

Working Memory for Spatial Sequences: Developmental and Evolutionary Factors in Encoding Ordinal and Relational Structures

He Zhang (张贺),^{1,2*} Yanfen Zhen (甄艳芬),^{1*} Shijing Yu (余诗景),¹ Tenghai Long (龙腾海),^{1,2} Bingqian Zhang (张冰倩),^{1,2,3} Xinjian Jiang (姜新剑),¹ Junru Li (李俊汝),¹ Wen Fang (方文),¹ Mariano Sigman,^{4,5} Stanislas Dehaene,^{6,7} and Liping Wang (王立平)¹

¹Institute of Neuroscience, Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, People's Republic of China, ²University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China, ³School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, People's Republic of China, ⁴Laboratory Neuroscience, Universidad Torcuato di Tella, C1428 Buenos Aires, Argentina, ⁵School of Language and Education, Universidad Nebrija, 28015 Madrid, Spain, ⁶Collège de France, 75231 Paris Cedex 05, France, and ⁷Cognitive Neuroimaging Unit, CEA DSV/I2BM, INSERM, NeuroSpin Center, Université Paris Sud/Université Paris-Saclay, 91191 Gif-sur-Yvette, France

Sequence learning is a ubiquitous facet of human and animal cognition. Here, using a common sequence reproduction task, we investigated whether and how the ordinal and relational structures linking consecutive elements are acquired by human adults, children, and macaque monkeys. While children and monkeys exhibited significantly lower precision than adults for spatial location and temporal order information, only monkeys appeared to exceedingly focus on the first item. Most importantly, only humans, regardless of age, spontaneously extracted the spatial relations between consecutive items and used a chunking strategy to compress sequences in working memory. Monkeys did not detect such relational structures, even after extensive training. Monkey behavior was captured by a conjunctive coding model, whereas a chunk-based conjunctive model explained more variance in humans. These age- and species-related differences are indicative of developmental and evolutionary mechanisms of sequence encoding and may provide novel insights into the uniquely human cognitive capacities.

Key words: abstract pattern; evolution; sequence learning; working memory

Significance Statement

Sequence learning, the ability to encode the order of discrete elements and their relationships presented within a sequence, is a ubiquitous facet of cognition among humans and animals. By exploring sequence-processing abilities at different human developmental stages and in nonhuman primates, we found that only humans, regardless of age, spontaneously extracted the spatial relations between consecutive items and used an internal language to compress sequences in working memory. The findings provided insights into understanding the origins of sequence capabilities in humans and how they evolve through development to identify the unique aspects of human cognitive capacity, which includes the comprehension, learning, and production of sequences, and perhaps, above all, language processing.

Received Mar. 22, 2021; revised Nov. 8, 2021; accepted Nov. 19, 2021.

Author contributions: X.J. and L.W. designed research; H.Z., Y.Z., S.Y., T.L., B.Z., and X.J. performed research; H.Z., Y.Z., S.Y., T.L., J.L., W.F., and L.W. analyzed data; H.Z., Y.Z., M.S., S.D., and L.W. wrote the paper.

This work was supported by the Key Research Program of Frontier Sciences (Grant QYZDY-SSW-SMC001), the Strategic Priority Research Program (Grant XDB32070200), the Pioneer Hundreds of Talents Program from the Chinese Academy of Sciences, the Shanghai Municipal Science and Technology Major Project (Grant 2018SHZDX05), and the National Science Foundation of China (Grant 31871132) to L.W. We thank Danni Chen and Yiang Xu for experimental assistants. We also thank Guofang Ren and Yafang Xie from Far East Horizon Education Group for help in the data collection of children participants.

*H.Z. and Y.Z. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Liping Wang at liping.wang@ion.ac.cn.

<https://doi.org/10.1523/JNEUROSCI.0603-21.2021>

Copyright © 2022 the authors

Introduction

Most human behavior, from the way we move our eyes or walk, dance, or speak, to abstract cultural inventions such as reading or mathematics, are organized in sequences. As a consequence, the general ability to identify and learn sequences is a widespread feature across species and throughout development (Terrace and Mcgonigle, 1994; Saffran et al., 1996; Graybiel, 1998; Dehaene et al., 2015), but the specific ways by which sequences are learned can show substantial differences. Several studies converge to the rather intuitive idea that children have a less refined system to assimilate the structure of sequences (Orsini et al., 1987; Pickering et al., 1998; McCormack et al., 2000; Farrell Pagulayan

et al., 2006; Botvinick and Watanabe, 2007). For example, 7- to 11-year-old children perform worse than adults (>80%) in an immediate serial recall task (McCormack et al., 2000). Using a similar spatial sequence task in animals, the ability of monkeys to memorize the temporal order of a sequence has also been found to be relatively poor, with a performance that was <40% correct responses when the sequence length was 3 or 4 (Botvinick et al., 2009; Fagot and De Lillo, 2011).

Completely distinct changes could account for these observations. Just to name two categorically different possibilities, younger children could dispose of the same resources and functions to identify sequential structure, but operating at a lower resolution, or, alternatively, it could be that the operations by which sequences are identified are all together distinct. In other words, the variability underlying computational mechanisms of sequence learning across human groups and species remains largely unknown (Terrace and Mcgonigle, 1994). By exploring sequence-processing abilities at different human developmental stages and in nonhuman primates, we can begin to understand the origins of such capabilities in humans, and how they evolve through development to identify the unique aspects of human cognitive capacity, which includes the comprehension, learning, and production of sequences, and perhaps above all, language processing (Martin and Gupta, 2004; Dehaene et al., 2015). Comparative studies produce different patterns of sequence learning, and the challenge is to infer, from these patterns, the algorithms used to extract sequences by individuals of different species or ages.

Some computational modeling studies have suggested that sequences can be encoded through a conjunctive coding in human adults, which crosses the item with ordinal information (Botvinick and Watanabe, 2007; Oberauer and Lin, 2017). This idea has been primarily supported by electrophysiological studies; single neurons in the prefrontal cortex and caudate nucleus have been reported to respond selectively to particular items (i.e., shapes or locations), but their response to these items also depends on the ordinal position of items (Barone and Joseph, 1989; Kermadi et al., 1993; Kermadi and Joseph, 1995; Funahashi et al., 1997; Ninokura et al., 2003, 2004; Inoue and Mikami, 2006). The representational code processed by these neurons is conjunctive, in that the neurons respond maximally to a particular conjunction of item and ordinal position. It has been proposed that this conjunctive coding underlies how the brain associates individual items with individual serial positions to encode and maintain sequences. According to the conjunctive coding model, the precision of items and ordinal representations are fundamental factors that determine the accuracy of sequence encoding and memory. Thus, it can be hypothesized that, compared with adult humans, the limited capability of sequence encoding and maintenance in young children and nonhuman primates may be because of a lower precision of temporal ordinal or item representations.

