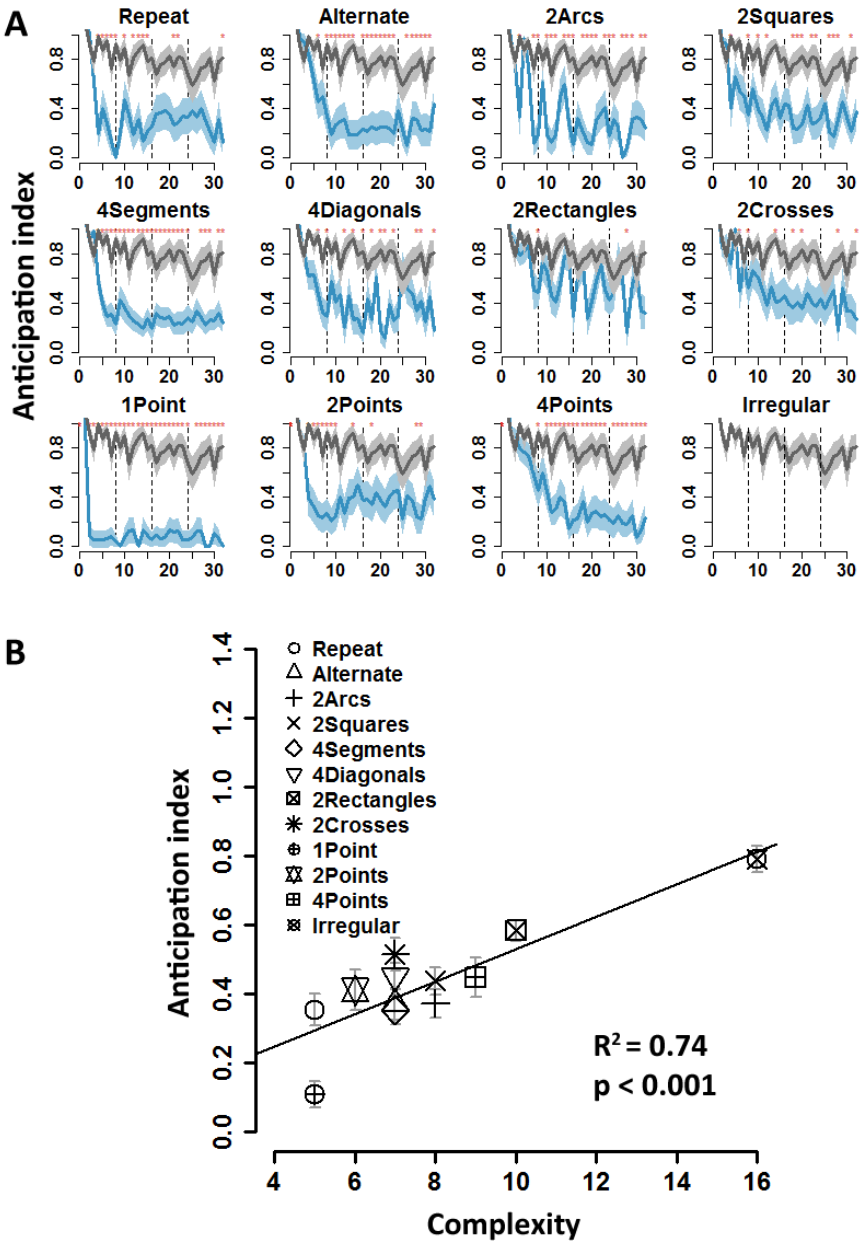

Figure S2. Behavioral performance collected during the fMRI experiment. (**A**) Sequential progression of saccade anticipation across the thirty-two successive locations. (**B**) Sequence complexity was significantly correlated with the anticipation index ($p < 0.001$, $R^2 = 0.74$). Note that the learning of 2nd and 3rd-level of structures in the "2rectangles" and "2crosses" did not show significant difference with the irregular baseline (data points 1, 3, 5 and 7, all $ps > 0.1$, signed rank test). Same format as **Figure 2**.

