An abstract language of thought for spatial sequences in humans

Fosca Al Roumi*[,1,5], Sébastien Marti[1,†] , Liping Wang[3], Marie Amalric[1,4] and Stanislas Dehaene*[1,2]

1. Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France

2. Collège de France, 11 Place Marcelin Berthelot, 75005 Paris, France

3. Institute of Neuroscience, Key Laboratory of Primate Neurobiology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China

4. CAOS Lab, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213

* Correspondence: fosca.alroumi@gmail.com and stanislas.dehaene@cea.fr
[5] Lead Contact
[†] Decease January 11, 2019

Abstract

How do humans encode spatial and sequential information in working memory? We tested the hypothesis that participants do not merely store each consecutive location in a distinct memory slot, but instead compress the whole sequence using an abstract language-like code that captures geometrical regularities at multiple nested levels. We exposed participants to sequences of fixed length but variable regularity, while their brain activity was recorded using magneto-encephalography. The entire sequence could be decoded from brain signals. Anticipation signals could also be decoded, and both behavior and anticipatory brain signals were modulated by sequence complexity, defined as the minimal description length provided by the formal language. Furthermore, abstract primitives of rotation and symmetry could be decoded both in isolation and within the sequences. These results suggest that humans encode the transitions between sequence items as rotations and symmetries, and compress longer sequences using nested repetitions of those primitives.

Introduction

Compared to other primates, the human species seems to be endowed with a peculiar ability to extract the abstract rules that underlie the structure of the information present in the environment (Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015). Such rule learning, indeed, is essential in order to acquire language, science, mathematics or music – activities which are all unique to humans. Furthermore, it is present in early childhood: even infants are able to quickly extract patterns from a stream of syllables, infer abstract regularities from a small number of examples, and generalize them to new items (Kabdebon, Pena, Buiatti, & Dehaene-Lambertz, 2015; Marcus et al., 1999; Saffran, 2003; Xu & Tenenbaum, 2007).

A prominent hypothesis is that the capacity to represent embedded structures may be specific to the human species (Hauser, Chomsky, & Fitch, 2002; Fitch, 2014). Indeed, several comparative studies found that, while human and non-human primates exhibit similar performance in sequence processing when sequential relations, statistical properties or simple non-adjacent dependencies suffice to perform the task (Hauser, Newport, & Aslin, 2001; Milne et al., 2016; Milne, Petkov, & Wilson, 2018; Newport, Hauser, Spaepen, & Aslin, 2004; Ravignani, Sonnweber, Stobbe, & Fitch, 2013; Sonnweber, Ravignani, & Fitch, 2015; Wilson, Marslen-Wilson, & Petkov, 2017; Wilson et al., 2013), important differences are observed whenever embedding or recursion are present (Fitch, 2004; Wang, Uhrig, Jarraya, & Dehaene, 2015). For instance, in a spatial motor task, although macaque monkeys could learn a supra-regular grammar and generalize it to novel sequences, they need a training period of thousands of trials to achieve the performance level that preschool children reach in only a few trials (Jiang et al., 2018).

Thus, the difference between human and non-human primates may originate, at least in part, in their fast abilities to acquire complex sequential rules. In the present research, which is part of series of behavioral and brain-imaging studies of spatial sequence learning (Amalric et al., 2017; Wang et al., 2019), we test the specific hypothesis that, when they learn a spatial sequence, human subjects make use of a language-like system of nested rules of variable complexity. To boost intuition, imagine that you have to view a sequence of 8 locations which are successively flashed on screen, in order to reproduce this sequence after a delay (i.e. the classical “Corsi blocks” neuropsychological test of spatial working memory).

Since the sequence length of 8 exceeds the typical working-memory span, it would be difficult to remember it after a single presentation. Imagine, however, that you discover a strong regularity in the sequence -- perhaps the first four items form a square, and the next four another square. Mentally remembering this sequence as “2 squares”, thus compressing the information into an internal language-like expression, would considerably facilitate its memorization.

In previous research (Amalric et al., 2017; Wang et al., 2019), we formalized this idea by proposing a hypothetical “language of thought” (Fodor, 1975) for sequences. Our language is akin to a mini computer language: it consists of primitives and rules whose combination can express any sequence of locations on an octagon (see figure 1). The language comprises simple primitives of rotation and symmetry (e.g. next item on the octagon, symmetrical item over the vertical axis, etc; figure 1C) as well as an operator of repetition with variation (see Supplementary Materials for a complete description). Since the language allows these instructions to be nested, it can express multiple levels of embedded repetitions and represent concepts such as “square”, “rectangle”, “two squares”, etc.

Our hypothesis is that, whenever they perceive a spatial sequence, human subjects attempt to represent it in memory by searching for the simplest mental program that can generate it. Thus, a simple prediction is that the difficulty of memorizing a sequence should not be proportional to its length, but to the length of the shortest mental program in the language of thought. Thus, sequences with a compact program should be easily memorizable, even though their length may exceed the typical memory span of ~7 items, whereas at the other extreme, sequences with no compact description other than the mere list of locations would be harder or even impossible to learn. The underlying hypothesis is that the brain operates as a “compressor” of incoming information that tries to select the minimal description for incoming stimuli, and that minimal description length is a good predictor of psychological complexity (Chater & Vitányi, 2003; Feldman, 2000; M. Li & Vitányi, 1993; Romano, Sigman, & Figueira, 2013).

In a first behavioral study, we evaluated the hypothesis that participants encode spatial sequences using this language of geometry (Amalric et al., 2017). Participants were presented with sequences of 8 dots organized as a regular octagon and spanning a large range

of complexities. They only saw the beginning of these sequences and had to predict the next locations. The results indicated that performance was highly sensitive to sequence complexity. Specifically, error rate increased linearly as a function of minimal description length, and the error patterns across ordinal positions in the sequence reflected the nested structure of the expression postulated by the formal language. Another group (Yildirim & Jacobs, 2015) also showed how a similar compositional language for spatial sequences could account for the transfer of abstract sequence knowledge from the visual to the auditory modality. However, unlike the present work, they did not model participants' sensitivity to geometrical regularities since the presented locations did not form regular geometrical shapes.

In a follow-up fMRI experiment on geometrical sequences (Wang et al., 2019), we merely asked participants to move their eyes to each item while the same 8-location sequences were presented on an octagon. Again, behavior indicated that gaze anticipation was inversely related to minimal description length and reflected each sequence's nested structure. Activity in the dorsal part of inferior prefrontal cortex correlated with the amount of compression, while the right dorsolateral prefrontal cortex encoded the presence of embedded structures. Those brain regions belonged to a network that is distinct from but close to the areas involved in natural language processing, consistent with previous results indicating a dissociation of the neural circuits involved in mathematical thinking and in natural language processing (Amalric & Dehaene, 2017; Maruyama, Pallier, Jobert, Sigman, & Dehaene, 2012; Varley, Klessinger, Romanowski, & Siegal, 2005). Although the content of sequences could not be decoded from fMRI signals, multivariate pattern analyses provided indirect evidence that the activity patterns in dorsal prefrontal cortex became increasingly differentiated as the sequences were learnt.

fMRI is notoriously insensitive to time, and thus failed to directly probe the precise temporal unfolding of language-like rules that our theory predicted. In the current study, we aimed to further probe the existence of an abstract, language-like representation of geometrical sequences in the human brain by using magneto-encephalography (MEG), a sensitive technique with much higher temporal resolution than fMRI. We used time-resolved multivariate decoding (King & Dehaene, 2014) in order to examine if the postulated primitives and rules could be decoded at the precise moment in the sequence where the postulated

language of geometry suggests that they should be deployed. Thus, we exposed human participants to several repetitions of geometrical sequences of variable complexity. To keep them focused, we asked them to click as soon as they had identified the repeated sequence; furthermore, we also asked them to detect occasional sequence violations. We then tested if MEG signals were sensitive to the postulated primitives, sequence complexity, and hierarchical structure. We identified markers of an anticipated representation of the sequence items and assessed their modulation by sequence complexity. Moreover, we exposed participants to multiple exemplars of each primitive operations in isolation, and we probed if those primitives could be decoded from the MEG signals. This approach allowed us to determine if the sequences were indeed represented as nested combinations of primitive rules.

Results

Primitive rules: behavioral results

During the primitive part of the experiment, participants were exposed to a succession of two dots forming a pair illustrating a given hypothetical primitive operation (e.g. all pairs had a vertical symmetry). For instance, in a given block, for all pairs, the second dot was always vertically symmetrical to the first one, thus testing the postulate primitive of vertical symmetry. All 11 primitives of the language were tested (see the list in figure 1C, and the Methods section). We asked participants to click a button whenever they felt that they could predict the location of the second item in each pair. Those responses were converted into a measure of "encoding time" (figure 2A), expressed as the number of pairs that needed to be presented before the response occurred (if participants failed to respond, the maximum number of presentations, i.e. 32, was used). To assess if participants had understood the primitive rule and did not simply memorize the 8 pairs by rote, we introduced a control condition in which the pairs were not driven by any general rule, and participants therefore had to memorize each of them. Participants performed very poorly in this condition, most of them failing to respond before the end of the run (i.e. after 32 pairs were presented). T-tests showed that, for all of the 11 proposed primitives, encoding time was shorter than for the control condition (all ps $< 10^{-5}$). Such savings indicate that subjects detected all regularities.

