

Current Biology

Production of Supra-regular Spatial Sequences by Macaque Monkeys

Highlights

- Monkeys can flexibly reproduce a spatial sequence in forward or mirror order
- Monkeys generalize the supra-regular mirror grammar to novel sequences
- Preschool children learn the mirror grammar much faster using a chunking strategy
- Human uniqueness may lie in the speed and strategy for sequence learning

Authors

Xinjian Jiang, Tenghai Long,
Weicong Cao, Junru Li,
Stanislas Dehaene, Liping Wang

Correspondence

liping.wang@ion.ac.cn

In Brief

Jiang, Long, et al. demonstrate that macaques can be trained to produce center-embedded spatial sequences. Preschool children quickly learn the grammar using chunking and geometric structure to compress the information. The human brain may possess additional computational devices to efficiently represent sequences during inductive learning.



Production of Supra-regular Spatial Sequences by Macaque Monkeys

Xinjian Jiang,^{1,2,5} Tenghai Long,^{1,5} Weicong Cao,^{1,2} Junru Li,¹ Stanislas Dehaene,^{3,4} and Liping Wang^{1,6,*}

¹Institute of Neuroscience, Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, 200031 Shanghai, China

²Key Laboratory of Brain Functional Genomics, Institute of Cognitive Neuroscience, School of Psychology and Cognitive Science, East China Normal University, 200062 Shanghai, China

³Collège de France, Paris, France

⁴Cognitive Neuroimaging Unit, CEA DSV/I2BM, INSERM, Université Paris Sud, Université Paris-Saclay, NeuroSpin Center, 91191 Gif-sur-Yvette, France

⁵These authors contributed equally

⁶Lead Contact

*Correspondence: liping.wang@ion.ac.cn

<https://doi.org/10.1016/j.cub.2018.04.047>

SUMMARY

Understanding and producing embedded sequences in language, music, or mathematics, is a central characteristic of our species. These domains are hypothesized to involve a human-specific competence for supra-regular grammars, which can generate embedded sequences that go beyond the regular sequences engendered by finite-state automata. However, is this capacity truly unique to humans? Using a production task, we show that macaque monkeys can be trained to produce time-symmetrical embedded spatial sequences whose formal description requires supra-regular grammars or, equivalently, a push-down stack automaton. Monkeys spontaneously generalized the learned grammar to novel sequences, including longer ones, and could generate hierarchical sequences formed by an embedding of two levels of abstract rules. Compared to monkeys, however, preschool children learned the grammars much faster using a chunking strategy. While supra-regular grammars are accessible to nonhuman primates through extensive training, human uniqueness may lie in the speed and learning strategy with which they are acquired.

INTRODUCTION

A major issue for cognitive neuroscience is determining how human computational capacities differ from those of other species [1–5]. Here, we investigate whether human and nonhuman primates can produce temporal sequences whose complexity approximates that found in human languages. There is much evidence that nonhuman animals can learn sequences based on statistical transition probabilities [6], chunking [7], ordinal knowledge [8], or algebraic patterns [9, 10]. However, attempts to teach them sequences with nested or recursive structures, which are characteristic of human languages, have mostly

been met with negative results [1]. So far, the generative algorithms acquired by animals seem mostly restricted to the lowest level of the Chomsky hierarchy [11–13]—that is, regular languages [14, 15]. Thus, it has often been proposed that a crucial distinction lies between the levels of regular or “finite-state” grammars, which are accessible to nonhuman animals, and supra-regular grammars or “phrase-structure” grammars, which may only be available to humans [2, 14–16]. Indeed, neuroimaging studies in humans have shown that the processing of embedded grammars specifically relates to left-hemispheric areas of the inferior frontal gyrus (“Broca’s area”) and the posterior superior temporal sulcus [1, 17–20], two regions whose size, connectivity, and lateralization are distinctly different in humans [21, 22].

Recently, several behavioral experiments suggested the successful acquisition of context-free, center-embedded grammars by songbirds and baboons [23–25]. However, these claims were subsequently criticized [26–29]. The main critique is that, given the indirect perceptual classification and novelty detection tasks used, instead of acquiring a full set of abstract grammatical rules, animals could have applied simpler strategies based on lower-level sensory features [15, 26, 29]. The gap between human grammar learning and nonhuman sequence learning therefore remains unclear, and so far, there is no evidence showing a genuine grasp of supra-regular grammars in nonhuman animals [30, 31].

RESULTS

Here, to sidestep those issues, we designed a novel behavioral paradigm, delayed-sequence reproduction task that required the animal to explicitly generate sequences according to the instructed grammars. We compared two grammars: (1) a “mirror” (context-free) grammar of the form ABC|CBA, which in formal language theory entails recursive time-symmetrical center embedding or, equivalently, a push-down stack, and (2) a “repeat” (context-sensitive) grammar of the form ABC|ABC, whose main feature (repetition in serial order) is described by formal language theory as involving tail recursion with cross-dependencies (Figure 1A). Both grammars require more computational power to parse than the regular languages generated by



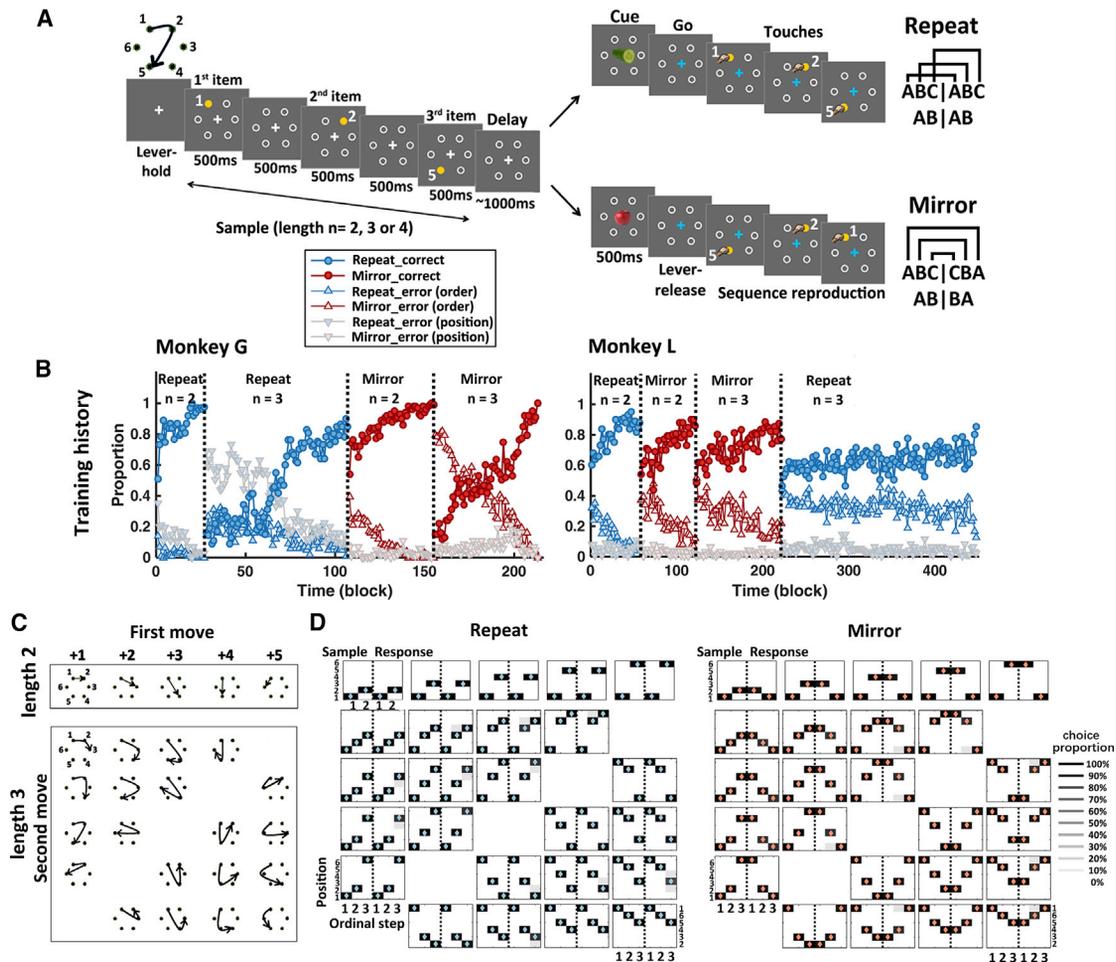


Figure 1. Acquisition of the Delayed-Sequence Reproduction Task

(A) Example trial with sequence “125.” Monkeys initiate the trial by pulling a lever. A series of two, three, or four sample stimuli (flashing colored dots), chosen from the six spatial locations, is presented at a fixed rate (1/s). After a delay, a 500-ms visual cue is presented at the center of the screen to indicate which grammar (“repeat” or “mirror”) to use in the production phase. When the “go” cue appears, monkeys have to release the lever and touch the locations by either copying (repeat) or reversing (mirror) the sample sequence.

(B) Learning dynamics during the training period. Each block represents data averaged over 100 trials. Correct trials are depicted with solid blue or red dots, order errors with light blue or red triangles, and position errors with gray triangles.

(C) Faithful reproduction of all sequences of length 2 and 3 in forward and backward order. The dataset comprises a total of 150 sequences, but for simplicity, performance is averaged after normalization for the starting location and is displayed according to distance (clockwise) from the previous position. Thus, 25 sequence patterns are displayed with the horizontal axis indicating the distance of the first move (for length 2 and length 3 sequences) and the vertical axis denoting the distance of the second move (length 3 only).

(D) Gray level indicates the percentage of responses at a given location of each step averaged over two monkeys after the acquisition of grammars (see individual animals in Figure S2). Diamonds indicate the correct position. Vertical dashed lines mark the transition between the sample (visual presentation) and production (hand touch) phases.

See also Figures S1 and S2 and Video S1.

finite-state automata [13]. However, the capacity to repeat a sequence previously seen or heard has been frequently demonstrated—for instance, in birdsong [32, 33] or in primate behavior [34, 35]—and may simply indicate that imitation is a computational primitive in several species. We therefore concentrated our efforts on the mirror grammar, which exhibits an undeniable center-embedding supra-regular organization.

In our test, each sequence item—A, B, and C—could be one of six spatial locations, resulting in a large number of combinations (30 length 2 and 120 length 3 for a total of 150 sequences per grammar; each location was only sampled once, without

replacement, in a given sequence). On each trial, monkeys visually perceived the first half of the sequence (of length 2 or 3), then a cue indicating which grammar was required (mirror or repeat), and finally produced the second half by successively touching the appropriate locations on screen in order to get a reward (Figure 1A). Touching either at a wrong position (position error) or in a wrong order (order error) terminated the trial.

Two animals (monkeys G and L) learned to perform the sequence task with both grammars (Figure 1B; monkey G, correct rate [CR] > 92%; monkey L, CR > 82%; data averaged from the last three blocks in training; compared with

chance-level, chi-squared test, $p < 10^{-16}$ in all blocks). The training history was controlled, which slightly affected learning speed but not ultimate acquisition (Figure S1). The training set included only a fraction of the sequences and differed across monkeys. Monkey G learned the repeat grammar (both length 2 and 3 sequences) first, while monkey L learned the mirror grammar (both length 2 and 3 sequences) first. Both animals showed high accuracies across all ordinal positions (Figure S2).