Figure S3

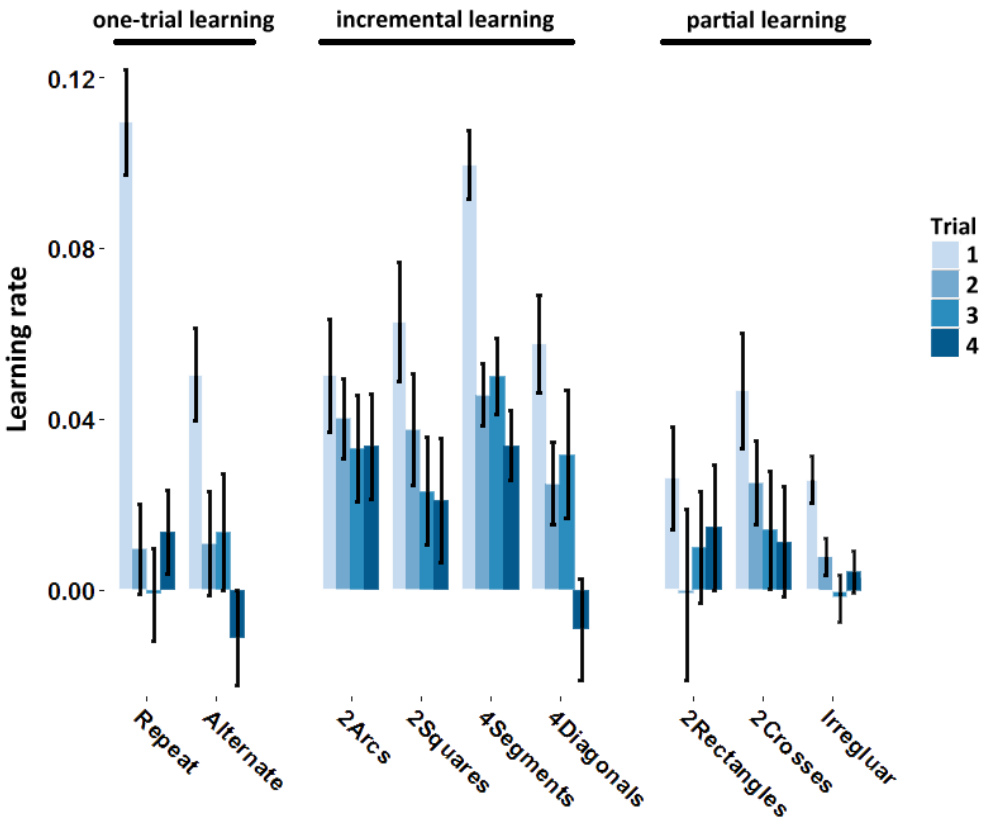

Figure S3. Learning rate within each trial. The learning rate was calculated as the slope of the evolution of the anticipation index AI across the eight data points forming a trial (Fig. 2A). The "repeat" and "alternate" sequences, with a single level of regularity, showed a pattern of "one-trial" learning. The learning rate was significantly positive only in the first trial, and the first-trial learning rate was significantly higher than the following ones (Tukey's HSD (honest significant difference) test, ps < 10^{-5}). Another set of sequences ("2arcs", "2squares", "4segments" and "4diagonals"), with second-level regularities, showed a profile of incremental learning in which the second and third trials continued to show a significant learning rate (Tukey's HSD test, ps < 0.05). Consistent with anticipation results (Fig. 2A), the "2rectangles" and "2crosses" sequences did not show any significant learning rate for each trial (Tukey's HSD test, ps > 0.1) and no between-trials differences (one-way repeated measures ANOVA, "2rectangles": $F(3,129)=0.6$, $p > 0.8$; "2crosses": $F(3,129)=2.1$, $p > 0.6$; "irregular": $F(3, 129)=2.2$, $p > 0.6$).