In the second half of each block, we introduced a violation detection task: participants were asked to press another button as fast as possible whenever they detected that the second dot of a pair was misplaced (figures 2B and C). In the control condition, participants missed 95% of those violations, where the miss rate never exceeded an average of 53% for the 11 genuine geometrical primitives. Again, all the differences relative to control were significant (all ps < 10^{-4}).

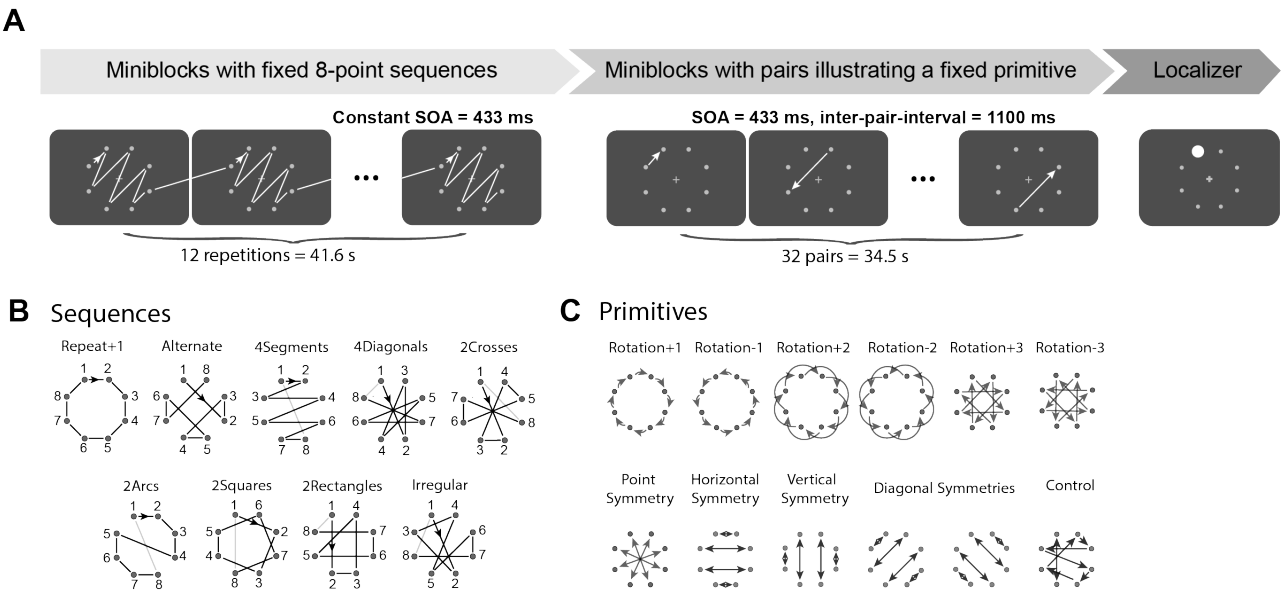

Figure 1: Experimental paradigm and stimuli

A. The experiment was divided into three parts: (1) In the sequence part, during each mini-block, a fixed spatial sequence of 8 locations was repeated 12 times; (2) In the primitive part, during each mini-block, 32 pairs of two successive locations illustrating a given primitive were repeated times; In both cases, participants were asked to report when they had identified the sequence or rule governing the pairs, and to click whenever they detected a violation. Finally (3), in the localizer block, dots were flashed at random locations on the octagon, and the data was used to train a location decoder.
B. The nine 8-location sequence templates used in the sequence part. Presentation order is indicated by arrows. Actual sequences were generated by varying the starting point, rotation direction and/or symmetry axis.
C. The pairs of locations illustrating each of the eleven primitive rules presented in the primitive part. Arrows indicate the first and the second element of each pair.

Taken together, those results show that participants identified all of the primitive rules used to form the pairs and could use them to detect outliers. To determine if these rules were processed with the same ease, we ran repeated-measures ANOVA on three dependent behavioral measures: encoding time, violation detection time, and violation miss rate. This

analysis revealed significant differences among the primitives (respectively $F(10) = 5.26$, $p < 10^{-4}$; $F(10) = 3.31$, $p = 0.001$; and $F(10) = 4.00$, $p = p < 10^{-4}$). Tukey post-hoc tests on encoding time and miss rate indicated that the primitives of rotation ±3 were significantly more difficult than the counterclockwise rotation-1, the point symmetry and the vertical symmetry.

In conclusion, behavioral measures indicated that the participants could detect all of the postulated geometrical primitives, but that, contrary to our initial assumptions (Amalric et al., 2017) those primitives may not be strictly equivalent in complexity, with rotation±3 being more difficulty to detect (for a similar conclusion, see Romano et al., 2013).

Geometrical sequences: behavioral results

During the geometrical sequence block, subjects were repeatedly exposed to 8-item sequences. Two behavioral measures of sequence complexity were obtained. First, subjects were asked to press a button whenever they felt that they had identified the repeating sequence precisely enough to predict the next item. We again analyzed the encoding time, defined as the number of sequence repetitions that the participants needed before responding (if participants failed to respond, the maximum number of repetitions, i.e. 12, was used). As predicted, this number increased with our theoretical measure of sequence complexity, i.e. minimal description length in the proposed language (Spearman rank-correlation $\rho = 0.45$, $t(19)=9.6$, $p < 10\text{-}7$; Pearson correlation coefficient $r = 0.70$, $t(19)=9.6$, $p < 10\text{-}13$, $t(19) = 20.7$) (figure 2D). Again, after ten repetitions of a given sequence, i.e. during the last two presentations, a violation could occur (i.e. a single dot was misplaced, off by 3 locations on the octagon). Participants were asked to press a button as fast as possible when they detected it. The violation detection times again exhibited a positive correlation with complexity (figure 2E; Spearman rank-correlation $\rho = 0.38$, $t(19)=7.1$, $p < 10^{-5}$; Pearson correlation coefficient $r = 0.40$, $t(19)=6.7$, $p<10^{-5}$). This was also the case for the number of missed trials (figure 2F, $\rho = 0.33$, $t(19)=6.3$, $p<10^{-5}$); $r =0.43$, $t(19)=7.1$, $p < 10^{-6}$). Indeed, the effect on errors was very large: subjects missed less than 10% of deviants in the simplest sequence, but more than 50% of them in the most complex sequence, consistent with the idea that they had trouble memorizing it.