A second, alternative proposal emphasizes that sequence memory depends not only on the number of items to be stored, but also on the presence of relational regularities (Marcus et al., 1999; Endress et al., 2009; Dehaene et al., 2015; Amalric et al., 2017; Wang et al., 2019). Rather than encoding the complete series of individual items, the process of sequence memory is enhanced by compressing items into a small number of known groups or “chunks” (Miller, 1956; Ericsson et al., 1980; Chase and Ericsson, 1982; Feldman, 2000; Cowan, 2001; Gilchrist et al., 2008; Brady et al., 2009). The sequences that humans judge as “complex” are not necessarily longer, but are less regular and

therefore more difficult to compress in working memory (Planton et al., 2021). Indeed, in our previous behavioral study, we found that accuracy in sequence encoding and production tasks varied according to sequence complexity (Amalric et al., 2017; Wang et al., 2019). Thus, we proposed that the complexity of a sequence is related to the length of its compressed form when it is encoded using an internal language (i.e., symmetries, rotations in geometry, or combinatorial rules).

In a recent review, we distinguished the following 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 (Dehaene et al., 2015). We proposed that only humans possess a representation of nested tree structures, also described as a “universal generative faculty” (Hauser and Watumull, 2017) or “language of thought” (Fodor, 1975), which enables sequence encoding by “compressing” information using abstract rules. By contrast, macaque monkeys are thought to be more limited in their ability to spontaneously detect relational structures between items and compress sequence memory using an internal language.

These hypotheses, nevertheless, have yet to be directly investigated. Both the precision of temporal order or item recognition and the learning of structured representations could depend on the evolutionary history of a species or environmental pressures during childhood. Furthermore, it is not yet clear whether the spontaneous memory compression using relational structures is unique to humans. Here, we directly tested these hypotheses by using the same spatial sequence reproduction task (Jiang et al., 2018) in human adults, children (6–7 years old), and nonhuman primates (macaque monkeys). We then combined conjunctive coding models to investigate the computational mechanisms underlying developmental and evolutionary factors that contribute to the learning of both ordinal information and relational structure during sequence encoding and compression.

Materials and Methods

Participants

The adult group comprised 40 adults (mean age = 24.0 years, age range = 21–27 years, 17 males) who were recruited from the Institute of Neuroscience, Chinese Academy of Sciences, and the Fenglin campus of Fudan University. Six adult participants (mean age = 25.0 years, age range = 22–27 years, three males) were recruited for the multisession experiment. Participant recruitment and experimental procedures followed the requirements of the ethical committee of the Institute of Neuroscience, Chinese Academy of Sciences. Informed consent was obtained from all participants. The experimental program was installed on the Microsoft Surface Pro4 System with a touchscreen.

The child group comprised 154 children (mean age = 6.4 years, age range = 6–7 years, 83 male) who were recruited from Shanghai Pudong Hongwen School. The ethical committee of the Institute of Neuroscience, Chinese Academy of Sciences approved the experiments, and all children and their parents gave informed consent. Seventeen children dropped out of the experiment, and their data were excluded from the final analysis. One additional child was excluded because of a failure to complete any of the sequences in the test session of the task, which indicated that the child did not understand the task. The experiment was framed as a game, which children played on an iPad tablet computer in landscape orientation in a classroom. The experimental program was built in Python 3.6 using the iOS Pythonista application (<http://omz-software.com>).

The nonhuman primate group comprised two adult male monkeys [Macaca mulatta: monkey 1 (M1), 12 kg; monkey 2 (M2), 9 kg]. Experiments were performed in accordance with the Institute of