Based on the sequential movements between spatial targets, the sequences on the hexagon were categorized into 25 patterns (Figure 1C). The results showed that both monkeys faithfully reproduced all spatial sequences of lengths 2 and 3 in forward and backward orders. Performance was above chance in every pattern at every ordinal position and without any significant bias for particular patterns in terms of both CRs (Figure 1D; for individual monkeys, see Figures S2E and S2G) and response times (RTs; inter-response latency) (Figures S2F and S2H). Position errors were rare and tended to lie near the correct target (Figure S2C). Order errors showed a trend toward transpositions of temporally close items within the target list (Figure S2D).

To further exclude the possibility that monkeys used alternative strategies to perform the task, we compared the animals' performance with the chance-level performance predicted by two alternative strategies: (1) remembering correctly all locations but picking them in random order (chance = 0.5 (1/2) for length 2, 0.167 (1/3!) for length 3, and 0.042 (1/4!) for length-4 sequences) or (2) producing correctly the first location, which is the most recent and therefore most likely to remain available in short-term memory, then picking the next locations in random order (chance = 0.5 for length 3 and 0.167 (1/3!) for length 4 sequences). We found that performance at each step was significantly higher than those two chance levels ($p < 10^{-4}$ in all blocks; Figure 1D), thus refuting those alternative strategies.

To exclude the possibility of rote memorization, we next focused on the crucial center-embedding mirror grammar and further tested the animals' generalization abilities. Generalization was evaluated by measuring performance on the first test day, which only involved trial-unique novel sequences. Nevertheless, all trials were rewarded, allowing behavioral improvement over successive blocks during generalization. The tests were terminated when performance reached 75% correct within a block (100 trials) or when animals no longer responded to the tasks. We first tested generalization to novel sequences that were never shown to animals during training. Both animals quickly transferred the grammar to the new sequences (Figure 2A; chi-squared test relative to chance, $p < 10^{-4}$ for each monkey in all blocks, including the first block of the first day).

Part of the power of a generative grammar is its ability to produce sequences of arbitrary length. To evaluate whether the monkeys generalized the mirror concept from AB|BA (length 2) and ABC|CBA (length 3) to longer sequences, we tested them with sequences generated by the grammar ABCD|DCBA (length 4) and ABCDE|EDCBA (length 5). The animals showed fast transfer to the longer sequences (Figures 2B and S3; for length 4 sequences tested in both monkeys, $p < 10^{-4}$ in all blocks compared with the chance level, and for length 5 sequences only tested in monkey G, $p < 10^{-4}$ in all blocks).

Thus far, we showed that monkeys were able to apply the grammars to the six fixed locations in a regular hexagon. We

then assessed whether monkeys generalized the rule to untrained geometrical layouts. The results showed that the monkeys could promptly generalize the grammar to locations in the form of a pyramid with the same number of targets (Figure 2C; for both monkeys, $p < 10^{-4}$ in all blocks), in a form with a random number and location of targets (Figure 2D; the form changed randomly from trial to trial for both monkeys, $p < 10^{-4}$ in all blocks), and in a horizontal line (Figure 2E; both monkeys, $p < 10^{-4}$ in all blocks). It is worth noting that during training, on the hexagon, three locations were never aligned.

Our explicit production task has a very low level of perfect performance by chance (1/120 for length 3 sequences). The monkeys' high performance level, generalizing across a variety of locations, implies that their abilities cannot be imputed to low-level shortcuts and involve a genuine understanding of the mirror task. At a minimum, to produce a sequence with such center-embedded structure, additional computational machinery beyond a finite-state automaton, such as a push-down memory stack, is required. We thus propose that macaque monkeys may possess such computational ability, sequentially storing each spatial location during the visual presentation and then retrieving them from the stack in a last-in-first-out manner.

Next, we explored whether the behavior could be spontaneously transferred across spatial locations, because such analogical reasoning is vital to advanced cognition and sometimes thought to be uniquely human [36]. In this test, the sample locations were visually displayed on a hexagon in one quadrant of the screen (e.g., 123; Figure 2F), and monkeys had to generate the corresponding sequence on another hexagon in a different quadrant (e.g., 1'2'3'). The task required animals to attend to an abstract structural feature, the local spatial relations of individual targets. Both monkeys showed rapid transfer of the grammar, with performance reaching 80.1% correct for monkey G and 45.3% for monkey L (Figure 2F; for both monkeys, $p < 10^{-4}$ in all blocks).

To test how fast the monkeys spontaneously generalized the mirror rule to novel sequences, the performance in the first 150 trials during each generalization was analyzed. We divided those 150 trials into 5 mini-blocks and found that the performance was significantly better than chance level in every mini-block (all $p < 10^{-4}$), and there was no significant learning effect for the first five mini-blocks (linear regression, $p > 0.1$) (Figure 3).

The above experiments could still be criticized, because they did not involve the full production of a center-embedded string. Rather, the experimenter provided the first half-string, and the monkey produced the second half-string in reverse order. We therefore tested whether monkeys could produce a complete center-embedded sequence. When given the sample of "ABC," monkeys were required to generate the sequence of "ABC CBA." The results showed that the two monkeys successfully learned to generate the full six-item sequence (monkey G: CR > 68%, monkey L: CR > 73%; in four successive blocks, $p < 10^{-4}$ in all blocks).

Finally, we tested the ability to generate an entire hierarchical sequence endogenously through a combination of two abstract rules: "next" and mirror [5]. One monkey (G) was trained to produce sequences based on two levels of embedding. In this task, using the standard hexagon, the monkey was first habituated to the concept of "next item" using the six length 3 mirror sequences (ABC|CBA) whose first three consecutive sample items

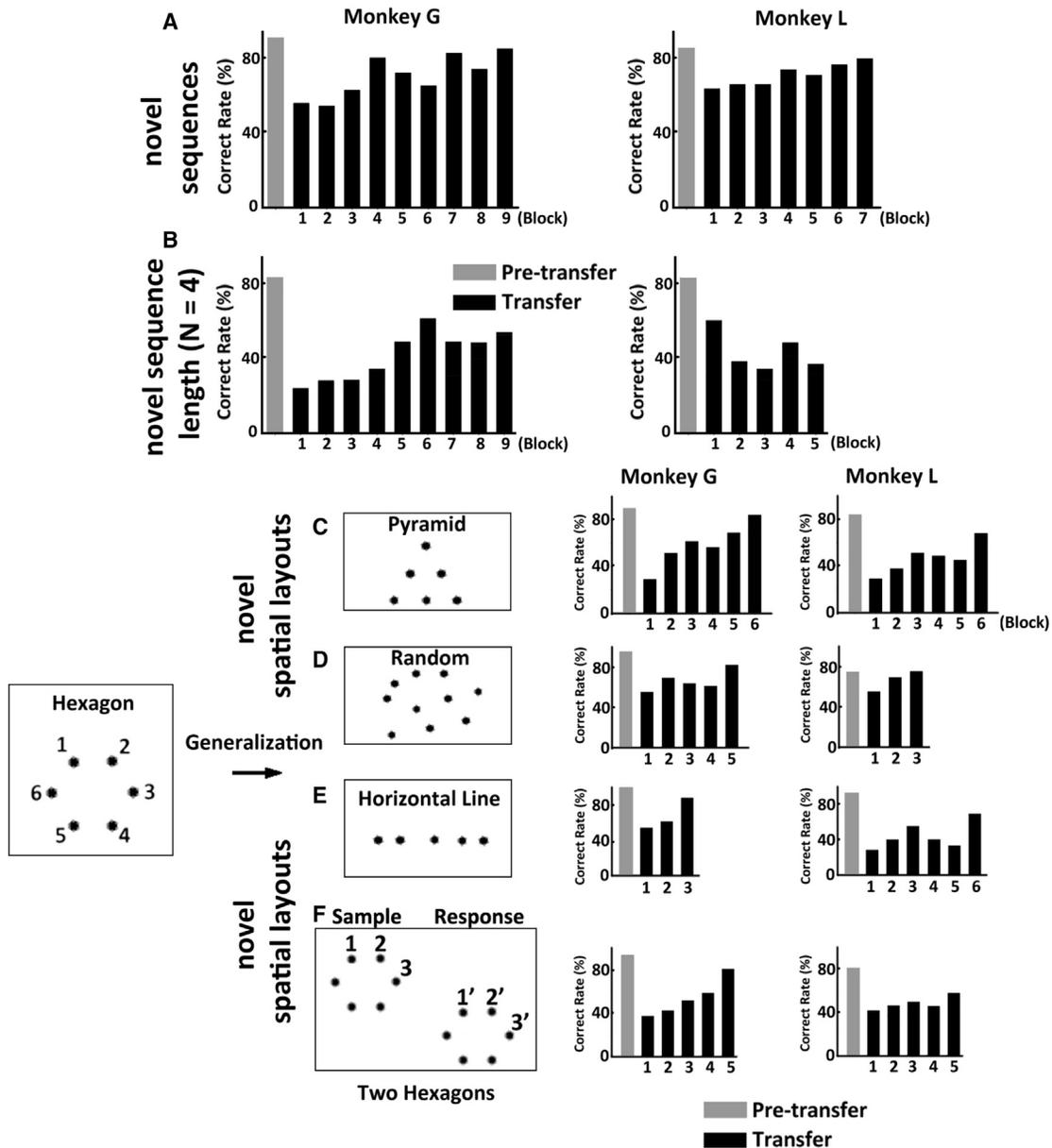


Figure 2. Generalization of the Mirror Grammar

Histograms show the success rate on the standard hexagon (in gray) before the transfer and during the transfer blocks (in black). Each bar represents performance averaged over 100 trials.

(A) Generalization to novel sequences on the standard hexagon.

(B) Generalization to longer length 4 sequences on the standard hexagon.

(C–F) Generalization to new geometrical shapes: (C) pyramid, (D) random locations, (E) horizontal line, and (F) two hexagons.