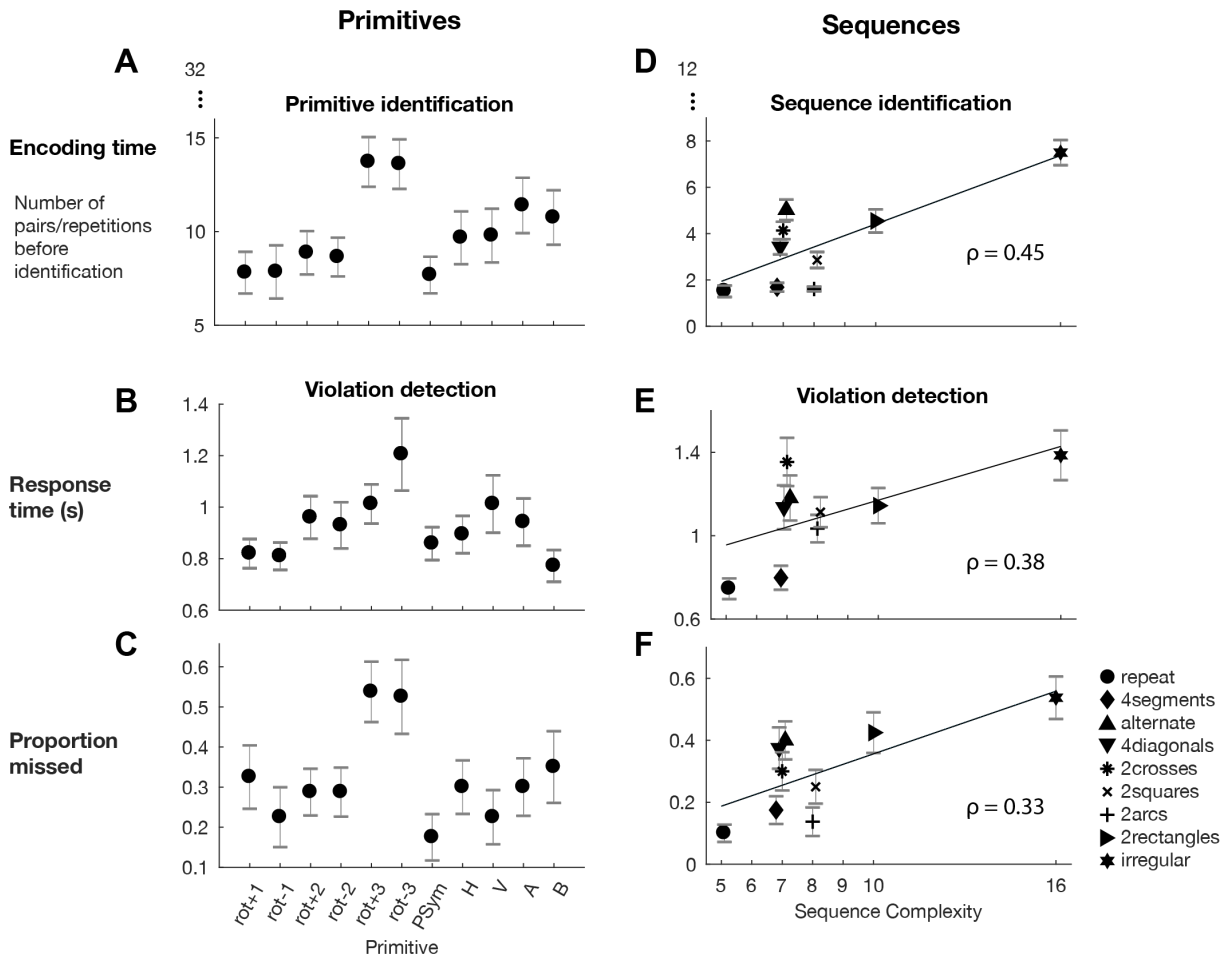

Figure 2: Behavioral performance and impact of sequence complexity

For primitive rules (left column) and geometrical sequences (right columns), graphs show (1) the mean encoding time, i.e. the mean number of repetitions that participants had seen before they reported identifying the rule or the sequence (A,D); (2) the performance in violation detection, as reflected in the mean response time to violations (top, B and E) and the proportion of missed trials (bottom, C and F). Error bars indicate one standard error of the mean (sem). For geometrical sequences, linear regression lines and Spearman coefficients indicate a significant effect of theoretical sequence complexity.

Overall, these results converge with the ones obtained using explicit predictions of the next item (Amalric et al., 2017) or using implicit eye-movement anticipations (Wang et al., 2019). They indicate that the more complex the sequences, the harder it is to predict the next locations, and therefore, the harder it is to detect violations. Nevertheless, the correlations with minimal description length were modest, suggesting that our measure of sequence complexity, based on the proposed language, may not be ideal. To examine this, we also computed the intercorrelation between our dependent measures. Both the detection time and the missed rates were well predicted by the encoding time, with correlations similar to

those obtained from our theory-driven measure (respectively $\rho=0.54$, $t(19)=10.5$, $p<10^{-8}$, $r=0.51$, $t(19)=9.8$, $p<10^{-8}$, and $\rho=0.44$, $t(19)=5.7$, $p<10^{-4}$, $r=0.48$, $t(19)=5.4$, $p<10^{-4}$). We extracted the first principal component of those 3 dependent behavioral measures (figure 2D, E and F). This empirical measure of complexity showed a robust correlation with theoretical complexity ($\rho=0.56$, $r = 0.75$), but also showed deviations for alternate, 4diagonals and 2crosses sequences, which could be explained by the fact that those sequences contained the rotation±3 primitives. In the rest of this paper, we shall use both theoretical and empirical complexity measures as predictors of brain activity.

Decoding the successive locations of each sequence item

We first tested whether MEG signals contained decodable information about the successive locations of each sequence item. At each time point, based on the 306 sensor measures, we trained an estimator to decode the angular position of the presented item (flashed at time 0s). This position decoder was trained on data independent of the 8-item sequences and for which no anticipation could be formed (i.e. data from a localizer block, where locations were randomly intermixed, as well as data from the first item of each pair in the pair block, which was unpredictable; see Methods for details).

As shown in figure 3A, position decoding was at chance prior to stimulus presentation, but rose suddenly ~70 ms after stimulus onset, and peaked at 140 ms. The generalization-across-time matrix (figure 3B) revealed both a diagonal, indicating an unfolding sequence of stages, and a partial square pattern indicating a sustained maintenance of location information in brain signals (King & Dehaene, 2014).

This location decoder, trained on independent data, was then applied to the 8-item sequence data. Successful generalization was observed: at each time step, the decoder successfully identified, with above-chance accuracy, the location of the current item in the sequence (figure 3C). To maximize the decodability of individual items in each sequence, we retrained a location decoder, but this time on the average brain responses in the time window of maximal decodability (from 100 to 200ms after the stimulus onset). We then tested it on each individual sequence. Figure 3D presents, for each sequence separately, the relative amount of times that the decoder predicted each location for each ordinal position. The pattern obtained

from these predictions tightly paralleled the actual spatiotemporal profile of each sequence (second line of figure 3D), with only some added spatial uncertainty (i.e. spreading of the decoding to the two locations surrounding the correct one on the octagon).

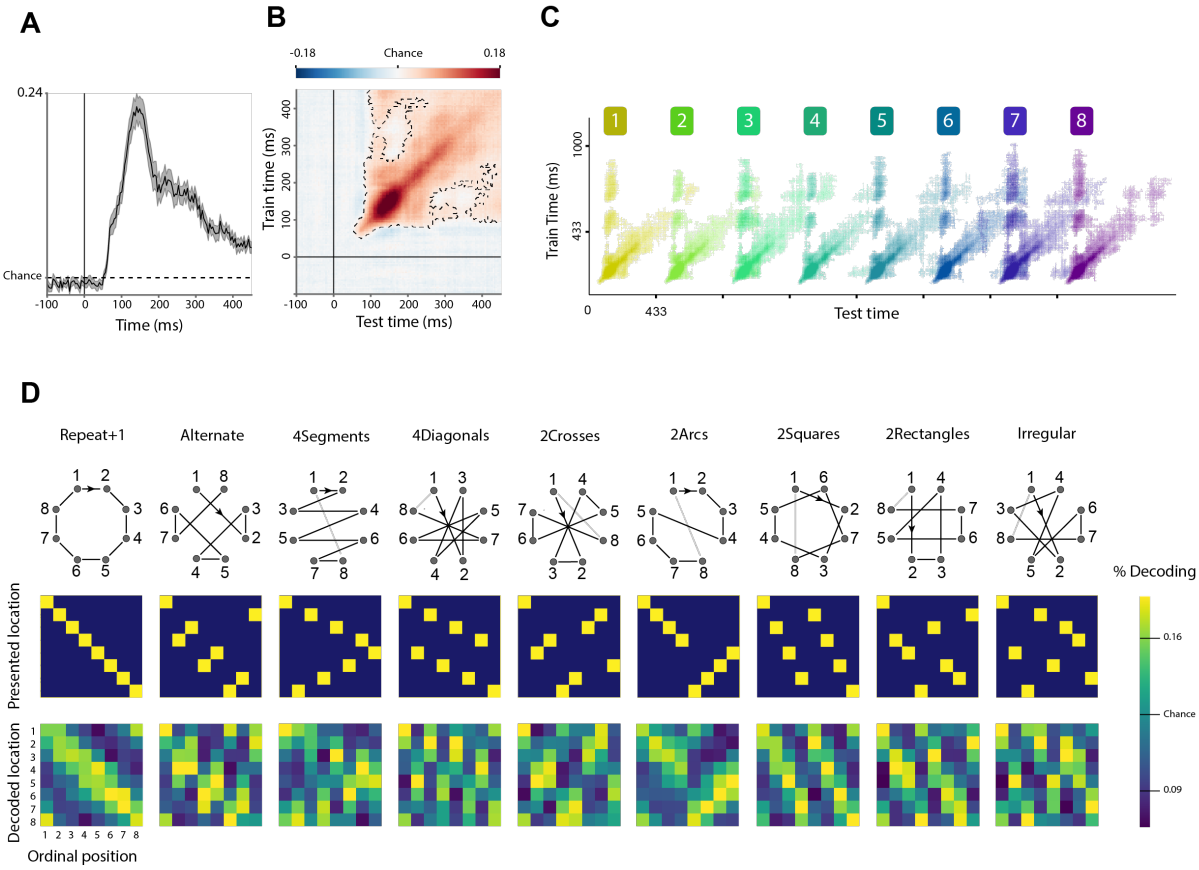

Figure 3: Decoding successive sequence locations from MEG signals

A. Performance in decoding stimulus location as a function of time following the flash of a dot at a given location. Maximal decoding performance is achieved at ~150 ms.
B. Average generalization-across-time matrix showing the location decoding score as a function of training time (y axis) and testing time (x axis). The dashed lines indicate $p < 0.05$ cluster-level significance, corrected for multiple comparisons (see *Methods* for details on the cluster-based permutation analyses).
C. Thresholded ($p < 0.05$, corrected) decoding matrix plot showing when each of the 8 successive sequence items could be decoded (averaged across subjects).
D. Decoding of each of the nine 8-location sequences. For each sequence, the top matrix shows the stimuli at each ordinal position, and the bottom matrix shows the proportion of times a given spatial location was decoded.

Prior research on predictive coding has demonstrated that predictable stimuli elicit a reduced brain response, yet a more faithful representation as reflected by a higher decoding accuracy

(Summerfield & de Lange, 2014). Here, we would expect this pattern to be present for easily predictable sequences with low complexity, but not for the more complex ones. To test this prediction, we examined if there was a linear modulation of the average decoding score as a function of complexity. We ran a linear regression on the decoding score as a function of theoretical and empirical complexity for each participant and determined if the regression coefficients associated to complexity were significantly negative, indicating that more complex sequences elicited a less precise internal representation of the successive locations. A small modulation was found as a function of empirical complexity (one tailed t-test $t(19) = -2.0$, $p=0.0297$) but not of theoretical complexity (one tailed t-test $p>0.1$).