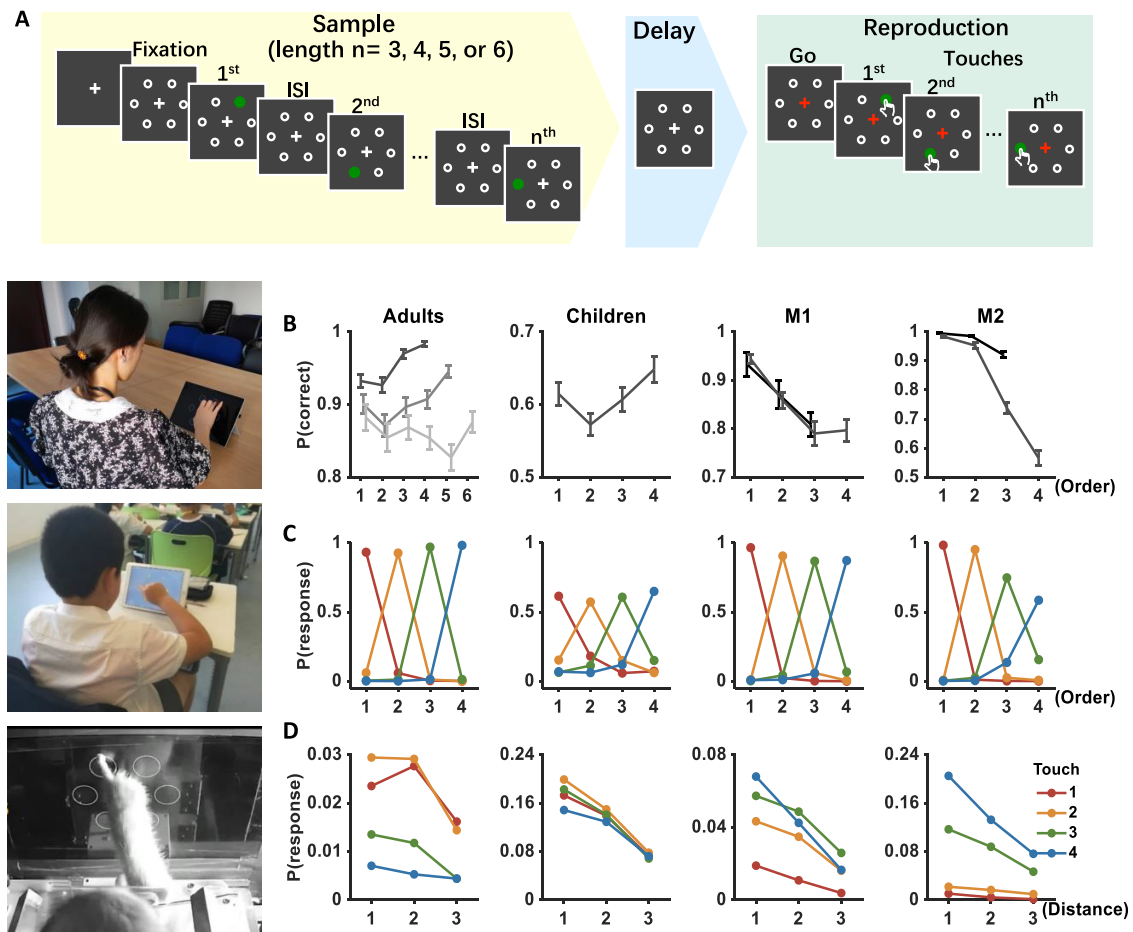


Figure 1. Delayed sequence reproduction task and behavior. **A**, Task. The task was highly similar between groups, and a trial was always initiated by the participants. A series of sample stimuli (adults, length = 4, 5, or 6; children, length = 4; monkeys, length = 3 or 4), chosen from six spatial locations, was presented at a fixed rate. After a delay, the go cue appeared, and participants were required to touch the screen to reproduce the sample sequence (for details, see Materials and Methods). **B**, Positional accuracy. Accuracies on each ordinal position during sequence reproduction. Error bars indicate SEM across participants (adults and children) or sessions (monkeys). **C**, Transposition gradient in the temporal order. *x*-axis, The temporal order; *y*-axis, the probability of response. The recall order of participants is more likely swapped with the neighbor orders. **D**, Transposition gradient in the spatial location. *x*-axis, The distance from target position on the hexagon; *y*-axis, the probability of error response. Participants wrongly select locations near target locations and the probability of selecting an error position decreases as its distance from target position increases.

Neuroscience, Chinese Academy of Sciences guidelines for the use of laboratory animals. The monkeys were housed individually and had *ad libitum* access to food but received water or juice on experimental days as rewards for correct responses during the tasks. During the experiment, the monkeys sat in a primate chair 30 cm from a computer monitor equipped with a touchscreen (model S2240T, DELL). Trial events, stimulus presentation, and data recording were controlled using MATLAB software (MathWorks).

Materials

The spatial sequences were created from six locations that formed a hexagon. Theoretically, the items in a sequence can locate on a continuous space (e.g., on arbitrary locations on a ring). To better control task difficulty and enable direct comparison between humans and monkeys, we adopted discrete locations in the current design. There were 360 sequences of the length 4, and 720 sequences of the length 5 and length 6 on the hexagon. Each location (a point on the hexagon) was sampled once within a given sequence (“without replacement”). Sequences were presented on the screen, and participants had to complete the sequence using a “repeat” or “mirror” rule. The repeat rule defined sequences in the form ABCD|ABCD (length-4), ABCDE|ABCDE (length-5), or ABCDEF|ABCDEF (length-6), and the mirror rule defined sequences in the form ABCD|DCBA, ABCDE|EDCBA, or ABCDEF|FEDCBA. The total of 360 length-4 sequences could be divided into 30 patterns

based on their geometrical relations. The pattern and the starting point for each sequence were randomly selected trial by trial. The procedure for testing human adults, children, and monkeys was essentially identical.

Procedure

The delayed sequence reproduction task was similar among groups (Fig. 1) but was tailored to be appropriate for each group.

Each trial was always initiated by the participants (clicking the mouse for human adults, touching the screen for children, and pulling a lever for monkeys). Once a trial was initiated, the six locations indicated by white circles (diameter, 3 cm) were always presented throughout the entire trial. The screen was blank between trials. The visual presentation of the target sequence was indicated using a dot with color (e.g., red: diameter, 3 cm) that flashed at each target location (duration: humans and M1, 250 ms; M2, 400 ms), with an intertarget interval of 250 ms for humans and 400 ms for monkeys. To render the experiment more attractive for children, cartoon figures (i.e., stars and a cartoon airplane) were used to indicate locations instead of the circles and flashing dot. After a short delay (duration: adults, 750 ms; children, 500 ms; monkeys, 400–800 ms), when the white cross turned to blue (the “go” signal, which was red for children), participants had to touch the screen to indicate the locations according to the order defined by the rule (repeat or mirror) to be used. Sequence productions with wrong locations (those not presented during the sample sequence) or wrong orders were considered as

errors. Feedback (a reward) was given to monkeys after the production of sequences. No feedback was given to human subjects, who were required to complete the sequence.

Familiarization/training phase

Humans. The experimental sessions were preceded by a familiarization phase. For adults, verbal instructions for the rule to be used were given and five practice trials of length-4 sequences were presented to familiarize participants with the task. For children, video-based instructions were given. Three example trials were presented together with verbal instructions via a video clip. Each example trial consisted of a full viewing of a length-4 sequence and sequential touches to reproduce the sequence according to the rule required. In the first example trial, stimuli presentation time was prolonged, and target locations were labeled with a number indicating its ordinal position. At the end of the video, experimenters verbally confirmed that children had understood the task. The video was played a second time when necessary.

Monkeys. Monkeys underwent a long-term training phase because verbal instructions could not be provided. The details of the training phase have been described previously (Jiang et al., 2018). During this phase, the monkeys pulled a lever to initiate a trial and were required to hold the lever down during the presentation of the sample stimuli. A release of the lever at any time during the visual presentation ended the trial. After a delay and a go signal had been presented, the monkeys had to release the lever and reproduce the sequence according to the rule to be used. Only the sequential touch of correct locations and orders was rewarded with water or juice. The intertrial interval was 2000 ms, after which the monkey was allowed to pull the lever to start the next trial. The intertrial interval was prolonged to 4000 ms as a punishment for errors.

Dataset

Adults completed 90 length-4 trials, 180 length-5 trials, and 180 length-6 trials with the repeat rule and the mirror rule, respectively. Sequences used were randomly selected. Participants performed the tasks of length-4, length-5, and length-6 in the same rule in three different blocks successively. Participants finished all blocks in the same rule then switched to the other rule. Children completed 90 repeat trials and 90 mirror trials on 2 separate days. On each day, participants finished one block (45 trials) in one rule and switched to the second block in the other rule. In each rule, three sequences were randomly selected from each sequence pattern. The order of rules was counterbalanced across participants in both adults and children. Only repeat trials were adopted in the current study. A total of 3600 length-4 trials, 7200 length-5 trials, and 7200 length-6 trials were completed by all adult participants. A total of 12,240 length-4 trials were completed by children.