See also [Figure S3](#).

involve clockwise movement (e.g., 123, 234, 345). During the test, the monkey was presented with only the first position (randomly chosen from any of the six locations) in the sample and required to produce the remaining five locations, e.g; 1→23 321 ([Figure 4A](#)). What the monkey was asked for could thus be described as a “program” involving two embedded operations in order to generate, from a single starting point, a complex sequence, first comprising three consecutive locations (two applications of next), then the same locations in reverse order (one application

of mirror to the preceding string). The monkey successfully learned to generate the entire spatial sequence (chi-squared test, $p < 10^{-4}$ in all blocks versus chance level; [Figure 4B](#)) and promptly generalized to new locations in untrained geometrical shapes, such as octagon, square, irregular polygon, and two hexagons where the first and second half of the sequences were spatially separated ($p < 10^{-4}$ in all cases) ([Figure 4C](#)). The RTs on the second half of sequence (the fourth, fifth, and sixth target) were significantly lower than on the first half (second and

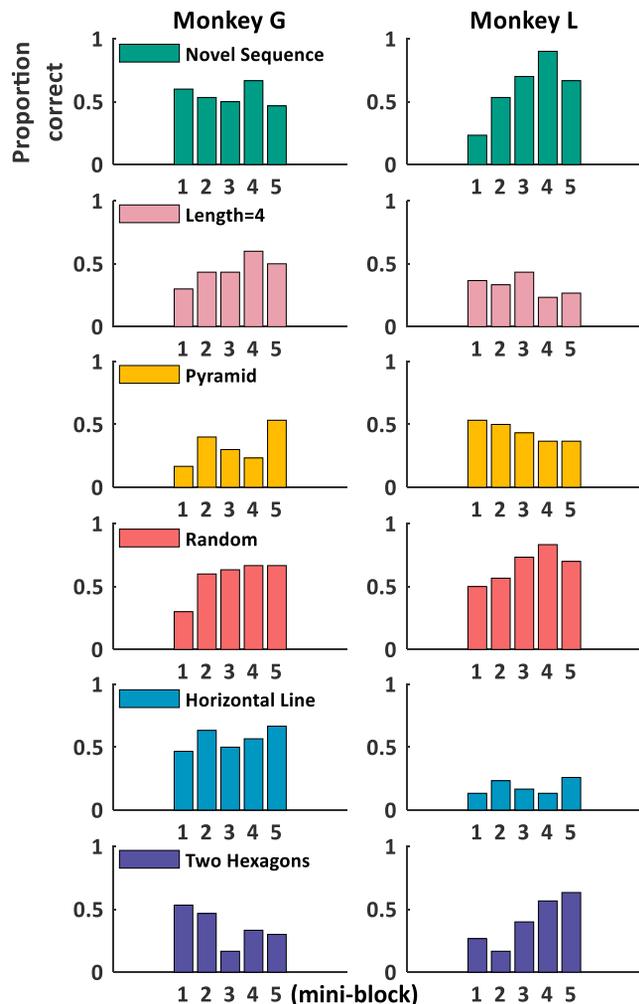


Figure 3. Fast Generalization of the Mirror Grammar to Novel Sequences

Histograms show performance in the first 150 trials during each generalization (Figure 2). Each bar represents performance averaged over 30 trials. The performance was significantly better than the chance level in every mini-block (chi-squared test relative to chance, all $p < 10^{-4}$), and there was no significant learning effect for the first five mini-blocks (linear regression, all $p > 0.1$).

third target) (Figure 4D; t test, $p < 0.001$, corrected), suggesting that, rather than using a serial search strategy [37], the monkey was performing the second half of the sequence using the previously learned mirror grammar.

Although monkeys' generalization was excellent (Figures 2 and 3), recall that they received thousands (10,000 to 25,000) of training trials before their performance ranged around 60% correct on length 4 sequences (Figure S3). To put this in perspective, we evaluated the grammar-learning abilities of 5- to 6-year-old preschool children using a very similar paradigm (Figure S4). Children vastly outperformed monkeys: most learned the repeat and mirror grammars after only five demonstrations, performed close to ceiling, and instantly generalized to various versions of novel stimuli (Figure S4).

Previous work suggested that human preschoolers and adults achieve such performance by taking advantage of the presence

of nested chunks of spatial locations [38]. We therefore compared the determinants of behavioral performance for length 4 sequences in monkeys and preschool children (Figure 5). The results revealed that monkey performance was primarily determined by a linear decrease with ordinal position (Figure 5A). Preschool children showed no such effect but rather gave evidence of chunking the sequence into two groups of two locations, with performance at the start of the second group decreasing as a function of distance to the preceding location (Figures 5B and 5D) and also when the trajectory of the second group crossed that of the first one (Figure 5F). Such a temporal and geometrical structuring of sequences was absent in monkeys (Figures 5C, 5E, and S5).

DISCUSSION

We demonstrated that macaque monkeys could explicitly produce center-embedded sequences whose description requires a supra-regular grammar. Various generalization tests, using untrained stimuli, excluded the possibility of mere memorization or retrieval of sequences from a look-up table. Monkeys spontaneously generalized to longer sequences, transposed spatial locations, and a new embedding. Most strikingly, by combining two abstract rules (next and mirror), one monkey could generate a complex sequence by starting from a single location and producing the next five locations in a purely endogenous manner.

Although rats [39], birds [40, 41], and several nonhuman primates [6, 42] may recognize statistical relationships and algebraic rules in patterned acoustic strings, their spontaneous production abilities (e.g., vocalizations or body movements) are limited and do not express such abstract regularities. In a recent review, we distinguished five levels of sequence knowledge with increasing degrees of abstraction: transition and timing knowledge, chunking, ordinal knowledge, algebraic patterns, and nested tree structures generated by symbolic rules [1]. Previous studies convincingly showed that animals could learn different types of sequences based on levels 1 to 4 [6–9]. Those levels, however, do not suffice to account for the most advanced form of sequence processing, namely human language. Even though a handful of animal experiments suggested the successful acquisition of language-like nested or center-embedded structures, these claims were judged premature and have been heavily criticized [26–29]. Compared to such prior sequence learning studies [43, 44], the present study presents several improvements: (1) the use of an explicit production task, which eliminates many of the shortcut strategies available to animals in perceptual classification and novelty detection tasks [29, 45], (2) generalization to novel configurations and sequence lengths (one of the hallmarks of generative grammar being its ability to process sequences of variable length), (3) transposition to a new location, demonstrating that behavior is based on an abstract rule rather than on shallow perceptual/motor properties of the stimuli, and (4) demonstration of the behavior in individual subjects, thus avoiding artifacts due to inter-subject averaging [45].

Our results question the view that the distinction between human and nonhuman animals, as regards the evolution of linguistic syntax, lies in the ability to learn regular versus supra-regular

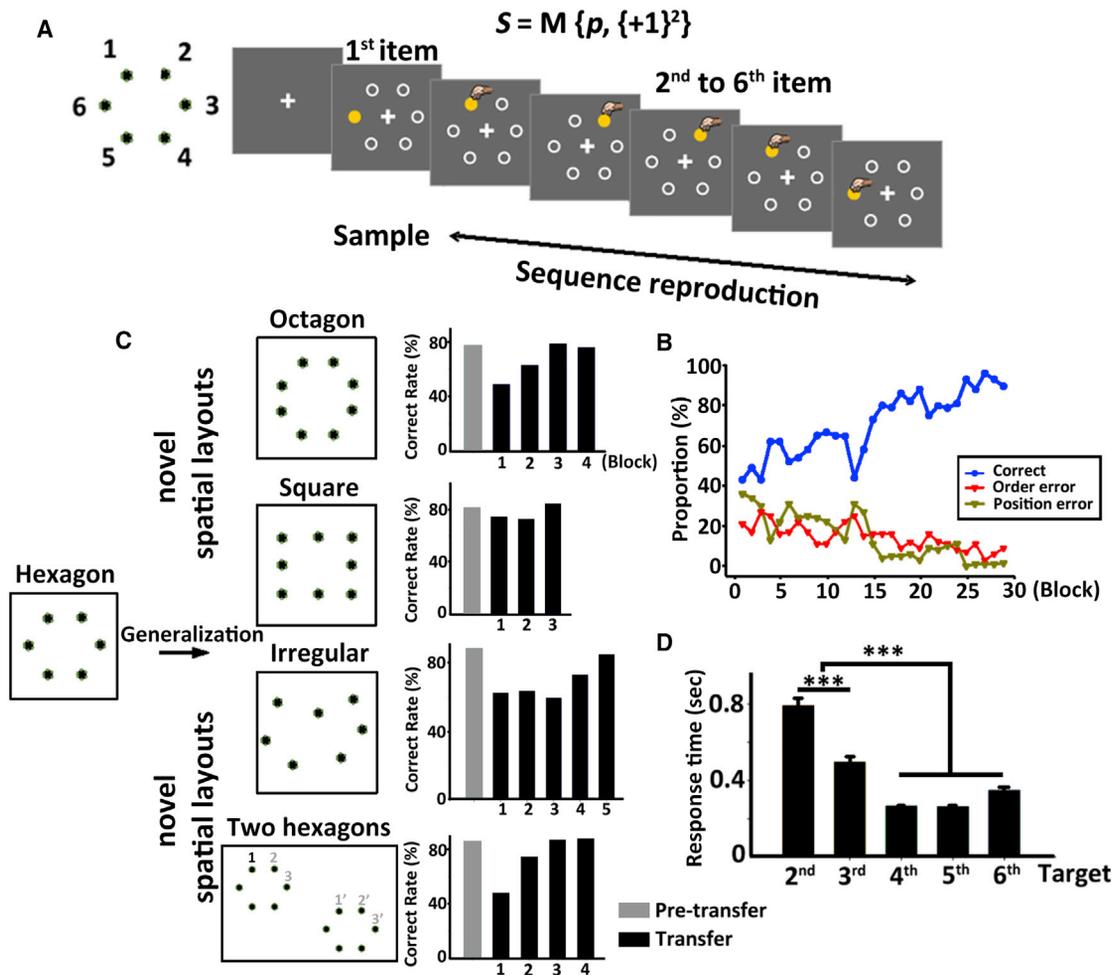


Figure 4. Combination of Abstract Rules and Grammar Generalization

(A) Task. On a given trial, a single item flashes for 500 ms at one of the six possible locations during the sample period. The monkey is then required to sequentially touch the remaining five positions based on the grammar $S = M \{p, \{+1\}^2\}$, where S is the generated sequence for a given starting position p , M indicates the global mirror transformation (“reverse”), and $+1$ denotes the local operation of clockwise movement (“next”).

(B) Learning dynamics. The monkey learned the grammar successfully, and performance reached 92% after 27 blocks ($p < 10^{-4}$ in all blocks).

(C) Generalization to novel sequences.

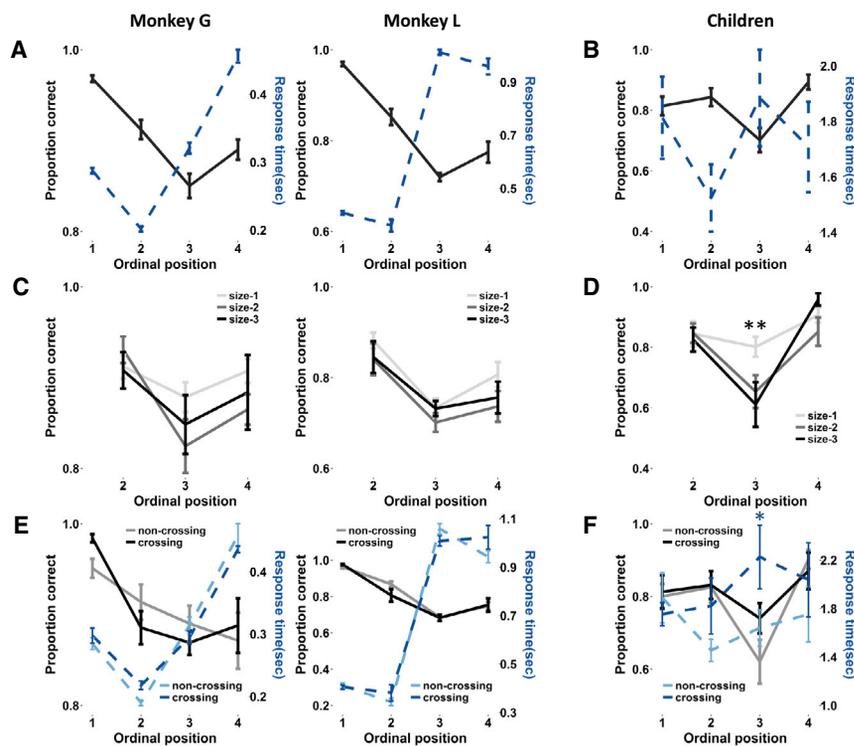
(D) Response time (RT, inter-response latency) for each consecutive target location in the hexagon condition (data from 2016 trials). RT to the third target was significantly lower than to the second target (t test, $p < 0.001$). RTs to second and third targets were significantly higher than to the fourth, fifth, and sixth targets (t test, $p < 0.001$, corrected). *** $p < 0.001$.