Anticipation and its modulation by sequence structure

Our next goal was to determine if anticipatory information was present in MEG even prior to actual stimulus presentation. Indeed, the two behavioral tasks that participants had to perform during the experiment encouraged them to anticipate the location of the next stimuli in order to detect violations. The theory predicted that such anticipations would decrease with sequence complexity, indicating that more complex sequences were increasingly harder to predict.

To test for the presence of anticipation and its modulation by sequence complexity, we assessed the performance of the position decoder prior to the presentation of each sequence item (time 0ms). Importantly, since there was evidence of spillover of the decoding to nearby locations on the octagon, we controlled for distance to the previous item. To this aim, we defined an anticipation score as the difference in the decoding score at the upcoming location and at the symmetrical but non-stimulated location, which was equidistant from the previous location (see Methods and figure 4A). This anticipation score was computed for all sequence items (figure 4B) then averaged across the training time window 100-200 ms (figure 4C) which corresponds to the maximal performance of the position decoder (figure 3A). We ran a cluster-based permutation test in the temporal window between the presentation of the preceding item and the anticipated item (i.e. from -430ms to 0ms). The anticipation score was significantly positive in two time windows, from -220 to -115 ms and from -85 to 0 ms prior to stimulus presentation. Additional analyses, presented as supplementary materials, excluded a contribution of eye movements to those anticipation signals (see figure S1).

We then used the anticipation score to probe the internal representation of the sequences. If predictive mechanisms are modulated by sequence structure, then the anticipation score should be increasingly smaller as the sequence gets more complex. We computed the average anticipation score for each sequence (except 4diagonals and 2crosses; see Methods), and ran a linear regression as a function of complexity. We observed a significant decrease of anticipation score with theoretical complexity (one-tailed t-test $t(19)=-2.1$, $p=0.025$) and empirical complexity (one-tailed t-test $t(19)=-2.0$, $p=0.028$). This result provides indirect evidence for the hierarchical representation postulated by the language of geometry, as it shows that expectation mechanisms, measured by the anticipation score, are modulated by the overall complexity of the sequence. Note that this effect did not depend on distance, but solely on complexity, as a similar regression showed that the anticipation score did not vary significantly with the distance between the anticipated item and the preceding one ($t(19)=-1.4$, $p>0.1$).

Previous research has shown that brain activity in language areas is modulated by the nested structure of language, such that activity varies sharply at the boundary of sentence constituents such as noun phrases (Nelson et al., 2017). We thus wondered if a similar effect occurred with the language of geometry. To do so, we therefore compared the anticipation scores of the items that, according to our postulated language, open a constituent, to the ones that are inside a component (i.e. inside a series of repeated instructions). The analysis was run only on the 4segments and 2squares sequences. Repeat1, alternate and irregular sequences were excluded from it as their postulated representation does not involved a nested syntactic structure. We also excluded the 2arcs sequences since all constituent opening corresponded to distance-4 transitions (see Methods). Finally, we also discarded the complex 2rectangle sequence, since its anticipation score was not significantly different from zero. Perhaps due to this reduced data set, and the ensuing lack of statistical power, this analysis did not reveal any significant difference as a function of syntactic structure.

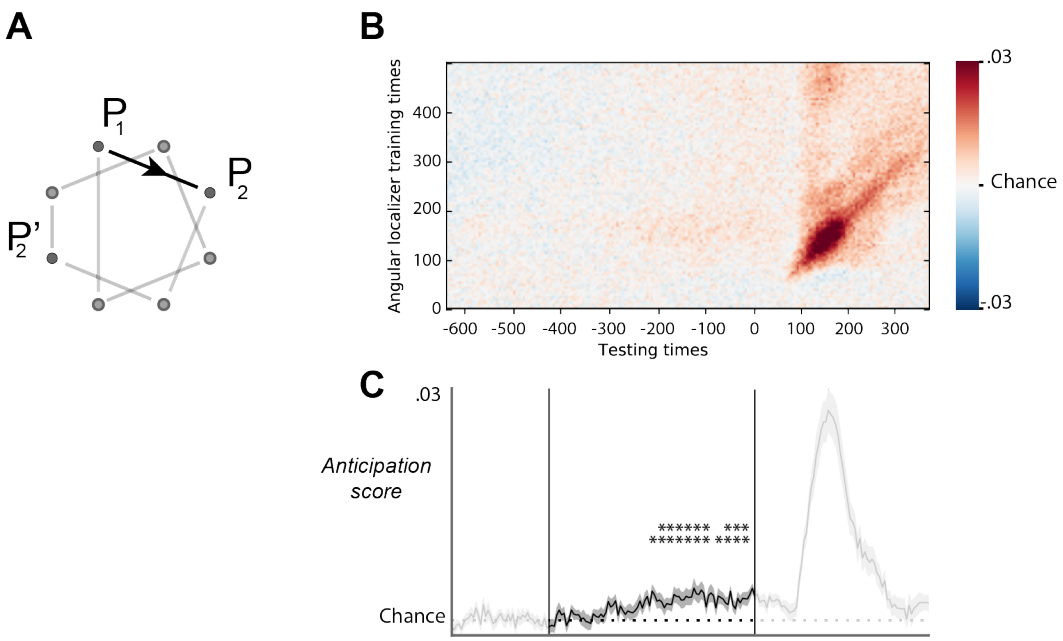

Figure 4: Detecting an anticipation of sequence locations from MEG signals

For each training time in the range 0-500 ms, we tested whether the location decoder could detect the stimulus location in a time window ranging from 650 ms before to 400 ms after stimulus presentation.

A. To control for distance from the previous stimulus location, an anticipation score was computed by comparing decoding at the correct stimulus location and at the equidistant, symmetrical, non-stimulated location.
B. The matrix representation indicates small generalization of the peak location decoder prior to stimulus presentation during the sequence block.
C. The temporal evolution of the anticipation score was obtained by averaging the anticipation score over training times ranging from 100 to 200 ms. The anticipation score was significantly above chance starting ~250 ms before stimulus presentation, thus demonstrating that participants could predict the next sequence location.

Decoding elementary primitive operations

The previous analyses were all based on the simple decoding of retinotopic location. The language of geometry hypothesis, however, predicts that participants encode spatial sequences, not only in terms of the specific locations involved, but also in terms of high-level primitives such as symmetries and rotations. To directly probe the representation of the sequences in terms of the language of geometry, we tried to decode from the brain signals the elementary primitives postulated by our formal language.

To do so, we first determined if those primitive operations could be decoded when they were presented in isolation. Using the miniblocks illustrating each primitive (see figure 1), we trained a decoder that took as input the average brain signals in the time interval 0-430 ms

between the presentation of the two elements of the pairs, and attempted to determine which of the 11 primitives was presented (*cf* Methods). Our rationale was that, during this time window, participants must be actively applying the corresponding transformation (e.g. symmetry around the vertical axis) on the first element of the pair in order to predict the location of the second. The performance of the cross-validated decoder was above chance (chance = 1/11=0.09; mean = 0.19, one-tailed t-test to chance $t(19) = 12.7$,$p < 10^{-10}$) and this was also the case for each participant individually. The 11x11 confusion matrix is presented in figure 5. We computed the relative amount of times that the decoder predicted the correct primitive for each primitive separately. Each primitive had an above chance score ($p<10^{-4}$) and a repeated-measures ANOVA did not reveal differences across them.

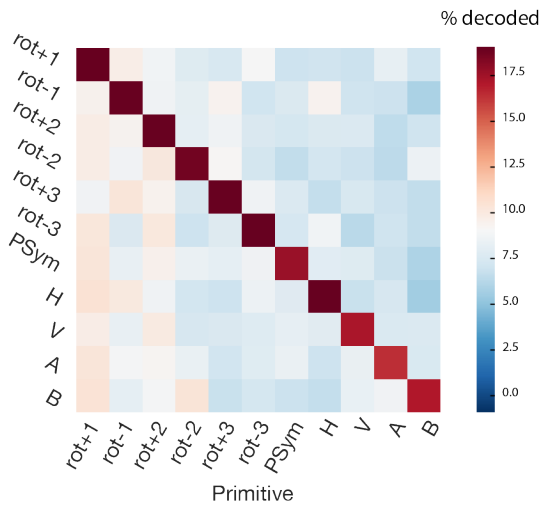

Figure 5: Decoding the 11 geometrical primitives

A decoder for each of the 11 geometrical primitives was trained on the average brain signals between the presentation of the two elements of the pairs (0-430 ms). The 11x11 confusion matrix shows the proportion of times that the decoder predicted the correct primitive for each primitive separately. Each primitive had an above chance score ($p<10^{-4}$) and a repeated-measures ANOVA did not reveal differences across primitives.