To examine the within-pattern difference in individual participants, six adult participants were recruited for the multisession experiment. Each participant completed a total of 3600 repeat trials in five sessions (720 trials/session/d) within 10 d. A daily session consisted of two blocks, and participants had a short break between blocks to avoid fatigue. In each block, each of the 360 length-4 sequences was presented once, and the order of sequences was randomized.

For the two monkeys, the data were collected after they had completely learned the sequence task (Jiang et al., 2018). To summarize, monkeys learned two rules, repeat and mirror, of reproduction and manipulation of spatial sequence. The data used here included only those obtained during the repeat task. All sequences in a session were of the same length. Monkeys completed test sessions across several days. For M1, test trials were intermixed randomly with “error stop” trials (i.e., whenever position or order was incorrect, the trial was terminated, and the program automatically moved onto the next trial) within sessions. M1 was tested with 13,034 trials, including 7573 error stop trials, in 26 sessions (days). M2 was tested with a total of 8948 trials in 15 sessions. Error stop trials were included only in analysis of the accuracy and reaction time (RT) of the whole-sequence recall [i.e., Fig. 2 (see also Fig. 4C)], but not of the accuracy and RT of each rank [i.e., Fig. 1 (see also Fig. 4A,B)].

In the current study, only sequences with the repeat rule were included in the analysis for the three groups of subjects. All data needed to evaluate the conclusions in the article are present in the article. The

data that support the findings of this study are available from the corresponding author on reasonable request.

Statistical analysis

Unfinished trials (i.e., trials with a reproduced sequence that was shorter than the sample sequence or error stop trials without any response) and trials with repetitive touches at the same location were excluded from the analysis. Trials with any RT that was not within the mean \pm 3 SDs on a per subject (adults and children) or per session (monkeys) basis were also excluded.

Friedman’s test was used to test for difference in accuracy and RT across ordinal position. To test for the significance of primacy and recency effect in accuracy, planned pairwise comparisons were conducted between the first and second item, as well as between the last two items in sequences. To test for the changes in RT between successive responses in a trial, planned pairwise comparisons were conducted between successive RTs. Bonferroni’s correction was applied to correct for multiple comparisons.

Sequences sharing the same geometrical structures were categorized into one pattern. For example, the sequence 1234, 2345, 5612, etc. had the same relationship between items and was termed as Pattern 1 (Fig. 2A). Across patterns, sequences were paired by matching the starting point and orientation (clockwise and counterclockwise) of the sequence, resulting in 12 matched sequences in each pattern. Friedman’s test was used to compare accuracy difference between patterns (“between-pattern difference”) based on accuracies of sequences (averaged over different trials of the same sequence). Within each pattern, the accuracy difference between sequences (“within-pattern difference”) was tested using the Kruskal–Wallis test based on the performance on each trial (correct or incorrect). A Bonferroni correction was applied for within-pattern difference tests. To quantify the similarity of structural learning strategies between the different groups, we used Spearman’s rank correlation to calculate pattern accuracy for each pair of groups.

Based on the gestalt principles of proximity and similarity, spatially and temporally adjacent items tend to be perceived as a chunk. The 30 patterns were divided into eight chunking modes (see Fig. 4B) and were defined as follows: “1-1-1-1” (patterns 19, 22, 23, 26, and 27), where none of the temporally adjacent items were located spatially adjacent to each other; “1-2-1” (patterns 13, 14, 15, 16, 18, 29, and 30), where the second and third items in the sequence were located in adjacent spatial locations and formed a chunk, and a sequence consisted of one single item, a length-2 chunk, and another item; “1-1-2” (patterns 20, 21, 24, and 25), where the last two items in the sequence formed a length-2 chunk; “2-1-1” (patterns 6, 7, 10, and 11), where the first two items in the sequence formed a length-2 chunk; “2-2” (patterns 4, 5, 8, 9, and 12), where the first two items (i.e., first and second items), as well as the last two items (i.e., third and fourth items), were located in adjacent spatial locations, and there were two consecutive length-2 chunks in a sequence; “1-3” (patterns 17 and 28), where the second, third, and fourth items formed a length-3 chunk; “3-1” (patterns 2 and 3), where the first, second, and third items formed a length-3 chunk; and “[\pm 1]³ⁿ” (pattern 1), where all items were spatially adjacent to the preceding item, and the sequences could be described as “repeat one-step movement three times.” The whole sequence was a length-4 chunk.

The complexity of each pattern was defined according to chunk size ($\sum_{i=1}^n \frac{1}{k_i}$, where k_i is the size of the chunk that contained the i th item, and n is the sequence length). For example, patterns in groups 1–3 consisted of a length-1 chunk and a length-3 chunk, and the chunk sizes of each item were (1, 3, 3, 3), respectively. Thus, the complexity of these patterns was $(1 + 1/3 + 1/3 + 1/3) = 2$. Note that the value of the complexity that was defined according to chunk size was equal to the chunk number in the sequence. RTs of each response were averaged over all correctly reproduced sequences. Spearman’s rank correlation was used to examine the relationship between pattern complexity and accuracy, as well as with average RTs of all touches in correct trials within each group. In each chunking mode, the RTs of each ordinal position were transformed to z scores with the mean and SD of all sequences. Wilcoxon tests were used to test for the significance of the within-chunk