Error bar indicates one SEM.

(context-free) grammars in the Chomsky hierarchy [2, 13, 16, 30]. The boundary between human linguistic rule-learning abilities and those of nonhuman animals needs to be reconsidered and is less clear than is currently assumed [31]. Our findings also argue against the hypothesis that the combination of symbolic operations into a hierarchical mental program is a human-specific ability [5]. However, we acknowledge that we only tested two supra-regular grammars (repeat and mirror). Both may be accommodated by a simple “stack,” i.e., a memory that stores information in sequential order and can operate in two distinct retrieval modes, namely last-in-first-out (mirror) and first-in-first-out (repeat). Producing a sequence in forward order was previously demonstrated in nonhuman primates (e.g., [34, 35]), and our main contribution is to show that they can also flexibly

produce it in backward order. Whether monkeys can learn other more complex supra-regular grammars remains to be tested.

It is also crucial to keep in mind that, in the present study, unlike previous animal studies with spontaneous discrimination [16, 23], monkeys received intensive training. Indeed, this was required in order to explore animals’ fundamental computational limits [46]. Compared to monkeys, preschoolers quickly learned the same grammars within a handful of trials. This finding is comparable to recent neural-network simulations indicating that conventional neural networks for sequence learning can eventually learn a recursive grammar, but only after exposure to a large and complex training set, whereas tree-based neural networks learn much faster [47–49]. Even after intensive training on length 4 sequences, behavioral analysis suggested that monkeys still



(F and E) When there is a presence of crossing in the spatial sequence, children (F), but not monkeys (E), showed the significant longer RTs at the ordinal position 3 (linear mixed model: RT, main effect of crossing, $F(1,150.74) = 4.18$, $p < 0.05$; contrasts at the third position, $t = 2.20$, $p < 0.05$; accuracy: $p > 0.1$; other comparisons: $p > 0.1$, also see Figure S5). * $p < 0.05$; ** $p < 0.01$. Error bar indicates one SEM. See also Figures S4 and S5.

relied on a simple ordinal memory encoding, whereas preschoolers spontaneously used chunking and global geometric structure to compress the information. Thus, the human brain may possess additional computational devices, akin to a “language of thought,” to efficiently represent sequences using a compressed descriptor during inductive learning [4, 38, 50, 51].

Taken together, our findings suggest that the neurological architecture of the monkey brain places no fundamental obstacle to the ultimate learning of supra-regular structures. Human uniqueness may lie in the speed with which such structures are learned, perhaps using a specific structure-sensitive algorithm [4, 38, 50], rather than in the mere capacity to acquire them. By introducing a task easily learnable by monkeys yet presenting a significant similarity to language, the present study paves the way to future electrophysiological studies of center embedding. Recordings in the monkey homolog of Broca’s area, the inferior frontal gyrus [52], or the dorsolateral prefrontal cortex [9], where abstract sequence coding was previously found, should shed light on the neural code for center-embedded structures and how it differs between human and nonhuman primates.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE

Figure 5. Comparison of Length 4 Sequence Performance in Monkeys and Preschool Children in the Mirror Task

(A) Both monkeys showed a significant memory decay with ordinal position during sequence reproduction as indicated by increasing RT (dotted line) and decreasing accuracy (black line; linear model, $p < 10^{-5}$ for each monkey and each dependent variable). Both monkeys also displayed classical primacy and recency effects (“U” shape of correct rate).

(B) Children ($n = 23$) showed no such memory decay (linear model: accuracy, $p > 0.1$; RT, $p > 0.1$), but decreased performance at the third position, consistent with chunking into two groups of two (multiple comparisons with Bonferroni correction after linear mixed model: third versus second position accuracy: $t = -4.34$, $p < 0.001$; fourth versus third position accuracy: $t = 5.87$, $p < 10^{-4}$; RT, third versus second, $t = 2.62$, $p < 0.05$).

(C and D) Consistent with this analysis, children (D), but not monkeys (C), showed an effect of step size specifically between ordinal positions 2 and 3, thus affecting performance on the third target but not on the second and fourth target (step-size and ordinal-position interaction $F(4,172.65) = 3.55$, $p < 0.01$, contrasts at the third position: size 1 versus size 2, $t = 2.85$, $p < 0.05$; size 1 versus size 3, $t = 3.67$, $p < 0.001$; size 2 versus size 3, $t = 1.6$, $p > 0.1$; other comparisons: $p > 0.1$; RT: see Figures S5A and S5B).

- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Monkey
 - Preschooler
- METHOD DETAILS
 - Stimuli
 - Behavioral training
 - Grammar generalization in monkey G and monkey L
 - Grammar learning and generalization in preschool children
 - Performance of length-4 sequences in preschool children
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and one movie and can be found with this article online at <https://doi.org/10.1016/j.cub.2018.04.047>.

ACKNOWLEDGMENTS

The research was supported by the CAS Pioneer Hundreds of Talents Program, the Key Research Program of Frontier Sciences, CAS (QYZDY-SSW-SMC001), and the Shanghai Key Basic Research Project (16JC14202001) to L.W. S.D. was supported by Collège de France, Fondation Bettencourt-Schueller, and an ERC grant “NeuroSyntax.” We thank Christophe Pallier, Nai Ding, and Tianming Yang for their comments on the manuscript and Wen Fang and Jing Wu for experimental assistants. We also thank Mrs. Guofang Ren

(the general headmaster) from Far East Horizon Education Group for the help in the kindergarten class.

AUTHOR CONTRIBUTIONS

L.W., S.D., and X.J. designed the experiments; L.W. and S.D. wrote the paper; X.J., T.L., and W. C. performed the experiments; L.W., X.J., T.L., J.L., and W. C. analyzed the data.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 15, 2017

Revised: March 4, 2018

Accepted: April 16, 2018

Published: June 7, 2018

REFERENCES

- Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron* 88, 2–19.
- Hauser, M.D., Chomsky, N., and Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Matsuzawa, T. (2013). Evolution of the brain and social behavior in chimpanzees. *Curr. Opin. Neurobiol.* 23, 443–449.
- Yang, C. (2016). *The Price of Linguistic Productivity: How Children Learn to Break Rules of Language* (Cambridge, MA: MIT Press).
- Deacon, T.W. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain* (W. W. Norton & Company Ltd.).
- Wilson, B., Slater, H., Kikuchi, Y., Milne, A.E., Marslen-Wilson, W.D., Smith, K., and Petkov, C.I. (2013). Auditory artificial grammar learning in macaque and marmoset monkeys. *J. Neurosci.* 33, 18825–18835.
- Fujii, N., and Graybiel, A.M. (2003). Representation of action sequence boundaries by macaque prefrontal cortical neurons. *Science* 301, 1246–1249.
- Brannon, E.M., and Terrace, H.S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science* 282, 746–749.
- Shima, K., Isoda, M., Mushiake, H., and Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. *Nature* 445, 315–318.
- Wang, L., Uhrig, L., Jarraya, B., and Dehaene, S. (2015). Representation of numerical and sequential patterns in macaque and human brains. *Curr. Biol.* 25, 1966–1974.
- Chomsky, N. (1957). *Syntactic structures* (The Hague, Netherlands: Mouton).
- Chomsky, N. (1965). *Aspects of the theory of syntax* (Cambridge, MA: MIT Press).
- Jäger, G., and Rogers, J. (2012). Formal language theory: refining the Chomsky hierarchy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1956–1970.
- Fitch, W.T. (2014). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys. Life Rev.* 11, 329–364.
- Fitch, W.T., and Friederici, A.D. (2012). Artificial grammar learning meets formal language theory: an overview. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1933–1955.
- Fitch, W.T., and Hauser, M.D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.
- Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I., and Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. USA* 103, 2458–2463.
- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., and Weiller, C. (2003). Broca's area and the language instinct. *Nat. Neurosci.* 6, 774–781.
- Pallier, C., Devauchelle, A.D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. USA* 108, 2522–2527.
- Tyler, L.K., Marslen-Wilson, W.D., Randall, B., Wright, P., Devereux, B.J., Zhuang, J., Papoutsi, M., and Stamatakis, E.A. (2011). Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain* 134, 415–431.
- Neubert, F.X., Mars, R.B., Thomas, A.G., Sallet, J., and Rushworth, M.F. (2014). Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron* 81, 700–713.
- Rilling, J.K., Glasser, M.F., Preuss, T.M., Ma, X., Zhao, T., Hu, X., and Behrens, T.E. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.
- Abe, K., and Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* 14, 1067–1074.
- Gentner, T.Q., Fenn, K.M., Margoliash, D., and Nusbaum, H.C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.
- Rey, A., Perruchet, P., and Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: evidence from baboons (*Papio Papio*). *Cognition* 123, 180–184.
- Beckers, G.J.L., Berwick, R.C., Okanoya, K., and Bolhuis, J.J. (2017). What do animals learn in artificial grammar studies? *Neurosci. Biobehav. Rev.* 81, 238–246.
- ten Cate, C., and Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1984–1994.
- Poletiek, F.H., Fitz, H., and Bocanegra, B.R. (2016). What baboons can (not) tell us about natural language grammars. *Cognition* 151, 108–112.
- van Heijningen, C.A., de Visser, J., Zuidema, W., and ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl. Acad. Sci. USA* 106, 20538–20543.
- Fitch, W.T. (2017). Empirical approaches to the study of language evolution. *Psychon. Bull. Rev.* 24, 3–33.
- Ten Cate, C. (2017). Assessing the uniqueness of language: Animal grammatical abilities take center stage. *Psychon. Bull. Rev.* 24, 91–96.
- Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *J. Comp. Physiol. Psychol.* 71, 1–25.
- Thorpe, W.H. (1956). *Learning and instinct in animals* (Methuen).
- Inoue, S., and Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Curr. Biol.* 17, R1004–R1005.
- Kawai, N., and Matsuzawa, T. (2000). Numerical memory span in a chimpanzee. *Nature* 403, 39–40.
- Gentner, D. (1999). Analogy. In *The MIT Encyclopedia of the Cognitive Sciences*, F.C.K. R.A. Wilson, ed. (Cambridge: MIT Press), pp. 17–20.
- Conway, C.M., and Christiansen, M.H. (2001). Sequential learning in non-human primates. *Trends Cogn. Sci.* 5, 539–546.
- Amalric, M., Wang, L., Pica, P., Figueira, S., Sigman, M., and Dehaene, S. (2017). The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLoS Comput. Biol.* 13, e1005273.
- Murphy, R.A., Mondragón, E., and Murphy, V.A. (2008). Rule learning by rats. *Science* 319, 1849–1851.
- Spierings, M.J., and Ten Cate, C. (2016). Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proc. Natl. Acad. Sci. USA* 113, E3977–E3984.
- Berwick, R.C., Okanoya, K., Beckers, G.J., and Bolhuis, J.J. (2011). Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* 15, 113–121.