Figure S4

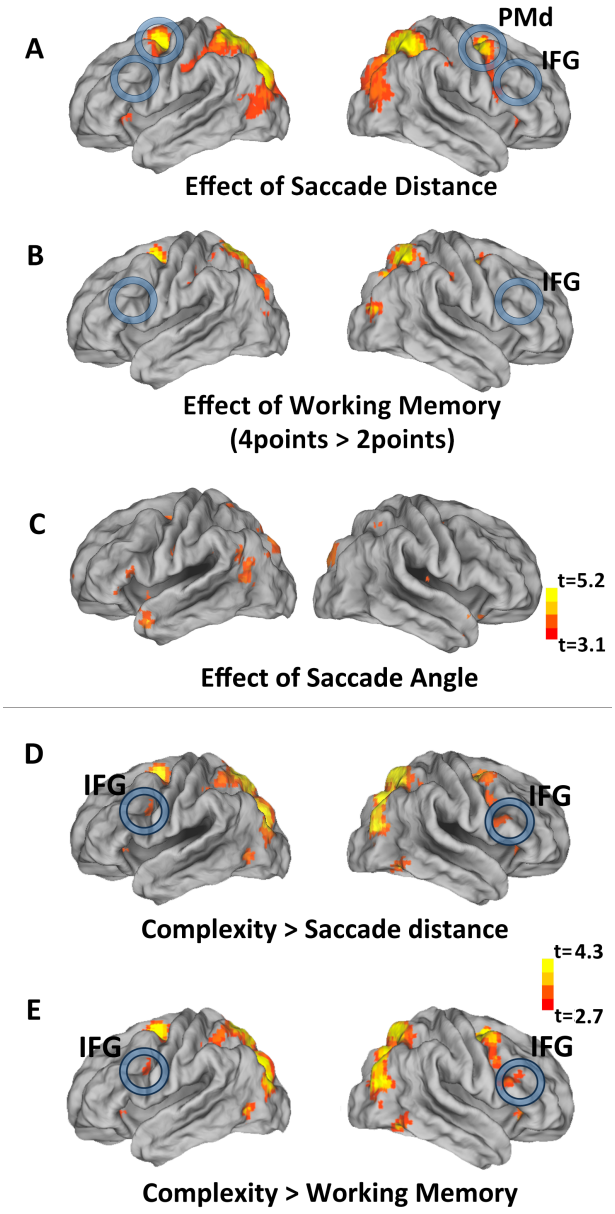

Figure S4. Brain activations related to saccade distance (A), working memory (B) and change of saccade angle (C). Encoding of the saccade distance (group analysis threshold at $t > 3.1$, cluster-level $p < 0.05$ FDR corrected) and working memory (contrast between sequence 4points and 2points at group level, cluster-level $p < 0.05$ FDR corrected) was shown in the bilateral PMd, but not in IFG, and the changes of saccade angle were most involved in the primary visual cortex (group analysis threshold at $t > 3.1$, cluster-level $p < 0.05$ FDR corrected). The differences in brain activities at group level ($p < 0.001$, cluster-level, uncorrected) between complexity and saccade distance effect and between complexity and working memory effect (the contrast of 4Points to rest) are shown in **(D)** and **(E)**, confirming a selective effect of complexity in bilateral IFG areas (same format as **Figure 3**).