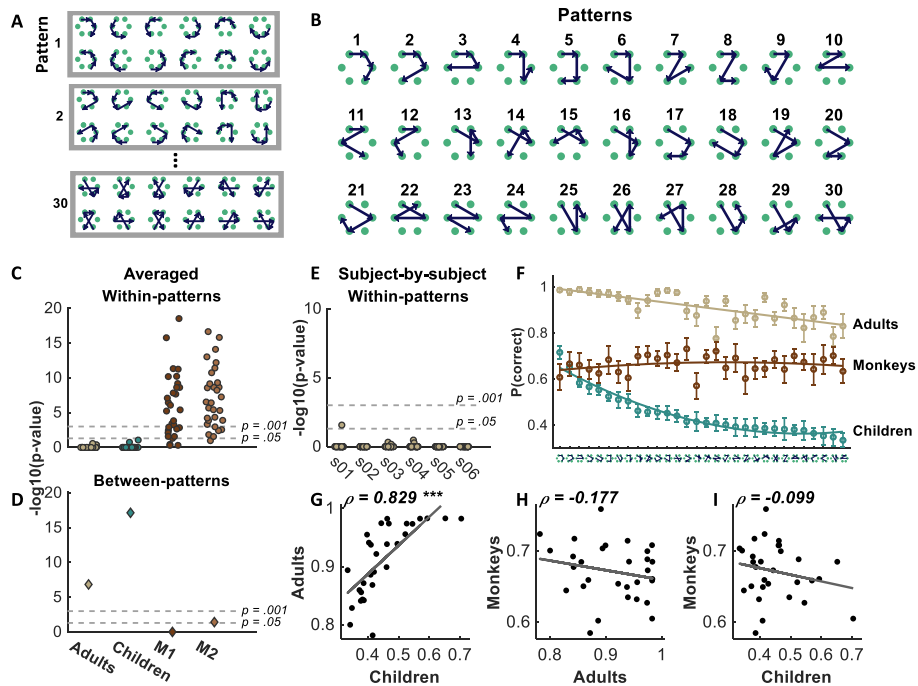


Figure 2. Extraction of relational structure in humans. **A, B**, Sequences (**A**) and patterns (**B**). The lines with the arrows on each hexagon mark the trajectory that connects the items in a sequence. Every 12 sequences, regardless of starting position and moving direction (clockwise or counterclockwise), shared the same relational structure and were grouped into one sequence pattern (surrounded with gray box). A total of 30 patterns was defined based on their relational structures. **C, D**, Within-pattern (**C**) and between-pattern (**D**) differences based on averaged data (averaged across participants in humans and averaged across sessions in monkeys). Within each pattern, the accuracy difference between sequences was tested with the Kruskal–Wallis test for each pattern, denoted by each dot in **C**. A Bonferroni correction was applied for within-pattern difference tests. Across patterns, sequences were grouped by matching the starting point and orientation (clockwise and counterclockwise) of the sequence, resulting in 12 matched sequences in each pattern. A Friedman test was used to compare the accuracy difference between patterns based on average accuracies of sequences (**D**). Log-transformed p values (y -axis) of within-pattern and between-pattern differences were plotted for each group. Larger log-transformed values correspond to lower p values. Horizontal dash lines mark the log-transformed values when $p = 0.05$ and $p = 0.001$. **E**, Within-pattern differences on subject-by-subject basis. Bonferroni's corrections for multiple comparisons were applied. The lack of significant within-pattern difference was highly consistent in each participant. **F**, Accuracy of each pattern in the three groups. Patterns were sorted according to children's accuracy in descending order. A quadratic polynomial fitting line is shown for each group. Error bars indicate SEM across sequences. **G–I**, Correlations of the mean accuracies of patterns between groups ($n = 30$). *** $p < 0.001$.

RT decrease (i.e., the difference between the first and second item within a chunk).

Spearman's rank correlation was used to test the correlation between children's spatial chunking strategies and learning performance at school. Children's scores in Chinese and math examinations ~ 2 months after test sessions were averaged and used as an index of examination performance. Outliers that exceeded the range of the median examination score ± 3 scaled median absolute deviations were excluded. Correlation analyses between children's accuracy in sequences with and without chunking strategies and examination performance were performed. Accuracies used in the analysis were the average of reproduction accuracies in sequences with and without chunking strategies.

Conjunctive coding model specifications

Simulations were implemented using MATLAB (MathWorks). The model consisted of (1) the encoding process for the input of sequence information, and (2) the retrieval process for the output of sequence information (Fig. 3A).

Encoding

The encoding matrix (EM) of the input sequence information S (EM_S) was determined according to the encoding process f_E , as follows:

$$EM_S = f_E(S),$$

where a sequence was defined as $S = (T_1, T_2, \dots, T_n)$ and $\forall i \in \{1, 2, \dots, n\} : T_i \in \{1, 2, \dots, N\}$. Only sequences with a length $n = 4$ were used in the current study. There were $N = 6$ potential locations for sequential stimuli, and each location was sampled

only once within a given sequence (without replacement: $\forall i, j \in \{1, 2, \dots, n\} : i \neq j \Rightarrow T_i \neq T_j$).

Several assumptions were made about the encoding process of sequence information. First, we assumed that the information of T_i is a conjunction of order information $i \in \{1, 2, \dots, n\}$ and item information $T_i \in \{1, 2, \dots, N\}$.

Second, for a specific target (T_i), the estimation of the model for i of T_i (marked as a random variable X) obeyed the scaled Laplace distribution (also known as the double exponential distribution), and the estimation of the model for T_i of T_i (marked as a random variable Y) was subject to a scaled von Mises distribution (also known as the circular normal distribution), as follows:

$$X \sim \text{Laplace}(i, \lambda^{-1}) : E_{\text{order}}(X|T_i, \lambda) = e^{-\lambda|X-i|} \quad (1)$$

$$Y \sim \text{von Mises}(T_i, \kappa) : E_{\text{item}}(Y|T_i, \kappa) = e^{-\kappa \left(1 - \cos \left(\frac{2\pi}{N} (Y - T_i) \right) \right)}, \quad (2)$$

where λ and κ are dispersion measures of the distributions, which controlled the estimated precision of the order information i and the target T_i ; a higher λ indicates a higher estimation accuracy of i , while a higher κ indicates a higher estimation accuracy of T_i .