42. Sonnweber, R., Ravignani, A., and Fitch, W.T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Anim. Cogn.* *18*, 733–745.
43. Terrace, H.S. (2012). The Comparative Psychology of Ordinal Knowledge. In *The Oxford Handbook of Comparative Cognition*, 2 ed. (Oxford University Press).
44. Fountain, S.B., Rowan, J.B., Muller, M.D., Kundey, S.M.A., Pickens, L.R.G., and Doyle, K.E. (2012). The Organization of Sequential Behavior Conditioning, Memory, and Abstraction. In *The Oxford Handbook of Comparative Cognition, Second Edition*, E.A. Wasserman, and T.R. Zentall, eds. (Oxford University Press).
45. Ravignani, A., Westphal-Fitch, G., Aust, U., Schlumpp, M.M., and Fitch, W.T. (2015). More than one way to see it: Individual heuristics in avian visual computation. *Cognition* *143*, 13–24.
46. Westphal-Fitch, G., Huber, L., Gómez, J.C., and Fitch, W.T. (2012). Production and perception rules underlying visual patterns: effects of symmetry and hierarchy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *367*, 2007–2022.
47. Zhu, X., Sobhani, P., and Guo, H. (2015). Long Short-Term Memory Over Recursive Structures. *Proc. Int. Conf. Mach. Learn* *37*, 1604–1612.
48. Linzen, T., Dupoux, E., and Goldberg, Y. (2016). Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies. *Trans. Assoc. Comput. Linguist.* *4*, 521–535.
49. Bowman, S.R., Manning, C.D., and Potts, C. (2015). Tree-structured composition in neural networks without tree-structured architectures. In *Proceedings of the 2015 International Conference on Cognitive Computation: Integrating Neural and Symbolic Approaches*, pp. 37–42.
50. Kemp, C., and Tenenbaum, J.B. (2008). The discovery of structural form. *Proc. Natl. Acad. Sci. USA* *105*, 10687–10692.
51. Miller, G.F. (1967). Project Gramamama. In *The psychology of communication* (New York: Basic Books), pp. 125–187.
52. Wilson, B., Marslen-Wilson, W.D., and Petkov, C.I. (2017). Conserved Sequence Processing in Primate Frontal Cortex. *Trends Neurosci.* *40*, 72–82.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw data	This paper	https://data.mendeley.com/datasets/b7pfsj5kxk/draft?a=4ade10bc-3650-43d3-9491-7628be2cb7d2
Experimental Models: Organisms/Strains		
<i>Macaca mulatta</i>	Institute of Neuroscience, Chinese Academy of Sciences	N/A
Software and Algorithms		
MATLAB	Mathworks	R2015b
Psychtoolbox	http://psychtoolbox.org/	Version 3.0.13
SPSS Statistics	IBM	Version 24
Other		
Touchscreen	DELL	S2240T

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Liping Wang (liping.wang@ion.ac.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Monkey

Two adult male monkeys (*Macaca mulatta*, G: 12.5 kg and L: 8.1 kg) were used in the experiments. Experiments were performed in accordance with Institute of Neuroscience, Chinese Academy of Sciences guidelines for the use of laboratory animals. The monkeys were housed individually and fed *ad libitum*, but received water or juice on experimental days as rewards for correct responses during the tasks. During the experiments, the monkeys sat in a primate chair 30 cm from a computer monitor equipped with a touchscreen (DELL S2240T). Trial events, stimulus presentation, and data recording were computer controlled with MATLAB software (MathWorks, MA, USA).

Preschooler

Thirty-seven children were recruited (thirteen female; minimum age = 5 years, maximum age = 6 years, mean = 5.54 ± 0.64) from the Shanghai Ping-He and Yuan-Dong International School to perform a similar experiment. Fourteen of them were tested to learn the repeat and mirror grammars (length-2 and length-3 sequences), and twenty-three of them were asked to perform length-4 sequences. The experimental program was installed in the Microsoft Surface Pro4 System with a touchscreen. The experiment was framed as a game, which children came one by one to play in a quiet classroom. The ethical committee of the Institute of Neuroscience, Chinese Academy of Sciences, approved the experiments and all participants and their parents gave informed consent.

METHOD DETAILS

Stimuli

The spatial sequences were created from six locations (set: {1, 2, 3, 4, 5, 6}) that formed a symmetrical hexagon (Figure 1A). There are 150 sequences of the lengths 2 and 3 and 360 sequences of the length 4 on the hexagon. Each location was only sampled once in a sequence. Sequences were presented on the screen, and monkeys had to complete the sequence using either a “repeat” or “mirror” grammar. The “repeat” grammar defined sequences of the form AB|AB (length-2) or ABC|ABC (length-3), and the “mirror” grammar defined sequences of the form AB|BA or ABC|CBA (Figure 1A). For example, a visually presented sequence of ‘1 6 5’ had to be completed with a generative sequence of ‘1 6 5’ defined by the repeat grammar, and ‘5 6 1’ defined by mirror grammar. For both grammars, the total 150 sequences can be divided into 25 patterns based on their geometrical relations (Figure 1C). The pattern and the starting position for each sequence were randomly selected trial by trial. No three consecutive sequences from the same pattern or the same initial position were given to animals during either training or test periods. The procedure of testing preschool children was essentially identical (details see the description below).

Behavioral training

The monkeys learned to pull a lever (15 cm away from the touchscreen) to initiate a trial and not to release it during the presentation of the sample stimuli. A release of the lever at any time during the visual presentation ended the trial. The six locations, indicated by white circles (diameter: 3 cm, for monkey L) or white crosses (length: 2 cm, for monkey G) were always presented throughout the entire trial. The screen was black between trials. The visual presentation of the target sequence was indicated using a colorful dot (yellow or red, diameter: 3 cm) that flashed on each target location for 500 ms with a 500 ms inter-target interval (Figure 1A). After an 800 to 1200 ms delay, a 500 ms cue then appeared at the center of screen; this was either a green cucumber (for monkey G, the cue was a green circle), which indicated that the sequence should be finished using the “repeat” grammar, or a red apple (for monkey L, the cue was a yellow triangle), which indicated that the “mirror” grammar should be used. In another version of the task, the grammar cue was shown throughout the whole sample period. That is, the cucumber or square was flashed together with the sample location (see Video S1). When the white cross on the center of the screen turned to blue (‘go’ signal), the monkeys had to release the lever and touch the screen to indicate the locations according to the order defined by the grammar to be used. Sequence productions with wrong locations (those not presented during the sample sequence) or wrong orders were considered as errors, and whenever there was wrong position or order occurred, the trial was terminated and the program automatically moved to the next trial. Only the sequential touch of correct locations and order was rewarded with water or juice. Thus, the performance on each ordinal position was conditional accuracy (including the results in the Figures 1D, 5, S2, and S5). The inter-trial interval was 2000 ms, after which the monkey was allowed to pull the lever to start the next trial. During the training period, the inter-trial interval was prolonged to 4000 ms as a punishment for errors.

During the training period, for monkey G, 11 sequence patterns (five length-2 patterns and six length-3 sequences) were used. The six length-3 patterns were as follows: ‘+1+1’, ‘+5+5’, ‘+2+2’, ‘+4+4’, ‘+1+3’, and ‘+3+1’. For monkey L, 12 sequence patterns were selected, of which five were length-2 patterns and seven (‘+1+1’, ‘+1+2’, ‘+5+2’, ‘+1+4’, ‘+5+5’, ‘+3+4’, and ‘+2+2’) were length-3 sequence patterns. The remaining patterns were used for the generalization tests and were never shown to animals during training (Figure 2A).

Before grammar training, the monkeys received several weeks of procedural training, such as pulling, holding, and releasing the lever, and touching one or two locations on the screen that had been indicated by flashing dots. After this procedural training, monkeys received training with the two grammars (Figures 1B and S1). Grammar training always began with length-2 sequences. Once these had been learned, the length-3 sequences were introduced. The training order of sequences for monkey G was “repeat length-2,” “repeat length-3,” “mirror length-2,” and “mirror length-3” sequences. For monkey L, the order was “repeat length-2,” “mirror length-2,” “randomly presented repeat (length-2) and mirror (length-2),” “mirror length-3,” “repeat length-3,” and “randomly presented repeat (length-2&3) and mirror (length-2&3).” For the early phases of training, some “help” trials were included to motivate monkeys to perform the task. For example, the monkeys were allowed to touch wrong locations until correct sequence reproduction. The behavioral data in “help” and mixed (repeat and mirror)-sequence trials were not included in the present study. After several months of intensive training, both monkeys completed approximately 800 trials per day.

Grammar generalization in monkey G and monkey L

Different types of generalization test were carried out on separate days. Tests were terminated when the monkey’s performance reached 75% of the correct rate within one block (100 trials). Correct responses of the generalization trials were rewarded with water or juice.

Novel sequences on the standard hexagon

Once monkeys had learned the two grammars (correct rate > 80% for at least three consecutive training days, which corresponds to more than 2000 trials), the remaining untested sequences were used for the new sequence test (Figure 2A). On the test day, there was one training block (approximately 100 trials, which comprised 20% length-2 sequences and 80% length-3 sequences) with previously trained sequences as the pre-transfer control (Figure 2A). The sequences presented in the transfer blocks were all new. Both monkeys learned new sequences quickly, within the first transfer day. This generalization test ended when monkeys no longer responded, which lasted approximately 900–1000 trials. After the test, all 150 length-2 and length-3 sequences were given to the animals for consecutive 5–7 days training to acquire enough behavioral data for detailed analyses (e.g., Figures 1C and S2).

Longer-length sequences

For the length-4 higher-order sequence (ABCD|DCBA), ten patterns were randomly selected from the total possible 60 patterns on each test day on monkey G (Figure S3A). The generalization test was performed over six consecutive days. Before the generalization test, one block of the length-2 (20%) and length-3 (80%) sequences was first presented as the pre-transfer control. The sequences presented to monkeys in the transfer blocks were all length-4. Figure 2B shows the performance for the first day only. To avoid the learning effect of sequence patterns during each test day, for monkey L, the total 360 sequences were randomly assigned to the six test days. That is, the 60 sequences in each day could be from any patterns. After the length-4 sequence test, monkey G was further examined with length-5 higher-order sequences (ABCDE|EDCBA), for which ten patterns of length-5 sequences were chosen randomly for a 1-day test (Figure S3C). In this test day, the sequences presented to monkeys in the transfer blocks were all length-5, and the test was ended when monkey G stopped performing the task, due to the difficulty of the task and the low motivation.