Figure S5

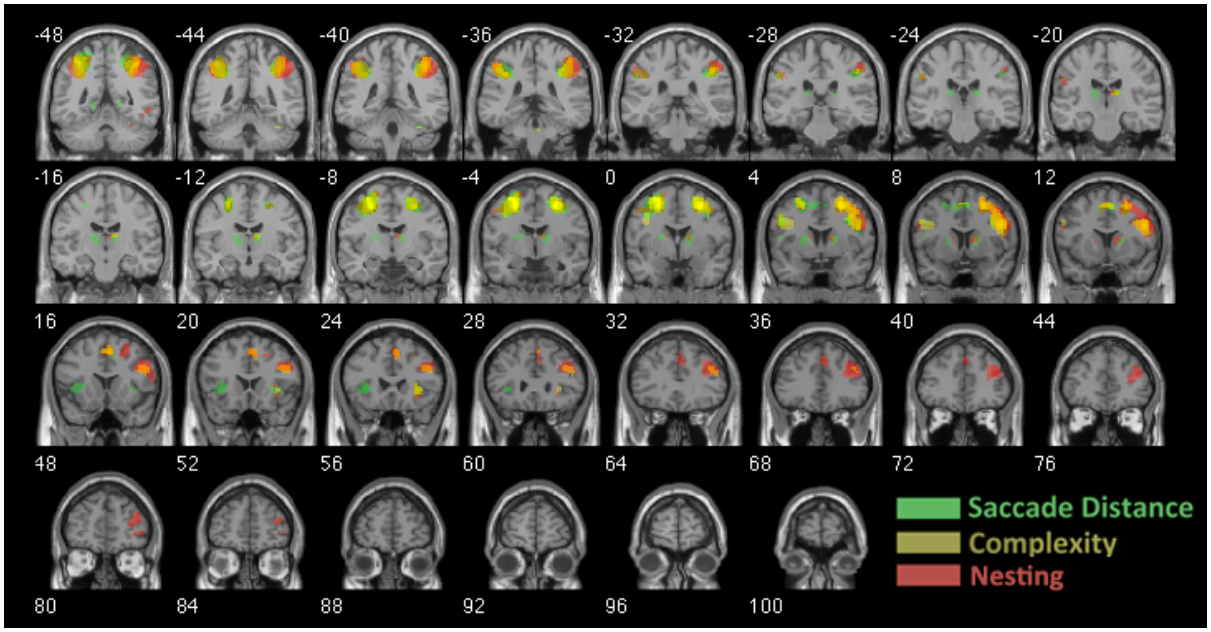

Figure S5. Brain activations related to saccade distance (green), sequence complexity (brown) and sequence nesting (red) are displayed on multiple coronal slices (group level, voxel $p < 0.001$, cluster-level FDR $p < 0.05$ corrected). The inter-slice-spacing was 4 mm.