We chose the Laplace distribution to describe the representations of ordinal information on the basis of previous works (Nosofsky, 1986; Shepard, 1987; Brown et al., 2007). We chose the von Mises distribution to describe the representations of item information as it is a continuous probability distribution on the circle. It is a close approximation to the wrapped

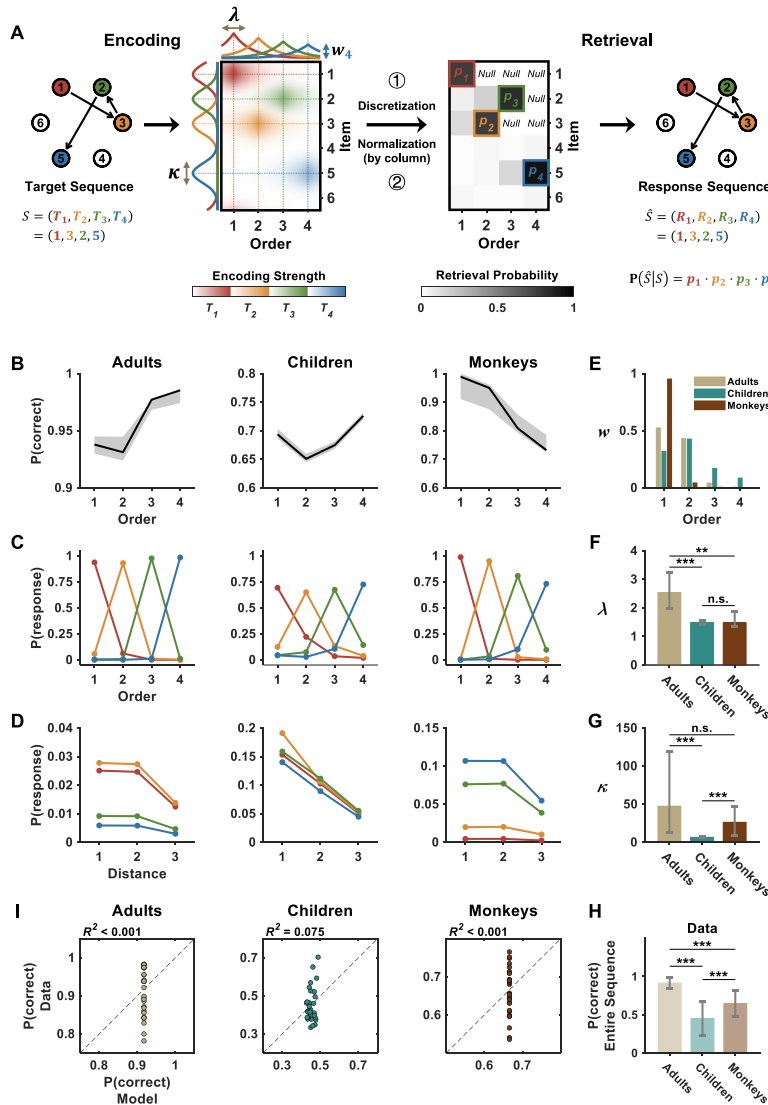


Figure 3. Fitting behavioral data to the conjunctive coding model. **A**, Model schema. The model assumed that the representation code of a spatial sequence is a conjunction of the spatial items (six locations on the hexagon) and ordinal positions (e.g., first, second, third, and fourth); for details, see Materials and Methods). The binding space shown is the encoding matrix of example sequence 1-3-2-5, as the location “1” at first (red), the “3” at second (orange), the “2” at third (green), and the “5” at fourth (blue). During sequence retrieval, the decoding matrix was obtained via a discretization and normalization of the encoding matrix (see Materials and Methods). In addition, conditional probabilities were used. Each retrieved item was removed in the subsequent step because in the current task, each location was sampled only once in a sequence. The probability of correct retrieval of the target sequence 1-3-5-2 was shown (for details, see Materials and Methods). **B–D**, Positional accuracy (**B**), transposition gradient in temporal order (**C**), and transposition gradient in spatial location (**D**) fitted from the model. Shaded error bars (**B**) indicate the 95% CI of model fitting results from 1000 bootstrap resamples, and the solid lines (**B–D**) are the results of the model using the medians of the best fitting parameters from 1000 bootstrap resamples. **E–G**, The precision of temporal order (λ ; **E**), spatial location (κ ; **F**), and weights (w) assigned along with the ordinal ranks (**G**) from the model fitting of each group. The bars are the medians of the best fitting parameters from 1000 bootstrap resamples, and the error bars indicate the 95% CI. **H**, The accuracy of sequence reproduction in human adults, children, and monkeys. Error bars indicate the SD across participants (in adults and children) or sessions (in monkeys, averaged over all the sessions across two monkeys). **I**, The performance of 30 patterns was not predicted by the model. Each dot corresponds to one pattern. R^2 was the coefficient of determination to evaluate the goodness-of-fit of simulated versus measured $p(\text{correct})$ of the 30 patterns (see Materials and Methods). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. n.s., Nonsignificant.

normal distribution, which is the circular analog of the normal distribution. These encoding probabilities can be regarded as a negative exponential function of the “distance” between the target and the estimated information in a psychological space based on the universal law of generalization (Shepard, 2012).

Finally, E_{order} and E_{item} were integrated as the sum of the joint distributions of all targets, weighted by constants w_i ($\sum_{i=1}^n w_i = 1$), which are the parameters that the model needs to learn. The integration resulted in a graded conjunctive representation of the sequence in memory for later

retrieval. With the background noise η added in the rank and item binding space, the encoding matrix of S (EM_S) is given by the following:

$$EM_S(X, Y) = f_E(X, Y|S) = \sum_{i=1}^n w_i \cdot$$

$$E_{\text{order}}(X|T_i, \lambda) \cdot E_{\text{item}}(Y|T_i, \kappa) + \eta, \quad (3)$$

where the background noise η is a constant, so it could be seen as the uniform distribution under the random choice. We make no specific assumption about the source of background noise. It could be arisen from attention to all locations or passive priming of all location-rank combinations. The background noise becomes important when the w_i value is small enough. This could help interpret the disappearance of the recency effect in monkeys.

The encoding matrix was then discretized. The probability mass function of Y (discrete analog) retains the form of the probability density function of X (continuous random variable; Botvinick and Watanabe, 2007; Brown et al., 2007).

Retrieval

The output of sequence information \hat{S} was determined according to the retrieval process f_R , as follows:

$$P(\hat{S}|S) = f_R(\hat{S}|EM_S),$$

where a retrieved sequence was defined as $\hat{S} = (R_1, R_2, \dots, R_n)$ and $\forall i \in \{1, 2, \dots, n\} : R_i \in \{1, 2, \dots, N\}$, and its subsequence was defined as $\hat{S}^{(i)} = (R_1, R_2, \dots, R_i)$ and $\hat{S}^{(0)} = \emptyset$.