The above decoding may suffer from a visual confound: the average distance between the dots in a pair varies with the primitive operation. Therefore, if sensitive to this visual feature, a decoder could perform above chance without capturing the abstract geometrical operation. To overcome this limitation, we capitalized on the fact that the same pairs of dots could appear in the context of either a rotation block or a symmetry block. For instance, a dot moving from the top left to the top right location can be construed as either a rotation around

the octagon, or as a symmetry around the vertical axis. Thus, we selected all such pairs and asked whether their brain signals allowed to decode whether they were encoded as a rotation or as a symmetry (see Methods). A decoder fed with the average brain signals in the time interval 0-430 ms from the onset of the first element of the pair performed above chance (mean = 0.58, $t(19) = 11.0$,$p = <10^{-8}$; all participants were above chance). Thus, over and above any location information, the neural patterns associated with abstract geometrical operations could be disentangled.

We then determined if the primitive operations that were identified in the blocks with pairs only could also be detected during the 8-location sequence blocks. The above decoder, trained on the 11 primitive pairs, indeed had a slightly above-chance score when tested on all the sequences (mean = 0.0973, $t(19) = 3.45$,$p =0.003$). However, the decoder trained only on the balanced set of rotation and symmetry pairs, as explained in the previous paragraph, did not ($p>0.7$). We also trained a new decoder from the sequence data itself. Again, we built a similarly balanced set of rotation and symmetry trials from all the sequences except the irregular one. Their label (rotation or symmetry) was attributed based on our formal language. The decoder performance was above chance (mean = 0.57, $t(19) = 6.9$, $p <10^{-5}$), suggesting that participants indeed distinguished the rotation and symmetry segments of the sequences, and encoded them differently.

It could be argued that, in this analysis, rotation and symmetry were partially confounded with sequence complexity, because most distance-1 rotation trials came from the simplest rotation+1 and rotation-1 sequences. To avoid such a confound, we therefore restrained the analysis to rotation segments drawn from the alternate sequences, versus symmetry segments from the 4segments sequences. Both sequences had the same complexity (7), yet the decoder performance was again above chance (mean = 0.67, $t(19) = 8.8$, $p <10^{-7}$).

In summary, the MEG decoding analysis provided evidence for an abstract encoding of rotations and symmetries, independently of the visual features of the stimuli, both in the primitive and in the sequence parts of the experiment. However, this code did not fully generalize from the primitive to the sequence part.

Discussion

The goal of the present study was to probe the internal code that humans use to represent geometrical sequences of varying regularity. We used MEG decoding to determine whether and how primitives and rules were used to encode the geometrical sequences. This technique also allowed us to find evidence for covert anticipation of sequence items and to assess its modulation by sequence complexity.

To determine how participants represented elementary primitive rules, we exposed them to multiple exemplars of each rule in isolation and to a control condition in which the pairs were not driven by any general rule. Differences relative to the control condition indicated that participants did not simply memorize the 8 pairs by rote, but were able to quickly grasp all the underlying rules and detect violations.

The next step was to decode from the brain signals the abstract representation of these primitive rules. To do so, we trained a decoder to predict the primitive that was shown to the participant. Even when carefully controlling for visual features, such a decoder was able to do so. In the sequence part of the experiment, participants were exposed to several repetitions of the geometrical sequences. This allowed us to determine if they represented sequences as nested combinations of primitive rules. To do so, we trained a decoder to predict the primitive participants used to encode the sequence segments, i.e. the transitions between two consecutive sequence items. Even when controlling for visual features, this decoder had an above chance performance, indicating that participants used the postulated primitives to represent transitions in the sequence context. Previous behavioral results indicate that these geometrical primitives are present even in young children and adults with limited access to formal schooling in mathematics (Amalric et al., 2017). They participate in an intuitive system of mental geometry which has been termed “core knowledge” and also comprises basic capacities for spatial navigation, shape recognition, topology and arithmetic (Dehaene, Izard, Pica, & Spelke, 2006; Spelke, Lee, & Izard, 2010). Our primitive decoding results on sequences show that, on top of the Euclidean concepts of distance, angle or sense, the human brain is highly sensitive to higher-level concepts of rotation and symmetry.

To access the representation of the sequence internal structure, a previous study (Wang et al., 2019) asked the participants to overtly anticipate with the eyes the sequence

items. In the current study, in order to avoid MEG artifacts, participants were asked to fixate a central cross while viewing the sequence items, which precluded a measure of overt attention. Instead, we looked for covert anticipation of sequence items. To do so, we trained a localizer decoder on a set of trials where no prediction was possible and tested it on the sequence part of the experiment. Decoding performance was above chance even before sequence items were presented. Contributions of eye movements to the anticipation signals could be excluded (See supplementary materials and figure S1). Moreover, similarly to the results obtained in (Wang et al., 2019) on gaze anticipation, the anticipation score calculated from the decoding performance was modulated by sequence complexity, suggesting that expectation mechanisms are engaged in the task in direct proportion to the sequence simplicity, given by the compactness of its internal description in the proposed language.

The observed anticipations fit within the predictive-coding framework, which proposes that the brain constantly projects stimulus-specific templates in advance of the stimulus itself, as previously found for instance during associative learning tasks (Albright, 2012; Kok, Mostert, & de Lange, 2017; Sakai & Miyashita, 1991), or during sequences of spatial locations with no structure (Ekman, Kok, & De Lange, 2017). Our own previous behavioral studies showed that human participants could anticipate the geometrical sequence items by overt pointing or looking to their spatial location before they were actually presented, and that such anticipation varied according to the hierarchical structure of the sequences (Amalric et al., 2017; Wang et al., 2019). Furthermore, in fMRI, sequence complexity and anticipation of the nested rules activated the IFG and Mid-DLPFC (Wang et al., 2019). These regions are plausible sources for the internal model of sequences that generates the observed top-down anticipations, and were shown to be engaged in the parsing mechanisms of spatial sequences (Bor, Duncan, Wiseman, & Owen, 2003; Desrochers, Chatham, & Badre, 2015) and in the representation of abstract rules (Badre, Kayser, & D'Esposito, 2010; Koechlin & Jubault, 2006). Human prefrontal cortex is functionally organized in a hierarchical way and represent rules of increasing abstractness (Badre, 2008; Badre & D'Esposito, 2009; Badre & Nee, 2018; Koechlin, Ody, & Kouneiher, 2003). Consistently with this description and with the cascade model for executive control, Wang et al. (2019) showed that saccade movements, sequence complexity and anticipation of the nested rules were represented respectively in the PMd area, in the IFG and in the Mid-DLPFC. We suggest that these regions may encode information about the sequence structure and send top-down signals in order to pre-activate circuits in posterior

cortices, as also observed in sequential memory for images (Tomita et al., 1999) and auditory patterns (Chao, Takaura, Wang, Fujii, & Dehaene, 2018).

The decoding of the primitives in the sequences and the modulation of the anticipation by sequence complexity suggest that participants represented the geometrical sequences in agreement with the proposed formal language of geometry, i.e. as nested combinations of elementary rules. The idea that humans encode complex concepts as nested combinations of a finite set of elementary primitives goes back at least to the 19th century (Boole, 1854) and has been termed the "language of thought" hypothesis (Fodor, 1975). It was initially supported by behavioral studies using classification tasks based on logical rules to investigate human rule learning abilities (Shepard, Hovland, & Jenkins, 1961). The results showed that task difficulty varied with the rule complexity, which was later formalized as the Boolean complexity of the rule, i.e. the length of its shortest logical expression as a formula with elementary logical operators (and, or), and parentheses (Feldman, 2000). Within this framework, the present experiment suggests that the human brain also encodes geometrical sequences by selecting the shortest expression that accounts for it and hence maximizes its compressibility (Chater & Vitányi, 2003; Feldman, 2000, 2003; M. Li & Vitányi, 1993; Romano et al., 2013).

Several observations suggest that the formal language of geometry that we proposed can be improved. An empirical measure of complexity, derived from participants' behavioral performance, departed slightly from the complexity measure provided by the formal language, and provided a better predictor of MEG anticipatory activity. This deviation from theory may arise simply from the fact that some primitives were more difficult to process than others. In an augmented version of the theory, weights could be assigned each primitive before computing sequence complexity (Romano et al., 2018).

Remarkably, the decoder of geometrical primitives, trained on the data from the primitive part, did not generalize to the sequence part, hinting at different primitive representations for the two tasks. This result could be due to peripheral factors such as the different timings of the two parts (figure 1), possible head position changes across the experiment (the primitive and the sequence part being at the beginning and end of the experiment). However, a major difference also exists between the block types: in the primitive part, a single primitive rule is considered at the time whereas in the sequence part, the primitives have to be nested to form more complex rules. Moreover, contrary to primitive pairs, sequences were repeated several

times. These repetitions facilitate memorization and may foster an integrated sequence representation, which may considerably enrich their internal representation, over and above the proposed minimal set of rotations and symmetries. For instance, subjects may encode sequences using visual aids. Even if vertices or edges are never explicitly displayed, lines may be mentally constructed linking the separate sequence items and further integrated to form shape percepts that may be encoded in early visual cortices (Gilbert & Li, 2013; Kok, Failing, & de Lange, 2014; W. Li, Piëch, & Gilbert, 2006; Roelfsema & Houtkamp, 2011). The complex shapes drawn from the full sequences may in turn be decomposed into elementary shapes such as squares, circles, triangles or rectangles (Leyton, 2001). A more complete version of the language of geometry should probably comprise these two types of representations: a visual one in terms of the shapes the sequences form, and an abstract linguistic one provided by the formal language.