Figure S6

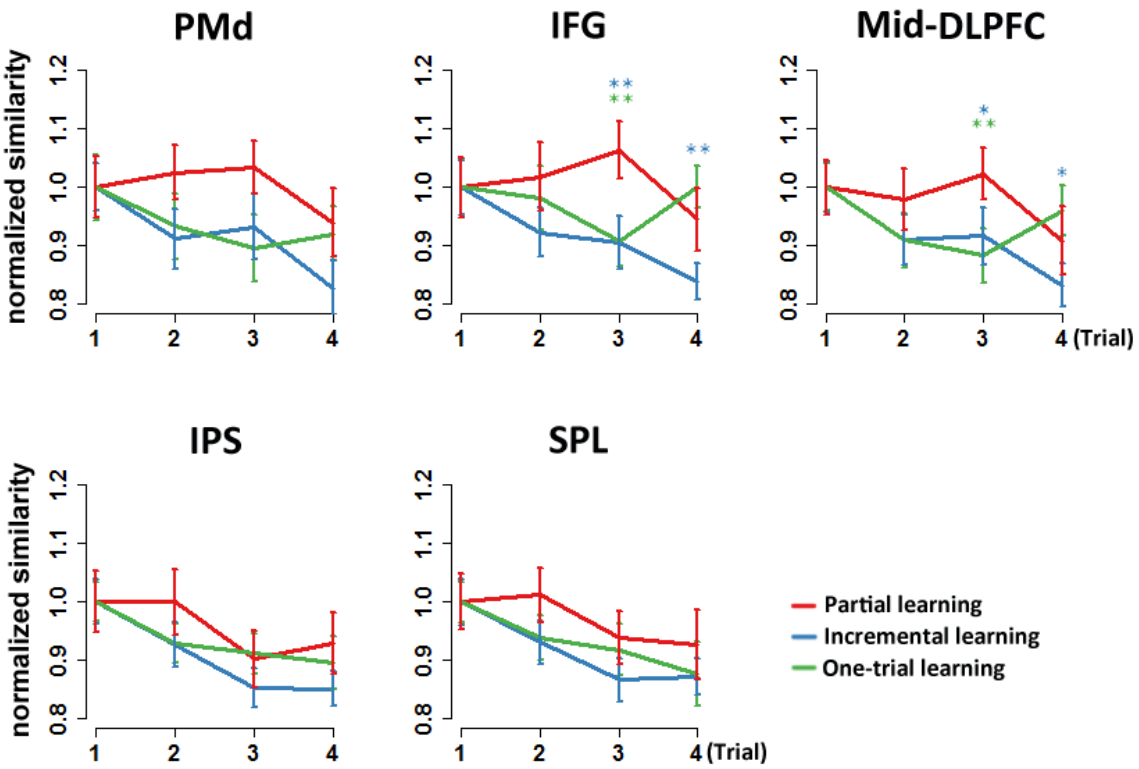

Figure S6. Evolution of representation similarity during learning. Representational similarities were extracted in each ROI using the RSA toolbox (www.mrc-cbu.cam.ac.uk/methods-and-resources/toolboxes/)(Kriegeskorte et al., 2008). Only similarities in frontal IFG and Mid-DLPFC showed significant differences in learning between groups. Multiple comparisons, performed separately for each trial, found that the similarity was significantly different in the third and fourth trial between the partial learning group and the incremental learning group (denoted by blue stars, ** $p < 0.01$, * $p < 0.05$, *t*-test), and between the partial learning group and the one-trial learning group (denoted by green stars, ** $p < 0.01$, * $p < 0.05$, *t*-test) in the frontal areas (IFG and Mid-DLPFC). Abbreviations: IPS, inferior parietal sulcus; SPL, superior parietal sulcus.

Supplementary Tables

Table S1

Regions showing brain activation positively correlated with sequence complexity

x	y	z	t Value	Brain Regions
-42	5	28	3.86	Left inferior frontal
48	8	28	4.94	Right inferior frontal
-24	-4	55	7.04	Left dorsal premotor
27	-1	52	6.72	Right dorsal premotor
6	14	52	4.26	Supplementary motor
-33	-46	37	4.70	Left inferior parietal
42	-37	43	4.98	Right inferior parietal
-18	-67	49	8.20	Left superior parietal
24	-64	52	7.26	Right superior parietal
-30	-76	25	7.41	Left middle occipital
33	-73	28	6.37	Right middle occipital
45	-58	-8	4.57	Right inferior temporal
33	23	1	3.84	Right insular

Significant peaks at a cluster level of $p<0.001$ corrected by FDR $p<0.05$.

Table S2

Regions showing brain activation positively correlated with sequence nesting

x	y	z	t Value	Brain Regions
42	38	25	4.67	Right middle frontal
39	45	15	4.09	Right middle frontal
-12	8	7	3.77	Left caudate
15	11	10	3.72	Right caudate
-51	10	28	3.70	Left inferior frontal
50	17	26	4.14	Right inferior frontal
-24	-4	55	7.04	Left dorsal premotor
27	-1	52	6.72	Right dorsal premotor
3	14	52	3.60	Supplementary motor
-42	-40	40	4.28	Left inferior parietal
42	-40	46	6.36	Right inferior parietal
-24	-61	55	7.65	Left superior parietal
27	-61	55	7.41	Right superior parietal
-30	-76	22	6.55	Left middle occipital
30	-76	19	7.04	Right middle occipital
54	-55	-8	3.40	Right inferior temporal
30	26	-5	3.18	Right insular

Significant peaks at a cluster level of $p<0.001$ corrected by FDR.