After all the generalization tests, both monkey G and L were intensively trained around ~10,000 trials on the length-4 sequences. The behavioral data were used for the comparison between the well-trained monkeys and untrained preschool children (Figures 5 and S5).

Sequences on new geometrical forms

The new geometrical forms included the “Pyramid” (six locations), “Random” (random number of locations in random positions), “Horizontal line” (four to six locations), and “Two-hexagon” (Figures 2C–2F). Only one of these new spatial layouts was presented to animals on any one test day. All geometrical forms were the first time exposed to the monkeys. The tests were stopped when the correct performance rate reached 75% within one block. For the test of the “Random” form, the position of each location and number (4–10 items) of locations were randomized from trial to trial. The distance between locations was constrained to at least 6.5 cm to avoid potential touch mistakes on the touchscreen. For the “Two-hexagon” form, monkey G was tested with two fixed hexagons. One was placed in the second quadrant as the sample form and the other one was on the fourth quadrant as the reproduction. For monkey L, the two hexagons were randomly placed from trial to trial. There was no further training after the test day to allow re-learning of the new forms.

Rule-combination and generalization

Monkey G was habituated with the length-3 sequences generated by the embedded program, as follows: $S = M \{p, \{+1\}^2\}$, where S is the generated sequence for a given starting position p, M indicates the global mirror transformation (‘reverse’) and +1 denotes the local operation of clockwise movement (‘next’). For example, for a starting location of ‘2’, the monkey had to generate the remaining sequence by sequentially touching the five locations ‘3’, ‘4’, ‘4’, ‘3’, and ‘2’ to get a reward. The habituation was completed on the standard hexagon. The generalization tests used locations and geometrical shapes that had not previously been trained, including eight locations in an octagon, a square, a random number of locations in an irregular polygon and two-hexagon (Figure 4C). For the two-hexagon, the monkey was presented one location on the sample hexagon and was required to generate the following 2nd and 3rd targets on the sample (same) hexagon and the 4th, 5th and 6th targets by reversing the first half of the sequence on another corresponding hexagon (Figure 4C). On each day, only one generalization test was carried out. No further training was given after generalization tests.

Grammar learning and generalization in preschool children

To render the experiment more attractive to children, the flashed dots used to indicate locations were replaced by cartoon animals (e.g., jumping fishes, Figure S4). Each grammar (“repeat” and “mirror”) was tested in separate blocks. For the fourteen children, half of them started with the “repeat” grammar block, while the other half started with the “mirror” grammar block (five children stopped the experiment after completing the mirror task). Each block started with two full viewings of one length-2 (the 1st trial) and one length-3 (the 2nd trial) sequence, and children were instructed to look at how each animal jumped from one location to another. In trials 3 to 5, the experimenter demonstrated how to play the game by viewing the sample sequence (two or three locations) and then touching the locations on the screen to reproduce the second half of the sequence according to the grammar required. Then, the child was asked to play the game as the experimenter did in order to guess (reproduce) the positions after viewing the sample sequence. In each block, one type of form and one grammar were tested. Whenever children touched the wrong location or order, the program automatically restarted from the beginning of the trial, went on to correct the error, and asked children to guess the next location. The test was stopped after three consecutive successful trials. After the experiments, children were asked to comment on the game. The children explicitly reported their strategies to perform the “repeat” (14 out of 14 children) and “mirror” (9 out of 10 children) tasks after the experiment, which indicated that they clearly grasped the grammars.

Performance of length-4 sequences in preschool children

Twenty-three children were recruited to perform the length-4 sequences using the same computer program. The experimenter showed the children how to play the game until the children reported that they had understood the grammar (usually 3–5 demonstrations from the experimenter). For each grammar, there were thirty sequences of length-4 sequences (according to Figure S3A) given to the children and the order was randomized. Each grammar was tested in separate blocks. Half of them started to perform the mirror task and the other half performed the repeat task first. Each type of sequence was only allowed to test once. The initial position of a sequence from a pattern was randomly chosen from one of the six locations. Whenever children touched the wrong location or order, the trial was terminated and the program automatically continued to next sequence pattern. In each trial, children were given enough time to think to produce sequences. For the comparison between monkeys and preschool children, only behavioral performance in the mirror task was included and analyzed in the present study.

QUANTIFICATION AND STATISTICAL ANALYSIS

Chi-square tests were used to compare behavioral performance (accuracy, response time) at each ordinal position or overall sequential positions against what would be expected by chance. Chance performance on a standard hexagon was defined as 0.167 (1/6) at ordinal position one, 0.2 (1/5) at the second position, and so on, i.e., assuming that a location that had already been selected would not be repeated. For statistical comparisons involving generalization performance, we made the conservative assumption (backed up by Figure 1B) that the monkey’s spatial memory was perfect, and we tested for above-chance ordinal judgment against a chance

level for the whole sequence defined as $1/n!$ for length n , i.e., 0.5 ($1/2$) for length 2, 0.167 ($1/3 \cdot 1/2$) for length 3, 0.042 ($1/4 \cdot 1/3 \cdot 1/2$) for length 4 and 0.0083 ($1/5 \cdot 1/4 \cdot 1/3 \cdot 1/2$) for length 5 sequences. A linear model was used to test the monotonic relationship between ordinal position and behavioral performance (accuracy and response time) (Figures 5A and 5B). A linear mixed model was used for the analysis of the step-size and spatial grouping effects and their interactions with ordinal positions during sequence reproductions. (Figures 5C, 5D, and S5).

DATA AND SOFTWARE AVAILABILITY

Raw data are available on Mendeley dataset: <https://data.mendeley.com/datasets/b7pfsj5kxk/draft?a=4ade10bc-3650-43d3-9491-7628be2cb7d2>

Current Biology, Volume 28

Supplemental Information

**Production of Supra-regular Spatial
Sequences by Macaque Monkeys**

Xinjian Jiang, Tenghai Long, Weicong Cao, Junru Li, Stanislas Dehaene, and Liping Wang

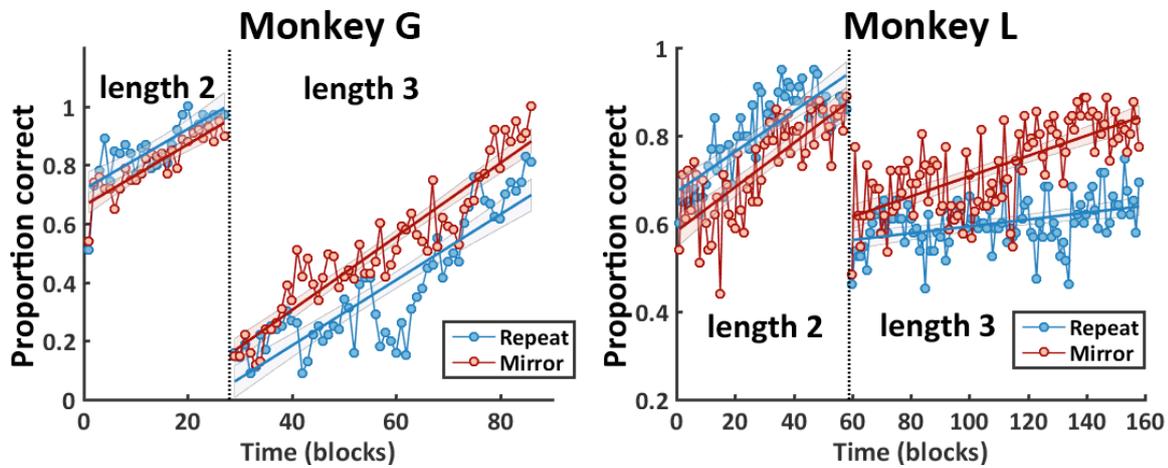
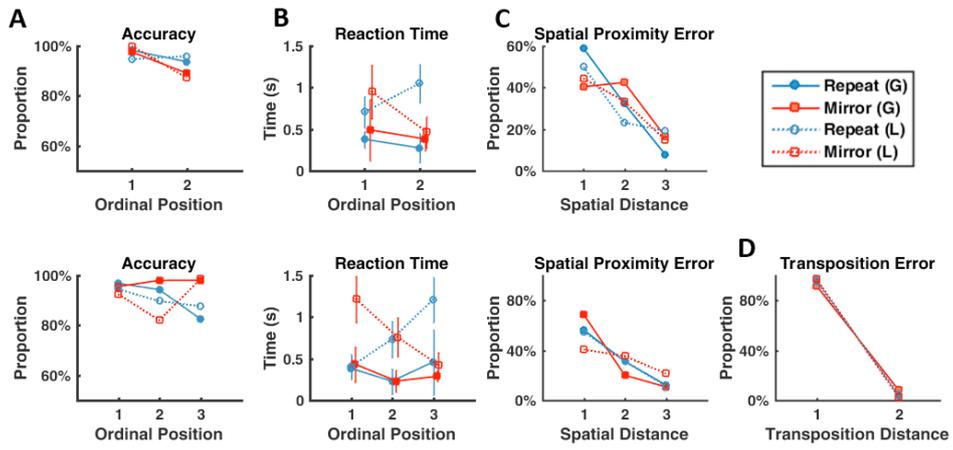
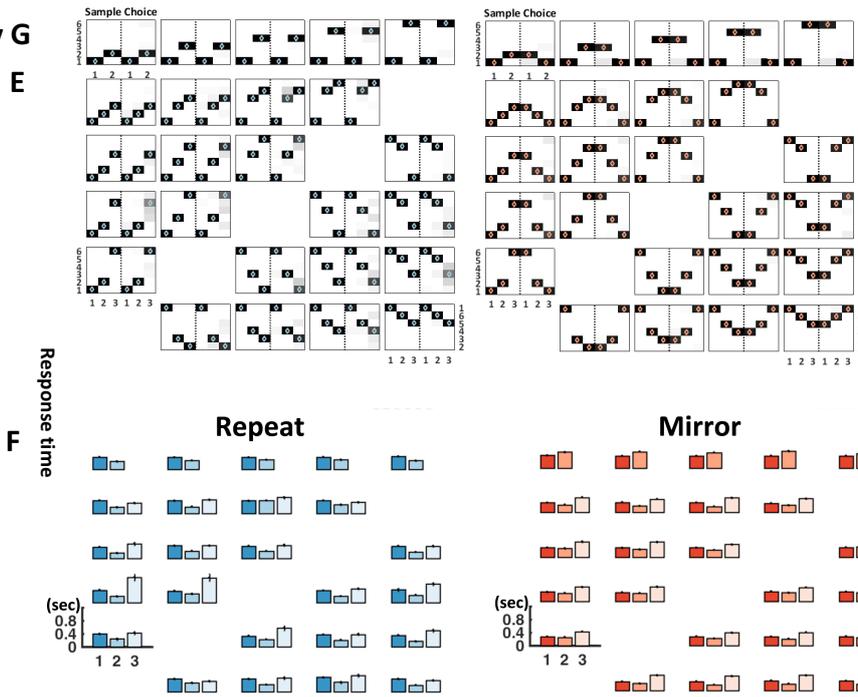


Figure S1. Learning dynamics, related to Figure 1. There was no significant difference in learning rate between the “mirror” and “repeat” grammar in monkey G (linear regression, length-2: $t = 0.09$, $p > 0.1$; length-3: $t = 1.31$, $p > 0.1$). Monkey L showed a lower learning rate in the “repeat” than that in the “mirror” on length-3 sequences (linear regression, length-2: $t = 0.58$, $p > 0.1$; length-3: $t = 5.00$, $p < 0.01$) was probably due to a long interval between the learning of “repeat” length-2 and length-3 sequences (Figure 1B). Each block contains 100 trials.