We use the conditional probabilities to describe the retrieval process as follows:

$$P(\hat{S}|S) = f_R(\hat{S}|EM_S) = \prod_{i=1}^n P(R_i|\hat{S}^{(i-1)}) = \prod_{i=1}^n \frac{EM_S(i, R_i)}{\sum_{j \notin \hat{S}^{(i-1)}} EM_S(i, j)}, \quad (4)$$

where $j \notin \hat{S}^{(i-1)}$ denotes that the items already retrieved would be removed in a subsequent step, given that each location was sampled only once (without replacement) within a sequence. Normalization was performed within each order by dividing the sum of probabilities of the remaining items. The 1-norm normalization is based on the choice axiom of Luce (1959). Taking $S = (T_1, T_2, T_3, T_4) = (1, 3, 2, 5)$, and retrieved sequence $\hat{S} = (R_1, R_2, R_3, R_4) = (1, 3, 2, 5)$ as an example, the retrieval process is shown in Figure 2A.

Chunk-based conjunctive coding model

The chunk-based model assumed that chunking processing would improve the order precision (λ).

There are several additional assumptions for this model. First, we assumed that items in a sequence could be grouped into chunks based on their spatial and temporal proximity and similarity. The chunked sequence information S_C was determined according to the chunking process f_C for a specific sequence $S = (T_1, T_2, \dots, T_n)$, as follows:

$$S_C = f_C(S) = (T_1^C, T_2^C, \dots, T_{n_C}^C)$$

$$T_i^C = \left(T \sum_{j=1}^{i-1} k_j + 1, T \sum_{j=1}^{i-1} k_j + 2, \dots, T \sum_{j=1}^i k_j \right),$$

where n_C is the number of chunks within S_C and k_i is the number of targets (spatial item) within T_i^C (chunk size).

Chunks were defined based on the gestalt principles of proximity and similarity. Spatially and temporally adjacent items were put into the same chunk according to the following rules:

$$\forall j \in \{1, 2, \dots, k_i - 1\} \text{ and } k_i > 1 : |T_i^C(j+1) - T_i^C(j)| = 1 \text{ or } N - 1$$

$$\forall i \in \{1, 2, \dots, n_C - 1\} : |T_{i+1}^C(1) - T_i^C(k_i)| \neq 1 \text{ and } N - 1.$$

Taking the sequence $S = (1, 3, 2, 5)$ as an example (Fig. 4D), since the second and the third item were spatially adjacent, it can be regarded as $S = ([1], [3, 2], [5])$, which consists of three chunks: $[1]$, $[3, 2]$, and $[5]$.

Second, we also assumed that items within the same chunk shared a common order precision λ_{k_i} , which is determined by the number of items in the chunk (i.e., chunk size k_i , where i represents the i th item in the sequence). For example, the chunk sizes for each item in $S = ([1], [3, 2], [5])$ are 1, 2, 2, and 1, respectively, and the precision of the temporal order for all targets is $\Lambda = ([\lambda_1], [\lambda_2, \lambda_2], [\lambda_1])$. We recalculated λ of each T_j based on chunk size k_i , as follows:

$$\forall T_j \in T_i^C : E_{\text{order}}(X|T_j, \lambda_{k_i}) = e^{-\lambda_{k_i}|X-j|}. \tag{5}$$

$\{\lambda_{k_i}\}$ were free parameters, and their fitting results can be seen in Figure 5C.

Path-based models

Path-based models use path characteristics of the sequences, such as the path length L and the path crossings number n_{crs} , to recalculate λ for a specific sequence $S = (T_1, T_2, \dots, T_n)$.

The path length-based model was as follows:

$$\forall T_j \in S : E_{\text{order}}(X|T_j, \lambda_L) = e^{-\lambda_L|X-j|} \tag{6}$$

$$L = \sum_{i=1}^{n-1} \left| \frac{2\pi}{eN} \iota T_{i+1} - \frac{2\pi}{eN} \iota T_i \right|,$$

where L refers to the length of the trajectory that connects the items of sequence S . ι is the imaginary unit.

The relationship between λ_L and L was assumed to be as follows:

$$\lambda_L = \lambda e^{-aL} + b,$$

where λ , a , and b are non-negative free parameters.

The path crossing-based model was as follows:

$$\forall T_j \in S : E_{\text{order}}(X|T_j, \lambda_{n_{\text{crs}}}) = e^{-\lambda_{n_{\text{crs}}}|X-j|}, \tag{7}$$

where n_{crs} represents the number of crossings in the spatial sequence. When there was no crossing in the sequence [e.g., $S = (1, 2, 3, 4)$], $n_{\text{crs}} = 0$ or when there was one crossing [e.g., $S = (1, 3, 2, 5)$], $n_{\text{crs}} = 1$.

Model fitting

There were $P_n^N = \frac{N!}{(N-n)!} = 360$ different length-4 sequences.

However, the absolute target locations have no effect on the retrieval results in our model. The sequences were further regrouped into 30 patterns starting from location 1 based on the sequential geometrical relationships among spatial targets (Fig. 2B) with rigid transformations, including rotations, reflections, translations, or their combinations.

Moreover, there were still 360 possible response sequences $\{\hat{S}_j\}_{j=1}^{360}$ for a specific stimulus pattern S_i in $\{S_i\}_{i=1}^{30}$.

We chose the square error as the loss function (quadratic loss function) to estimate the free parameters in a particular model, as follows:

$$L(P, Q) = \sum_{i=1}^{30} \sum_{j=1}^{360} (P_{ij} - Q_{ij})^2. \tag{8}$$

P_{ij} is the relative frequency of the data response sequence \hat{S}_j for a given stimulus pattern S_i , and can be expressed in terms of n_{ij} (the total occurrence of response sequences \hat{S}_j given S_i), as follows:

$$P_{ij} = f(\hat{S}_j|S_i) = \frac{n_{ij}}{\sum_{k=1}^{360} n_{ik}}. \tag{9}$$

Q_{ij} is the predictive probability of response sequence \hat{S}_j of the model for a given stimulus pattern S_i , as follows:

$$Q_{ij} = P(\hat{S}_j|S_i). \tag{10}$$

The *fminsearch* function was used to minimize the loss function in MATLAB.

In the model fitting for length-4 sequences, the original conjunctive coding model had the following seven free parameters: λ , κ , $\{w_i\}_{i=1}^4$, η . The chunk-based model had the following 10 free parameters: $\{\lambda_{k_i}\}_{k_i \in \{1, 2, 3, 4\}}$, κ , $\{w_i\}_{i=1}^4$, η . The path length-based model had nine free parameters, as follows: a , b , λ , κ , $\{w_i\}_{i=1}^4$, η . The path crossing-based model had eight free parameters, as follows: $\{\lambda_{n_{\text{crs}}}\}_{n_{\text{crs}} \in \{0, 1\}}$, κ , $\{w_i\}_{i=1}^4$, η .