Supplementary Materials

Statistical Analysis

At single-subject level, fMRI images were high-pass filtered at 128s and experimental effects in each voxel were estimated using a multi-session design matrix modeling the 60 conditions (15 sequences x 4 trials) and the 6 movement parameters computed at the realignment stage. 60 regressors of interest were thus obtained by convolution of a boxcar function lasting 9.2 seconds (i.e. the duration of any 8-locations sequence) with the standard SPM hemodynamic response function (HRF). For the second-level group analysis, individual contrast images for each of the experimental conditions relative to rest were smoothed with an isotropic Gaussian filter of 8 mm FWHM and entered into a whole-brain ANOVA with sequence as within-subject factor. Unless otherwise noted, the results are reported using a correction for multiple comparisons across the whole brain volume (cluster-level, $p<0.05$ False Detection Rate (FDR) correction, spatial extent > 10). All the effects reported survived FDR correction. Peak activations are reported with the coordinate system of the MNI template brain. Regions of interest (ROIs) were defined by selecting 15mm sphere surrounding the peak voxel in each interested cluster. Marsbar (http:// marsbar.sourceforge.net) was used to extract brain signal in ROIs.

At group level, to search for brain regions whose activation correlated with a given parameter, we used the normalized values of this parameter defined for each sequence as contrast weights. We tested various parameters: saccade distance, saccade direction changes, complexity and nested structure. For sequences in the following order: “repeat”, “alternate”, “2squares”, “2arcs”, “4segments_V”, “4segments_H”, “4segments_A”, “4segments_B”, “4diagonals”, “2rectangles”, “2crosses”, “1point”, “2points”, “4points” and “irregular”, the (Min-Max) normalized distance was [-0.21 -0.07 -0.11 0.06 0.13 0.13 0.13 0.13 0.39 0.03 0.15 -0.61 -0.22 -0.03 0.08]. For direction changes, the normalized angle-change equaled [-0.36 -0.02 -0.25 -0.08 0.20 0.20 0.20 0.20 0.25 -0.17 0.03 -0.67 0.33 0.06 0.07]. To probe the effect of sequence complexity, we used normalized minimal description length: [-0.25 -0.06 0.03 0.03 -0.06 -0.06 -0.06 -0.06 -0.06 0.21 -0.06 -0.24 -0.15 0.12 0.76]. Finally, to determine the effect of sequence nesting, we used the differences of anticipation index between the data point 5 (corresponding to the second-level rules) and the mean of

data point 3 and 7 (corresponding to the first-level rules), with normalized weights: [-0.15 0.04 0.57 0.57 -0.22 -0.29 -0.31 -0.29 -0.03 0.2 0.15 -0.10 -0.43 0.18 0.11].

Spatial relation of the nested-structure effect to the fMRI language and mathematical calculation localizers

Regions showing the effect of anticipation of nested-structure were compared with those involved in sentence processing and arithmetic calculation, as localized by a previously published functional localizer (for details, see Pinel et al.(Pinel et al., 2007). In this 6-min fMRI localizer, brain activations were identified during auditory and visual instructions of left/right hand motor actions, written and spoken language comprehension and mental calculation. Functional images were acquired on the 20 subjects who performed the sequential saccade task, with a TR of 2.4s and a voxel size of 3×3×3 mm. In the present study, we used two contrasts: sentence processing (spoken and written sentence relative to rest) to localize regions for language processing; and mental calculation (relative to sentence processing), which identified brain regions of mathematical thinking (Fig. 5A and 5B).

As shown previously (Pinel et al., 2007), calculation activated the bilateral intraparietal sulcus, SMA and IFG regions (Fig. 5B, denoted in cyan, $p < 0.001$, cluster-level FDR $p < 0.05$ corrected). The calculation-related regions largely overlapped (Fig. 5B, denoted in brown) with the regions exhibiting an effect of anticipation of nested structures (Fig. 5B, denoted in yellow, $p < 0.001$, FDR $p < 0.05$ corrected) in our spatial sequence paradigm. Fig. 5A does not reveal any such overlap with the bilateral temporal cortex, left inferior prefrontal gyrus and premotor cortex (Fig. 5A, denoted in red, $p < 0.001$, FDR $p < 0.05$ corrected) that were activated by sentence processing.