Monkey G



Monkey L

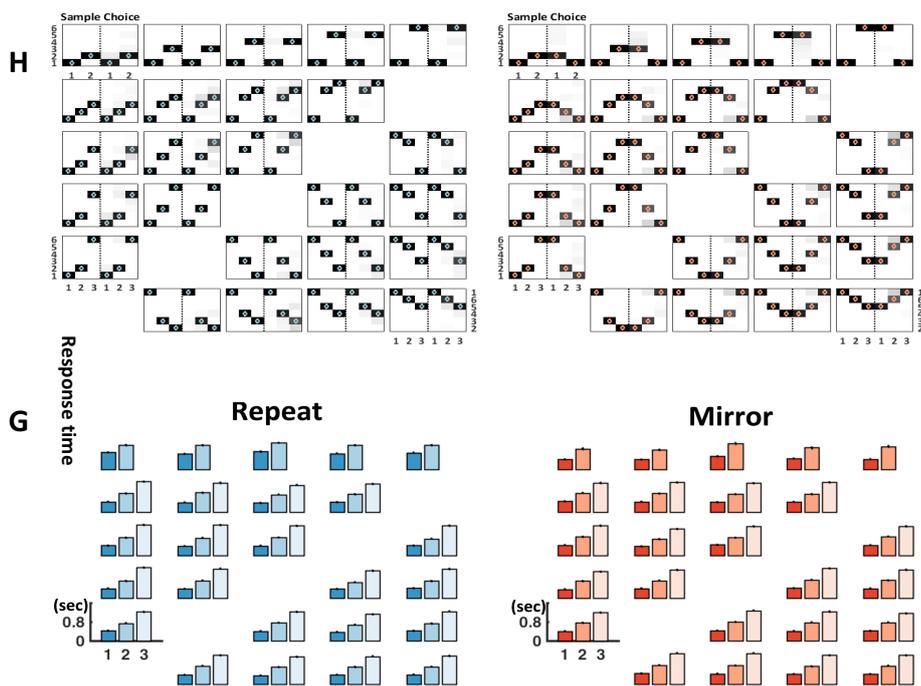


Figure S2. Behavioral performance on individual monkeys, related to Figure 1. (A, B)

Sample ordinal position curves for the “repeat” and “mirror” reproduction of spatial sequences in monkey G and L. Panels show data for reproduction (conditional) accuracy (**A**, *upper panel*: length-2; *lower panel*: length-3 sequences) and response time (RT, inter-response latency) (**B**). Both monkeys showed a primacy effect in the repeat task and monkey L showed primary and recency effect in the mirror task. The recall accuracy differed across ordinal positions in both tasks (Mantel-Haenszel chi-squared tests for each monkey per rule (“repeat” or “mirror”) per sequence length (2 or 3) with different patterns as stratum: $\chi^2 = 27.9$, $df = 1$, $p < 10^{-6}$ (monkey G, repeat, length-2); $\chi^2 = 250.6$, $df = 1$, $p < 10^{-16}$ (monkey G, mirror, length-2); $\chi^2 = 11$, $df = 1$, $p < 10^{-3}$ (monkey L, mirror, length-2); $\chi^2 = 0.2$, $df = 1$, $p > 0.1$ (monkey L, repeat, length-2); and $\chi^2 = 167.5$, $df = 2$, $p < 10^{-16}$ (monkey G, repeat, length-3); $\chi^2 = 75.4$, $df = 2$, $p < 10^{-16}$ (monkey L, repeat, length-3); $\chi^2 = 36.6$, $df = 2$, $p < 10^{-7}$ (monkey G, mirror, length-3); $\chi^2 = 460.4$, $df = 2$, $p < 10^{-16}$ (monkey L, repeat, length-3)). Note that the accuracies on all ordinal positions were much higher than the chance level ($\chi^2 > 72$, $df = 1$, $ps < 10^{-16}$ in all positions) (the performance was even higher than 95% in monkey G for the mirror task), indicating that animals did not perform the mirror task merely by relying on recency or primacy effect [S1]. Panel (**C**) shows the proportion of errors failing at different step sizes (spatial distances) (1: neighbor positions, 2: one-step movement further from the neighbor; 3: the point-symmetry position, the longest distance) from the position of the correct response ($\chi^2 > 36$, $df = 2$, $ps < 10^{-8}$ in all conditions of both monkeys). Panel (**D**) shows the proportion of errors involving a shift of one or two ordinal positions within the target list ($\chi^2 > 13$, $df = 1$, $ps < 10^{-3}$ in all conditions of both monkeys). The effects were consistent with previous serial recall tasks in humans [S2] and monkeys [S3]. Panels (**E, F, G, H**) show the evolution of choice proportion (**E, G**) and response time (**F, H**) of each step in the 25 sequence patterns (Figure 1C) for monkeys G and L. Both monkey G (data from 1704 “repeat” trials and 2494 “mirror” trials cross five consecutive training days) and monkey L (data from 1845 “repeat” trials and 3448 “mirror” trials cross six consecutive training days) showed significantly high correct rates on all sequences compared with chance-level (Chi-squared test, $ps \ll 10^{-8}$ in all cases), and consistent behavioral performance across different sequences within each monkey. Error bar indicates one SEM.

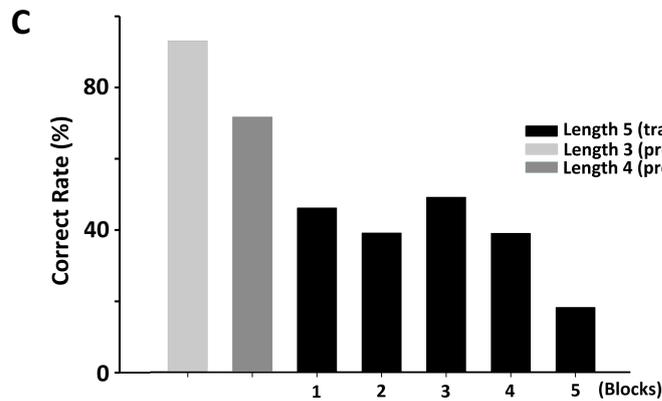
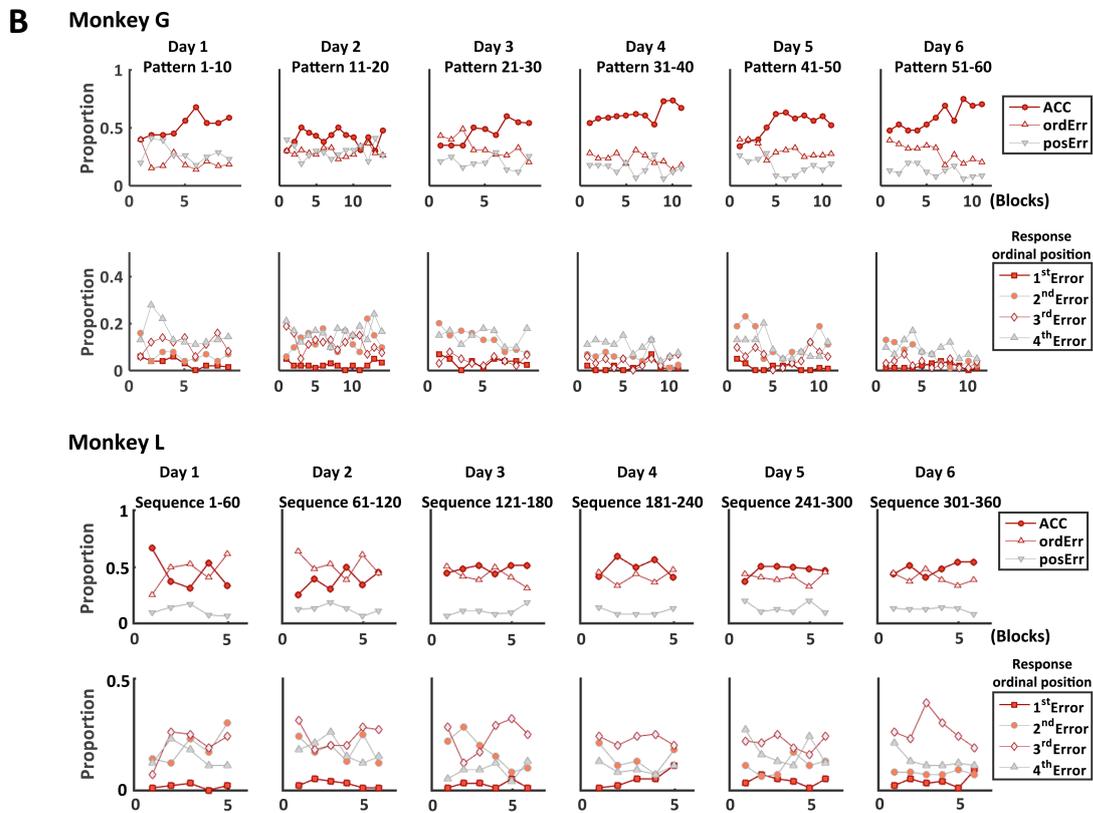
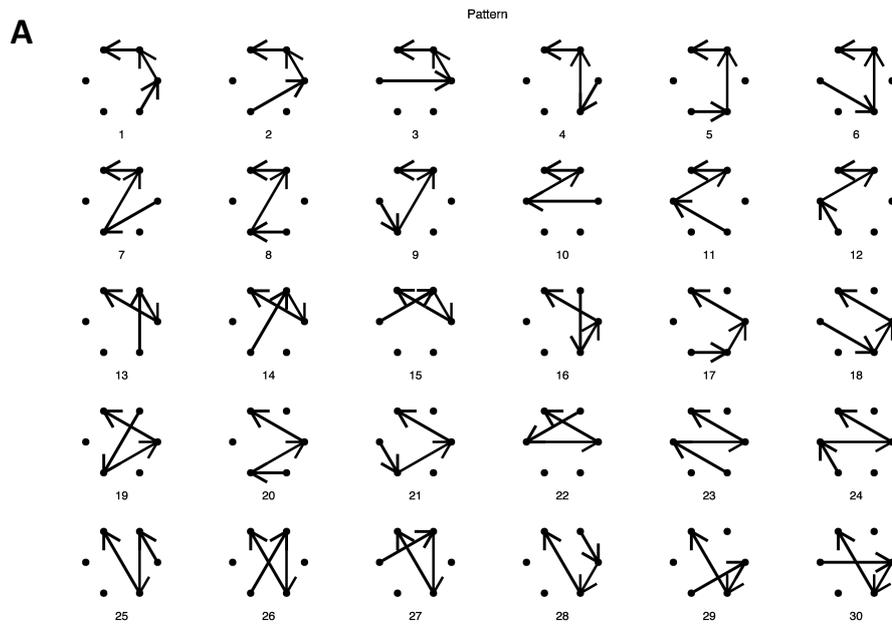
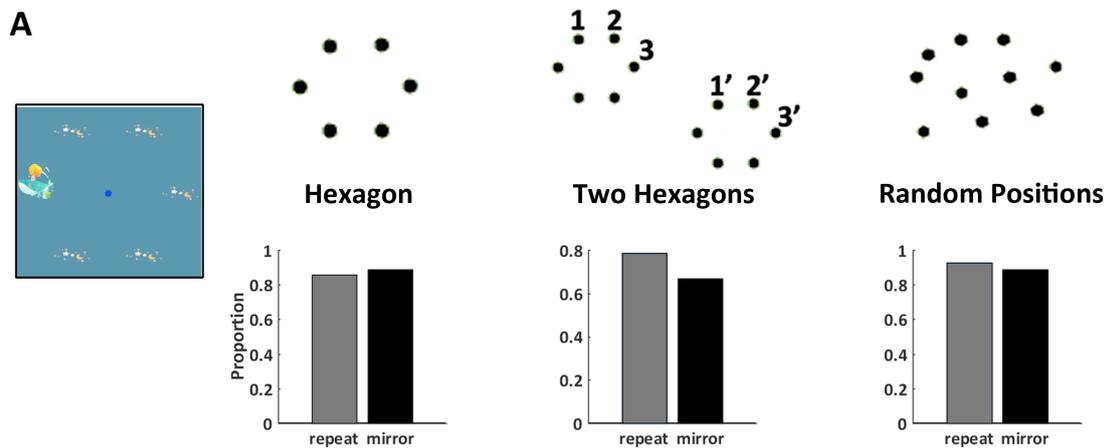