Acknowledgements

We acknowledge help from all the NeuroSpin support teams, and particularly Leila Azizi and Virginie van Wassenhove. We are grateful to Maxime Maheu, Pedro Pinheiro-Chagas and Darinka Trübutschek for their valuable comments. This research was supported by INSERM, CEA, Collège de France, the Bettencourt-Schueller Foundation and an ERC grant "Neurosyntax" to SD. It was performed on a platform of the France Life Imaging network, partly funded by Agence Nationale de la Recherche grant *ANR-11-INBS-0006,* and supported by the Bettencourt-Schueller Foundation and the Leducq Foundation.

Methods

Participants

20 participants (9 men; M_{age} = 24.6 years, SD_{age} = 3.7 years) with normal vision were included in the MEG experiment. In compliance with institutional guidelines, all subjects gave written informed consent prior to enrollment and received 90€ as compensation.

Experimental protocol

General structure of the experiment

The main task, completed in the MEG Elekta acquisition device, was subdivided into 3 parts. To avoid biasing subjects towards specific primitives, the sequence part was performed first. The second part was dedicated to each of the primitive operations, and the third part was a localizer task with unpredictable locations, designed to train a decoder for spatial locations. During the 3 parts of the MEG experiment, white dots were flashed for 100ms on the vertices of an octagon while the subject was fixating a cross at the center of the screen. The MEG experiment was preceded by a short training (c.a. 20 minutes) to the geometrical sequences.

Training on sequences

As initial training, outside the MEG, the geometrical sequences were presented with a slower pace than the rest of the experiment: a stimulus onset asynchrony (SOA) of 700ms between consecutive dots, and a dot duration of 200ms. Each sequence was repeated until the participants pressed the space bar to report that they had memorized it. They were then asked to type in the 8 locations that followed the last item displayed on the screen. If they succeeded, the word 'Bravo' (congratulations) was presented on the screen and the next sequence started. If they failed, the word 'Erreur' (error) was displayed and the same sequence restarted. Participants were instructed that the same sequences would be presented to them during the main experiment. No training was provided on the primitive part of the experiment. The training part was meant to select participants able to quickly encode the geometrical sequences. Only the ones that had finished the training in less than 20 minutes were qualified for the main MEG experiment. 5 out of 25 participants did not manage to do the training part of the experiment in less than 20 minutes.

MEG task

Geometrical sequences

The first part of the experiment was devoted to the geometrical sequences and was composed of 4 runs. 9 sequences (figure 1B) were composed of 8 non-repeating locations. The last 3, called 'Memory sequences', which were composed of 1, 2 or 4 spatial locations, were not analyzed in this study. During one run, each sequence of 8 locations was presented 12 times consecutively. The sequences presented in figure 1B were mere templates: to generate the actual sequences, the starting point and global direction of rotation were varied and balanced across runs. 4segment sequences were selected such that each of the 4 symmetry axes (horizontal, vertical and diagonal) would appear once. Participants had to perform two tasks. First, they had to report with a button press when they had identified the sequence and felt able to predict the next locations. Second, during the 11th or the 12th repetition, an item appeared at an unexpected location, and subjects had to report this violation with another button press as fast as possible. This task was added to maintain participants attention during the full block.

Primitive part

The second part of the experiment was devoted to the primitive operations. It was composed of 4 runs, subdivided in 12 mini-blocks for each of the 12 conditions. The 11 first ones followed elementary 'primitive' rules (figure 1C), because a simple geometrical operation allowed to determine the spatial location of the second item of a pair when given the first. The 12th condition was a control condition in which no minimal rule allowed to do so, and participants could only memorize the 8 unrelated pairs in order to perform the task. A mini-block was composed of 32 pairs with a SOA of 433ms between the items of the pair and an inter-pair-interval of 1100ms. Each of the 8 pairs appearing 4 times in the mini-block.

The task was similar to the sequence part. Participants reported with a button press when they had identified the rule that allowed them to predict the location of the second item given the first. In addition, they had to detect as fast as possible when the second item did not appear at the expected location. Violations could only occur during the presentation of the last 8 items of the mini-block.

Localizer part

The last part of the experiment was meant to train a spatial decoder for spatial position of the presented items. To do so, dots were flashed pseudo-randomly on the vertices of the octagon with SOA 433ms. Occasionally (1/20 dots on average) the color of the dot changed. The subject had to click as fast as possible to report this.

MEG acquisition and preprocessing

MEG recordings

Participants performed the tasks 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). 102 triplets, each comprising one magnetometer and two orthogonal planar gradiometers composed the MEG helmet. The brain signals were acquired at a sampling rate of 1000 Hz with a hardware highpass filter at 0.03Hz.

Eye movements and heartbeats were monitored with vertical and horizontal electro-oculograms (EOGs) and electrocardiograms (ECGs). 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.

Data cleaning : Maxfiltering

Bad MEG channels were identified visually in the raw signal and were provided to the MaxFilter software (ElektaNeuromag®, Helsinki, Finland) to compensate for head movements between experimental blocks by realigning all data to an average head position and to apply the signal space separation algorithm (Taulu, Kajola, & Simola, 2004) to suppress magnetic interference from outside the sensor helmet and interpolate bad channels.

Data cleaning: ICA

The rest of the analysis was performed with MNE Python (Gramfort, 2013; Jas et al., 2018). Oculomotor and cardiac artefacts were removed performing an independent component analysis (ICA). The components that correlated the most with the EOG and ECG signals were

automatically detected. We then visually inspected their topography and correlation to the ECG and EOG time series to confirm their rejection from the MEG data.

Multivariate Pattern Analysis

The goal of the Multivariate Pattern Analysis (MVPA) decoding analyses was to predict from single-trial brain activity (*X*) a specific categorical (e.g. primitive identity) or continuous (e.g. angular position) variable (*y*) that represents the neuronal state corresponding to the participant's mental representation. These analyses were performed following King et al's preprocessing pipeline (King & Dehaene, 2014; King, Pescetelli, & Dehaene, 2016) implemented in MNE-python version 0.16 (Gramfort, 2013). Prior to model fitting, each channel at each time-point was z-scored across trials. Each estimator was fitted on each participant separately, across all MEG sensors using the parameters set to their default values provided by the Scikit-Learn package (Pedregosa et al., 2011). When the estimator was trained and tested on two different conditions, the whole training and testing sets were used to respectively fit and test the estimator. By constrast, when the decoder was trained and tested on non-independent data, we used a 5-folds stratified cross-validation procedure to prevent overfitting. The estimator was trained on 80% of the trials and tested on the remaining 20%. This operation was repeated 5 times. The reported scores are the average across cross-validation folds.

Decoding of angular position

The spatial angular decoders were built from two ridge regressions used to decode the angular position Θ. One predicted sin(Θ) and the other cos(Θ). The angular decoding score was obtained by first computing the mean absolute difference between the predicted angle (Θ_{pred}) and the true angle (Θ_{true}). We subtracted to this score π/2 to obtain a score in the range of - π/2 and π/2 (chance = 0) (King et al., 2016). When we wanted to determine which of the 8 locations was predicted by the angular decoder, we binned its output into 8 evenly spaced angular bins.

To access the temporal organization of the neural representations, we computed the temporal generalization matrices. These matrices represent the decoding score of an

estimator trained at time t (training time on the vertical axis) and tested with data from another time t' (testing time on the horizontal axis).

Anticipation score

The anticipation score is a quantity that allows to determine if the location of a sequence item can be decoded from the brain signals before it was actually displayed. It is defined as the subtraction of the number of times the correct position was predicted in advance, and the number of times the other position at the same distance from the preceding item was predicted (figure S1). This measure was designed to overcome a potential confound that comes from the fact that successive sequence items tend to be close to each other (average angle between two sequence items is 73°, s.d. = 28°). As the angular decoder output spreads on neighboring angles, it may predict above chance the spatial position of the next item. The anticipation score was not defined when the next item is at distance 4 from the previous one. As this represents 50% of the transitions for 4diagonals and 2crosses sequences, we excluded these sequences from the linear regression anticipation analyses that we ran across participants.

Decoding of the primitive identity

We used Support Vector Machine classifiers to decode primitive identity. When controlling for visual confounds, rotation±2 primitive was excluded from the trial set as no symmetries involved a distance of 2. Moreover, when only two categories were considered (e.g. 'rotation' and 'symmetry'), the score of the decoder was provided in terms of the area under the curve. When decoding the 11 primitives, we used One-VS-rest multiclass classifiers and report their mean accuracy.