Analysis of language-related and Mathematic-related ROIs

We performed an analysis with individual region of interests (ROIs) for the 20 subjects who underwent the functional localizer. For each subject, within each of 7 language-related and 7 mathematics-related ROIs respectively reported by Pallier et.al. (2011) (Pallier et al., 2011) and Amalric et.al. (2016)(Amalric and Dehaene, 2016). The 7 language-related ROIs in the left hemisphere are: 1, TP - temporal pole; 2, aSTS - anterior

superior temporal sulcus; 3, pSTS - posterior superior temporal sulcus; 4, TPJ - temporal parietal junction; 5, IFGorb - inferior frontal gyrus pars orbitailis; 6, IFGoper- inferior frontal gyrus pars percularis; 7, IFGtri - inferior frontal gyrus pars triangularis. The 7 mathematics-related ROIs are: 1,2, left and right IPS – intraparietal sulcus; 3,4, left and right SFG – superior frontal gyrus; 5,6, MFG – medial frontal gyrus; 7, SMA – supplementary motor area. We first used the functional localizer to identify subject-specific voxels activated by sentence processing (voxel $p < 0.001$, uncorrected) and mathematical calculation (voxel $p < 0.001$, uncorrected). Then within each ROI, we extracted the subject-specific contrast values of each spatial sequence (12 sequences (Fig. 1B)) relative to rest at these identified voxels. These values were further tested statistically in contrasts independent from those used to defined them, thus avoid circularity and "double-dipping" (Fig. 5C and 5D).

Representational similarity analysis (RSA)

The analysis of neural activity within ROIs was conducted with the RSA toolbox (www.mrc-cbu.cam.ac.uk/methods-and-resources/toolboxes/). The three frontal areas (PMd, IFG and Mid-DLPFC) were selected as ROIs defined in the Figure 4, and spherical ROIs in right IPS and right SPL were generated centered on the peak voxels (IPS: [42 -40 46], SPL: [27 -61 55], from Table S2) with a 15 mm radius. We compared the sequence-wise (15 sequences defined in *Methods*) patterns amongst fMRI t-maps for the four trials (Trial 1, 2, 3 and 4). Per subject, the representational dissimilarity matrixes (RDMs) comprised correlation distances (1-correlation coefficient) between the images from the four trials for each sequence, which yielded a 60 × 60 matrix per region. The 15 sequences were grouped into three categories according to their learning rate (Fig. S3): one-trial learning (Repeat and Alternate), incremental learning (2Arcs, 2Squares, 4Segments_H, 4Segments_V, 4Segements_A, 4Segments_B, 4Diagonals) and partial learning (2Rectangles, 2Crosses and Irregular) sequences. We then performed a two-way ANOVA with two main factors: *Trial (1, 2, 3 and 4)* and *Sequence group* (one-trial learning, incremental learning, and partial learning) to compare the representational similarities (normalized by the first trial) between sequences within each trial.

Supplementary reference:

Amalric, M., Dehaene, S., 2016. Origins of the brain networks for advanced mathematics in expert mathematicians. Proc Natl Acad Sci U S A 113, 4909-4917.
Amalric, M., Wang, L., Pica, P., Figueira, S., Sigman, M., Dehaene, S., 2017. The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. PLoS Comput Biol 13, e1005273.
Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2, 4.
Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T., 2000. Automated Talairach atlas labels for functional brain mapping. Hum Brain Mapp 10, 120-131.
Pallier, C., Devauchelle, A.D., Dehaene, S., 2011. Cortical representation of the constituent structure of sentences. Proc Natl Acad Sci U S A 108, 2522-2527.
Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Le Bihan, D., Poline, J.B., Dehaene, S., 2007. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neurosci 8, 91.