Figure S3. Generalization on length-4 and -5 sequences, related to Figure 2. (A, B)

Generalization of the “mirror” grammar to length-4 sequences in six consecutive days. There are sixty patterns (three hundreds and sixty sequences in total) included in the test. **(A)** shows the 30 patterns during the sequence production (the other 30 patterns include the sequences with the same locations in a reverse direction). ACC: accurate trial. **(B)** For monkey G, the sixty patterns were randomly assigned in the six test days, with ten patterns each day. For monkey L, the 360 sequences were randomly distributed into six test days (60 sequences for each day), regardless of patterns. The result showed that monkeys generalized the “mirror” grammar to all the novel length-4 sequences. The performance on all the length-4 sequences was significantly higher than chance-level ($\chi^2 > 3189$, $df = 1$, $ps < 10^{-8}$ in all transfer blocks). **(C)** Generalization of the “mirror” grammar to length-5 sequences in monkey G. The monkey stopped the task after 5 blocks. The performance on the length-5 sequences (trial by trial, the 500 sequences were randomly chosen from the database) was significantly higher than chance-level ($\chi^2 > 37$, $df = 1$, $ps < 10^{-6}$ in all transfer blocks).



B

Repeat	id1	id2	id3	id4	id5	id6	id7	id8	id9	id10	id11	id12	id13	id14
Hexagon	0	5	0	0	0	0	0	0	0	0	0	0	7	0
Two Hexagons	0	8	0	0	0	0	4	0	0	0	4	0	0	0
Random Positions	0	0	0	0	0	0	0	0	0	0	0	5	0	0

Mirror	id1	id2	id3	id4	id5	id6	id7	id8	id9
Hexagon	0	NaN	0	0	0	0	0	0	0
Two Hexagons	1	NaN	0	0	0	5	0	0	0
Random Positions	0	NaN	0	0	0	0	0	0	0

Figure S4. Preschool children (5-6 years old, n=14) quickly learned the “repeat” and “mirror” grammars and generalized them to new stimuli on “Hexagon”, “Two Hexagons” and “Random Positions”, related to Figure 5. (A) The picture on the left top shows the game used for children experiments, as used in our previous study[S4]. Right panels display the proportion of subjects showing the perfect performance (three consecutive correct trials immediately after demonstrations and an explicit report of the “repeat” or “mirror” grammar after the experiment). After a 5-trial demonstration, most of children were able to produce the “repeat” and “mirror” sequences and perfectly transfer to new positions on standard hexagon (12 out of 14 children for the “repeat” and 8/9 for the “mirror” grammar), two hexagons (11 out of 14 for the “repeat” and 6/9 for the “mirror” grammar) and random positions (13/14 for the “repeat” and 7/9 for the “mirror” grammar). The remaining children could also transfer both grammars within 10 trials, except that one child (id2) did not master the “mirror” grammar. **(B)** Table shows the individual’s performance, which indicated by the number of trials required to learn the grammar. The order of testing on the “Hexagon”, “Two-hexagons” and “Random Positions” was randomized. Zero means that subjects grasped the grammar from the first trial after the demonstration. NaN: fail to learn the grammar.

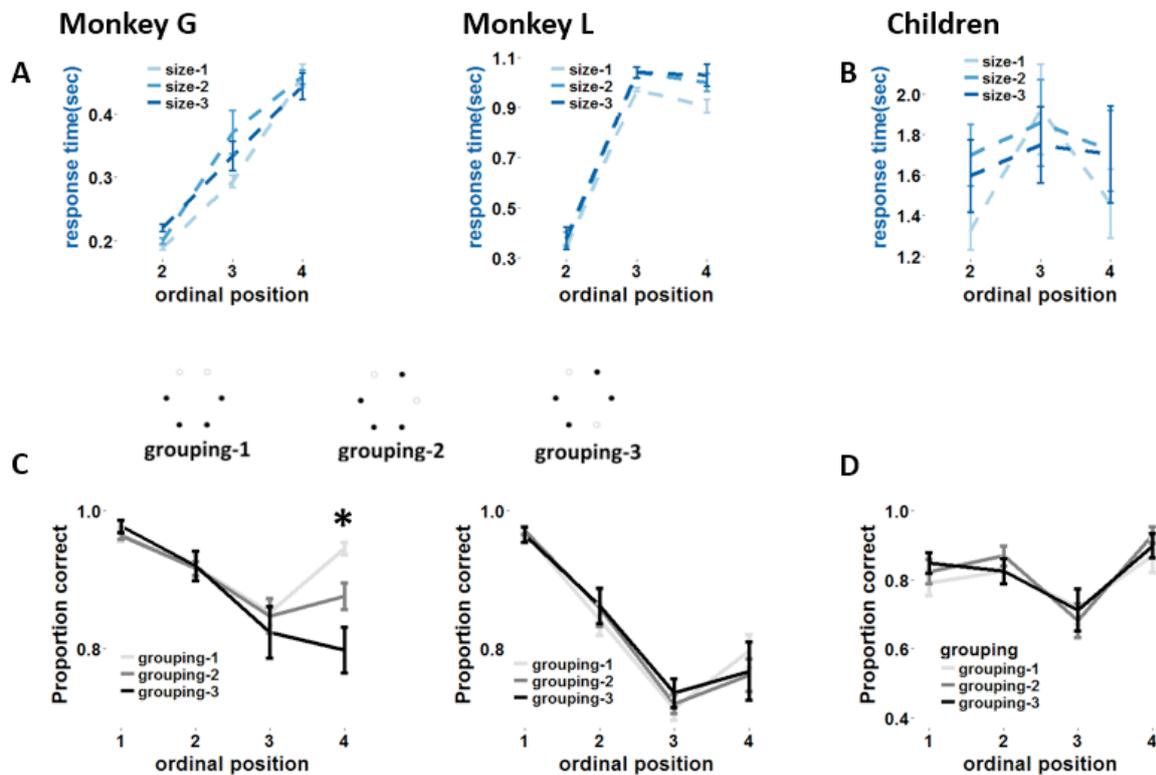


Figure S5. Comparison of length-4 sequence performance in monkeys and preschool children, related to Figure 5. For the two monkeys, after intensive training, the performance on the length-4 sequence was 57.5% correct on average (monkey G: 68%±2% correct, 5513 correct trials out of 8107, across 12 sessions (days) and monkey L: 47%±1% correct, 1754/3732, across 6 sessions). For the twenty-three untrained preschool children, their performance was 46%±5% correct (311 correct trials out of total 690, 30 trials per subject). The analysis in the Figures 5 and S5 was to examine the potential experimental factors including the step-size, spatial grouping and presence of crossing affecting the performance of sequence reproduction. The performance on each ordinal position was conditional accuracy (e.g. the accuracy at the 2nd position was only calculated with the trials in which the 1st position was correct, and the accuracy at the 3rd position was only included trials in which the 1st and 2nd positions were both correct, and etc.). In a standard hexagon, for each movement, there are three possible step-sizes: size-1, move to neighbor locations; size-2, move to locations two steps away; size-3: move to the point-symmetric location. For spatial grouping, we distinguished three possible spatial organizations: grouping-1, all four targets are grouped together; grouping-2, only three targets are spatially grouped together; grouping-3: targets are grouped in two separate groups of two locations.

(A) and **(B)** show the response times (RTs) on different step sizes at each ordinal position. No significant differences were observed on monkeys (Linear mixed model: main effect of step-size, monkey G, $F(2,70.55)=2.55$, $p > 0.1$; monkey L, $F(2,40)=1.39$, $p > 0.1$) or children ($n=23$) ($F(2,167.20)=1.89$, $p > 0.1$). **(C)** and **(D)** show the mirror task performance at each ordinal position when the sequences are divided according into three types of spatial organizations (grouping-1, -2 and -3). Spatial grouping did not affect the monkey L nor children (Linear mixed model: main effect of grouping, children, $F(2,239.69)=0.49$, $p > 0.1$; monkey L, $F(2,55)=0.04$, $p > 0.1$). Monkey G showed a significant grouping effect only at the 4th position ($F(2,110.79)=4.56$, $p < 0.05$), reflecting a possible spatial uncertainty of the last target. Error bar indicates one SEM.

Supplemental references:

- S1. Rey, A., Perruchet, P., and Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: evidence from baboons (*Papio Papio*). *Cognition* *123*, 180-184.
- S2. Hurlstone, M.J., Hitch, G.J., and Baddeley, A.D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological bulletin* *140*, 339-373.
- S3. Botvinick, M.M., Wang, J., Cowan, E., Roy, S., Bastianen, C., Patrick Mayo, J., and Houk, J.C. (2009). An analysis of immediate serial recall performance in a macaque. *Animal cognition* *12*, 671-678.
- S4. Amalric, M., Wang, L., Pica, P., Figueira, S., Sigman, M., and Dehaene, S. (2017). The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLoS computational biology* *13*, e1005273.