Statistical analyses

All statistics reported in the text refer to group-level analyses. Tukey post-hoc tests were performed using the R software packages *nlme* and *multcomp*. We used multiple comparison corrections to assess multivariate decoding performance for the two-dimensional time by time generalization score matrices and for the simple decoding score time-courses. We considered temporal clusters and non-parametric one-sample *t*-tests estimated on 4096

permutations (Maris & Oostenveld, 2007) implemented by the *permutation_cluster_1samp_*test available in *mne.stats* package. A cluster was defined by adjacent time points. The cluster-level statistic was the sum of the sample-specific *t*-statistics that belonged to a given cluster. The alpha level of the sample-specific test statistic and of the cluster-specific test statistic were 0.05. For decoding performance curves, '*' and '**' indicate that resulting p-values are respectively < 0.05 and < 0.01. Dashed contours on temporal generalization correspond to p-value < 0.05 resulting from the permutation test. The student t-tests and the linear regressions were performed with Matlab®.

Supplementary material

Minimal description and hierarchical structure of the geometrical sequences

The proposed language of geometry includes instructions for repeating operations and for more complex repetitions with variations. For example, the sequence [+0, [+1]^7] starts from the point 0 and repeats 7 times the operation rotation+1, thus describing a circle around the octagon. The sequence [[+0, H]^4{−1}] builds a horizontal segment by applying the horizontal symmetry to 0, then iterates it four times, each time shifting the starting point according to rotation-1 – thus describing a zigzag or 4-segment sequence. The sequence [[+0, P]^4 ⟨+1⟩] builds a diagonal from the point symmetry P connecting position 0 to 4, and then iterates it four times, shifting the two elements by rotation+1 to build each of the consecutive segments forming a star-shaped or 4-diagonals sequence. Further information on the language of geometry is provided in the supplementary material of (Amalric et al., 2017). Table S1 provides a minimal description of each of the 9 sequences under study. The * sign indicates that several same-length descriptions exist for the considered sequence, depending on starting point, rotation direction and/or symmetry axis.

Assessing a possible contribution of eye movements to spatial decoding

During the MEG acquisition, participants were instructed to fixate the central cross while the items were flashed. Their gaze was monitored online to make sure they did so. However, involuntary eye movements may contaminate MEG recordings and lead to decodable non-brain signals (Mostert et al., 2018; Quax, Dijkstra, van Staveren, Bosch, & van Gerven, 2019). We assessed if the spatial decoding and the anticipation results were partly due to participants eye movements during the sequence presentation. We conducted control analyses on the participants for whom we could collect eye tracking data (14 out of 20).

We first measured gaze position during the sequence trials to assess its distance to the central fixation cross. Figure S1B shows the gaze heatmap overlaid onto the fixation cross and the 8 octagon vertices presented to the participants. For each participant separately, we computed the standard deviation of the distribution of distances to the fixation cross. The mean standard

deviation was 38 pixels (SD = 23 pixels), which was larger than the fixation cross (30 pixels) but much smaller than the octagon radius (225 pixels). This confirmed that participants did not saccade to the sequence items.

Moreover, the performance of a position decoder trained on the eye tracking and EOG data was much lower than the one of a decoder trained on the MEG data (figure S1A and S1D). We conclude that, even if we cannot fully exclude that the late positive performance (from 220ms on) may correspond to eye movements and micro-saccades, they do not play a dominant role for the brain decoding results.

Finally, we ran a mutual information analysis to determine how much information was shared between the MEG decoding results and eye movement decoding results (figure S1C). The mutual information was calculated over the correctly classified MEG trials, therefore providing a measure of the amount of information provided by the eye movement decoding results about the MEG decoding results (Quax et al., 2019). Remarkably, the mutual information was lower in the time interval corresponding to the maximal performance of the MEG data decoder, suggesting that during that time window the MEG data decoder based its predictions on brain rather than eye signals.

SEQUENCE		COMPLEXITY	MINIMAL DESCRIPTION	DESCRIPTION	HIERARCHICAL STRUCTURE
REPEAT		5	$[+0, [+1]^7]$ *	**Rotation±1** is repeated. Simple clockwise or counterclockwise progression.	No nesting, mere repetition of primitives
ALTERNATE		7	$[[+0, +3]\ ^4\langle+2\rangle]$ *	2 primitives in opposite directions are alternatively applied: **rotation±3** and **rotation∓1**.	
4SEGMENTS		7	$[[+0, H]^4\{-1\}]$	The first segment is built from **axial symmetry**. It is translated four times shifting its starting point.	Two embedded levels of regularity First level where instructions build a geometrical shape Second level where the shape is repeated with a global transformation
4DIAGONALS		7	$[[+0, P\ ]^4\ \langle+1\rangle]$ *	**Rotational symmetry P** is repeatedly applied to four consecutive starting points.	
2CROSSES		7	$[[+0, P\ ]^4\ \langle-3\rangle]$ *	**Rotational symmetry P** is repeatedly applied to points shifted by rotation±3.	
2ARCS		8	$[[+0, [+1]^3]^2\langle V\rangle]$*	**Rotation±1** is applied 3 times. Global **axial symmetry** to determine the 4 remaining positions.	
2SQUARES		8	$[[+0, [+2]^3]^2\langle-1\rangle]$*	**Rotation±2** is applied 3 times then again 3 times shifting the starting point by **rotation±1**.	
2RECTANGLES		10	$[[[+0, V]^2\ \langle P\rangle]^2\ \langle-2\rangle]$*	A **rotational symmetry P** is applied to the segment, tracing a rectangle. A global **rotation±2** generates the 4 remaining points.	3 embedded levels of regularity
IRREGULAR		16	[+0,P,+3,+2,P,A,V,H]	Mere list of the successive transitions between locations.	No nesting

Table S1: Description of the geometrical sequences

The table presents one of the descriptions (*) or the unique description provided by the language of geometry for all the sequences we considered in the study. A verbal description of the sequence structure is also provided.

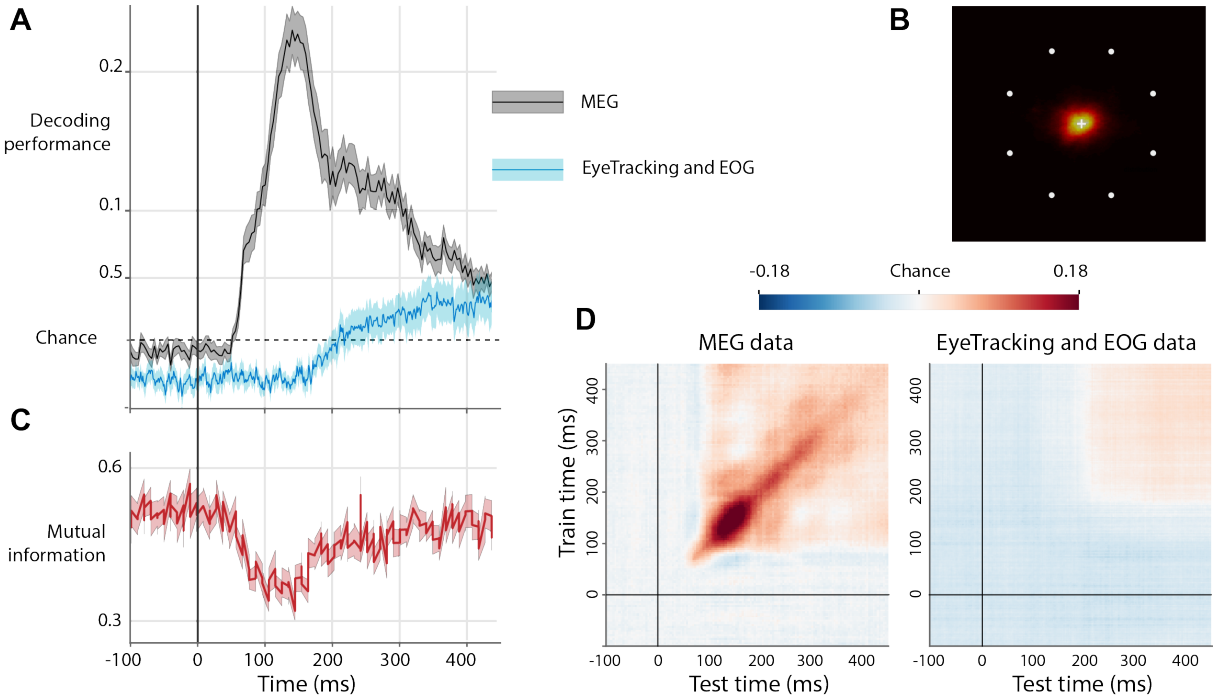

Figure S1: Eye movements do not account for brain decoding of stimulus location

A. Performance in decoding stimulus location from MEG data (black) and eye tracking and EOG data (blue) as a function of time following the flash of a dot at a given location. Decoding from MEG data was much better and significantly positive much earlier than decoding from eye tracking and EOG data.
B. The gaze heatmap during the sequence blocks confirms that participants did not saccade to the sequence items. Instead, in agreement with the instructions, they maintained the gaze on the central fixation cross.
C. When the MEG decoder had maximal performance, the mutual information, which quantifies the amount of information provided by the eye movements about the MEG decoding results, was reduced. This confirms that during that time window, the MEG data decoder based its predictions on brain rather than eye data.
D. Average generalization across time matrix showing the location decoding score for MEG data (left) and eye tracking and EOG data (right) as a function of training times (y axis) and testing times (x axis). The score for MEG data is much larger and significantly positive much earlier than the one for eye tracking and EOG data.

Bibliography

Albright, T. D. (2012). On the Perception of Probable Things: Neural Substrates of Associative Memory, Imagery, and Perception. *Neuron*. https://doi.org/10.1016/j.neuron.2012.04.001

Amalric, M., & Dehaene, S. (2017). Cortical circuits for mathematical knowledge: evidence for a major subdivision within the brain's semantic networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1740), 20160515. https://doi.org/10.1098/rstb.2016.0515

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 Computational Biology*, *13*(1). https://doi.org/10.1371/journal.pcbi.1005273

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2008.02.004

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn2667

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron*. https://doi.org/10.1016/j.neuron.2010.03.025

Badre, D., & Nee, D. E. (2018). Frontal Cortex and the Hierarchical Control of Behavior. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2017.11.005

Boole, G. (1854). *An investigation of the laws of thought: on which are founded the mathematical theories of logic and probabilities*. Dover Publications, Inc. https://doi.org/https://doi.org/10.1017/CBO9780511693090

Bor, D., Duncan, J., Wiseman, R. J., & Owen, A. M. (2003). Encoding Strategies Dissociate Prefrontal Activity from Working Memory Demand. *Neuron*, *37*(2), 361–367. https://doi.org/10.1016/S0896-6273(02)01171-6

Chao, Z. C., Takaura, K., Wang, L., Fujii, N., & Dehaene, S. (2018). Large-Scale Cortical Networks for Hierarchical Prediction and Prediction Error in the Primate Brain. *Neuron*. https://doi.org/10.1016/j.neuron.2018.10.004

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(02)00005-0

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science*. https://doi.org/10.1126/science.1121739

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron*. https://doi.org/10.1016/j.neuron.2015.09.019

Desrochers, T. M., Chatham, C. H., & Badre, D. (2015). The Necessity of Rostrolateral Prefrontal Cortex for Higher-Level Sequential Behavior. *Neuron*. https://doi.org/10.1016/j.neuron.2015.08.026

Ekman, M., Kok, P., & De Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*. https://doi.org/10.1038/ncomms15276

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*. https://doi.org/10.1038/35036586

Feldman, J. (2003). The Simplicity Principle in Human Concept Learning. *Current Directions in Psychological Science*. https://doi.org/10.1046/j.0963-7214.2003.01267.x

Fitch, W. T. (2004). Computational Constraints on Syntactic Processing in a Nonhuman Primate. *Science*, *303*(5656), 377–380. https://doi.org/10.1126/science.1089401

Fitch, W. T. (2014, September 1). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*. https://doi.org/10.1016/j.plrev.2014.04.005

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn3476

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). Neuroscience: The faculty of language: What is it, who has it, and how did it evolve? *Science*. https://doi.org/10.1126/science.298.5598.1569

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3). https://doi.org/10.1016/S0010-0277(00)00132-3

Jas, M., Larson, E., Engemann, D. A., Leppäkangas, J., Taulu, S., Hämäläinen, M., & Gramfort, A. (2018). A Reproducible MEG/EEG Group Study With the MNE Software: Recommendations, Quality Assessments, and Good Practices. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00530

Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S., & Wang, L. (2018). Production of Supra-regular Spatial Sequences by Macaque Monkeys. *Current Biology*. https://doi.org/10.1016/j.cub.2018.04.047

Kabdebon, C., Pena, M., Buiatti, M., & Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language*. https://doi.org/10.1016/j.bandl.2015.03.005

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/J.TICS.2014.01.002

King, J. R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron*, *92*(5), 1122–1134. https://doi.org/10.1016/j.neuron.2016.10.051

Koechlin, E., & Jubault, T. (2006). Broca's Area and the Hierarchical Organization of Human Behavior. *Neuron*, *50*(6), 963–974.

https://doi.org/https://doi.org/10.1016/j.neuron.2006.05.017

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*. https://doi.org/10.1126/science.1088545

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554. https://doi.org/10.1162/jocn_a_00562

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1705652114

Leyton, M. (2001). *A generative theory of shape*. Springer.

Li, M., & Vitányi, P. (1993). *An Introduction to Kolmogorov Complexity and Its Applications*. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer New York. https://doi.org/10.1007/978-1-4757-3860-5

Li, W., Piëch, V., & Gilbert, C. D. (2006). Contour Saliency in Primary Visual Cortex. *Neuron*, *50*(6), 951–962. https://doi.org/10.1016/j.neuron.2006.04.035

Marcus, G. F., Vijayan, S., Bandi Rao, S., Vishton, P. M., Saffran, J., Aslin, R., … Holyoak, K. J. (1999). Rule learning by seven-month-old infants. *Science (New York, N.Y.)*, *283*(5398), 77–80. https://doi.org/10.1126/science.283.5398.77

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*. https://doi.org/10.1016/j.jneumeth.2007.03.024

Maruyama, M., Pallier, C., Jobert, A., Sigman, M., & Dehaene, S. (2012). The cortical representation of simple mathematical expressions. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2012.04.020

Milne, A. E., Mueller, J. L., Mannel, C., Attaheri, A., Friederici, A. D., & Petkov, C. I. (2016). Evolutionary origins of non-adjacent sequence processing in primate brain potentials. *Scientific Reports*, *6*, 36259. https://doi.org/10.1038/srep36259

Milne, A. E., Petkov, C. I., & Wilson, B. (2018). Auditory and Visual Sequence Learning in Humans and Monkeys using an Artificial Grammar Learning Paradigm. *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2017.06.059

Mostert, P., Albers, A. M., Brinkman, L., Todorova, L., Kok, P., & de Lange, F. P. (2018). Eye movement-related confounds in neural decoding of visual working memory representations. *ENeuro*. https://doi.org/10.1523/ENEURO.0401-17.2018

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., … Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1701590114

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*(2), 85–117. https://doi.org/10.1016/J.COGPSYCH.2003.12.002

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*.

Quax, S. C., Dijkstra, N., van Staveren, M. J., Bosch, S. E., & van Gerven, M. A. J. (2019). Eye movements explain decodability during perception and cued attention in MEG. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2019.03.069

Ravignani, A., Sonnweber, R. S., Stobbe, N., & Fitch, W. T. (2013). Action at a distance: Dependency sensitivity in a New World primate. *Biology Letters*. https://doi.org/10.1098/rsbl.2013.0852

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Perception, and Psychophysics*, *73*(8), 2542–2572. https://doi.org/10.3758/s13414-011-0200-0

Romano, S., Salles, A., Amalric, M., Dehaene, S., Sigman, M., & Figueira, S. (2018). Bayesian validation of grammar productions for the language of thought. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0200420

Romano, S., Sigman, M., & Figueira, S. (2013). LT2C2: A language of thought with Turing-computable Kolmogorov complexity. *Papers in Physics*, *5*. https://doi.org/http://dx.doi.org/10.4279/PIP.05000

Saffran, J. R. (2003). Statistical Language Learning. *Current Directions in Psychological Science*, *12*(4), 110–114. https://doi.org/10.1111/1467-8721.01243

Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature*, *354*(6349), 152–155. https://doi.org/10.1038/354152a0

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*. https://doi.org/10.1037/h0093825

Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition*. https://doi.org/10.1007/s10071-015-0840-x

Spelke, E., Lee, S. A., & Izard, V. (2010). Beyond core knowledge: Natural geometry. *Cognitive Science*. https://doi.org/10.1111/j.1551-6709.2010.01110.x

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Taulu, S., Kajola, M., & Simola, J. (2004). Suppression of interference and artifacts by the signal space separation method. *Brain Topography*, *16*(4), 269–275. https://doi.org/10.1023/B:BRAT.0000032864.93890.f9

Tomita, H., Ohbayashi, M. M., Nakahara, K., Nature, I. H.-, 1999, undefined, Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, *401*(6754), 699–703. https://doi.org/10.1038/44372

Varley, R. a, Klessinger, N. J. C., Romanowski, C. a J., & Siegal, M. (2005). Agrammatic but numerate. *Proceedings from the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0407470102

Wang, L., Amalric, M., Fang, W., Jiang, X., Pallier, C., Figueira, S., … Dehaene, S. (2019). Representation of spatial sequences using nested rules in human prefrontal cortex. *NeuroImage*, *186*, 245–255. https://doi.org/10.1016/J.NEUROIMAGE.2018.10.061

Wang, L., Uhrig, L., Jarraya, B., & Dehaene, S. (2015). Representation of Numerical and Sequential Patterns in Macaque and Human Brains. *Current Biology*, *25*(15), 1966–1974. https://doi.org/10.1016/j.cub.2015.06.035

Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved Sequence Processing in Primate Frontal Cortex. *Trends in Neurosciences*. https://doi.org/10.1016/j.tins.2016.11.004

Wilson, B., Slater, H., Kikuchi, Y., Milne, A. E., Marslen-Wilson, W. D., Smith, K., & Petkov, C. I. (2013). Auditory artificial grammar learning in Macaque and Marmoset monkeys. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.2414-13.2013

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*. https://doi.org/10.1037/0033-295X.114.2.245

Yildirim, I., & Jacobs, R. A. (2015). Learning multisensory representations for auditory-visual transfer of sequence category knowledge: a probabilistic language of thought approach. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-014-0734-y