It is important to note that at least ≥ 30 trials were needed to obtain a trusted relative frequency P_{ij} for each pattern, and that at least $30 \times 30 = 900$ trials were needed to obtain a trusted $P_{30 \times 360}$ matrix for the model fitting. The trial number in the current study was insufficient for data from each individual participant to fit a model. Thus, data from all participants within the same group were modeled together, and 1000 bootstrap resamples from data were used to calculate statistical properties of the best fitting parameters (Fig. 3F,G). We used the medians of the best fitting parameters to fit the most representative models, and the behavioral benchmarks of these models are shown in Figure 3B–D. Then, we used a random permutation test ($N = 1000$) to examine the differences in parameters among the groups (Fig. 3F,G).

The number of trials also limited the fold number of the cross-validation. Therefore, repeated threefold cross-validation ($N_{\text{repeat}} = 100$) was used to evaluate different models. We chose the Bayesian information criterion (BIC) as the criterion of model evaluation because the chunk-based models had the most parameters and the BIC generally penalizes model fits with increasing numbers of free parameters more strongly than does the Akaike information criterion (AIC). Without the constant term $n \ln(n)$, AIC and BIC were calculated as follows:

$$\text{AIC} = n \ln(\text{RSS}) + 2k \tag{11}$$

$$\text{BIC} = n \ln(\text{RSS}) + k \ln(n), \tag{12}$$

where $\text{RSS} = \min(L(P, Q))$ is the residual sum of squares, $n = 30 \times 360$ is the number of data points, and k is the number of free parameters.

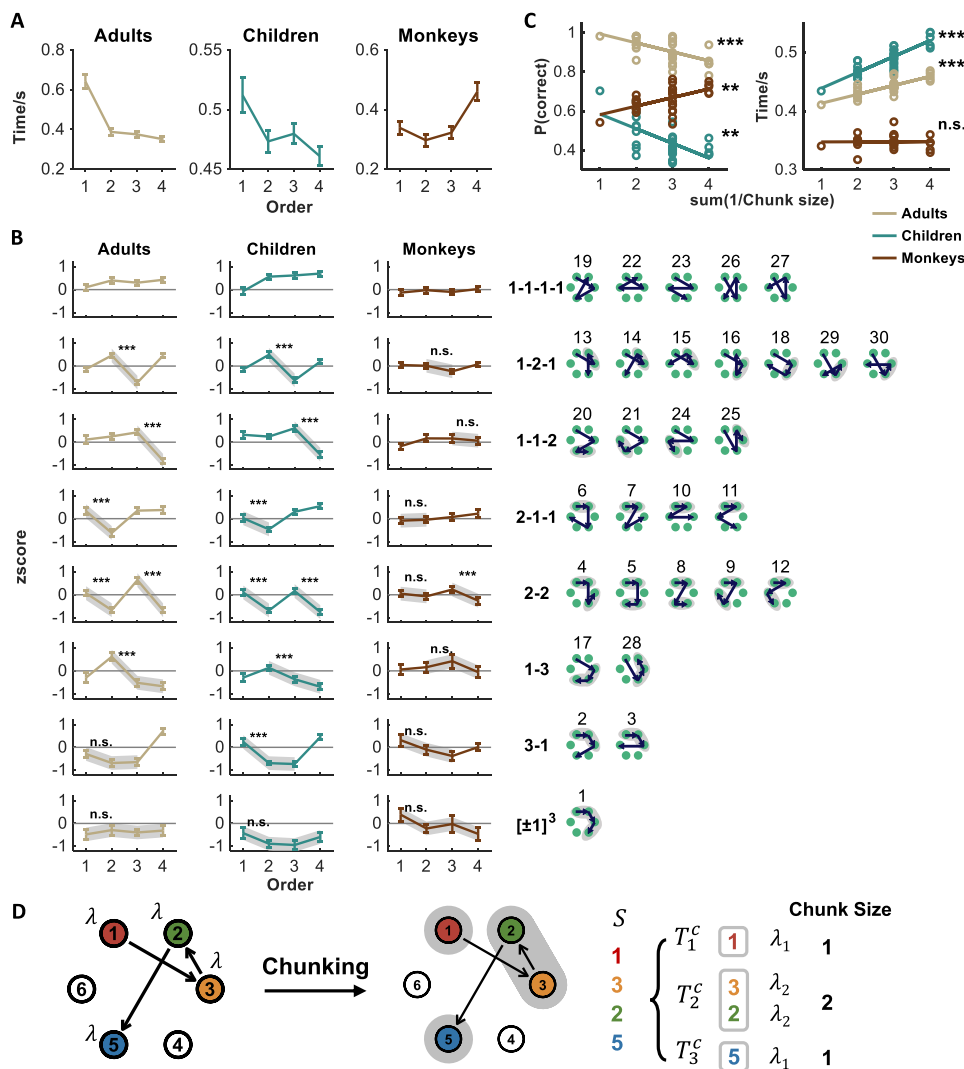


Figure 4. Sequence compression in humans. **A**, RT of each ordinal position from the correct trials. Error bars indicate the SEM across participants (in adults and children) or sessions (in monkeys). **B**, The eight chunking modes (for details, see Materials and Methods) and corresponding normalized RT. The RTs of each ordinal position were transformed to z scores with the mean and SD of all sequences (left panel). The significance of the pairwise Wilcoxon tests between the first and second item within a chunk are shown. Error bars correspond to the SEM across sequences. Patterns in each chunking mode are listed in the right panel. **C**, Correlation between pattern complexity and task performance of each pattern (left, accuracy of sequence reproduction; right, mean RTs). The complexity of each pattern was measured according to chunk size ($\sum_i^n \frac{1}{k_i}$, where k_i is the size of the chunk that contained the i th item, and n is the sequence length). The accuracy of sequence reproduction (left) and mean RTs (right) in humans (adults: khaki; children: cyan) was significantly predicted by sequence complexity. In contrast, the accuracy of monkeys (brown) showed a significant positive correlation with the complexity. The RT used in the analysis was the average of all touches in the correct trials. For each group, a regression line is plotted, and the significance of Spearman’s correlation is shown ($n = 30$). **D**, Illustration of the chunk-based conjunctive coding model. Chunks were defined based on the gestalt principles of proximity and similarity, and spatially and temporally adjacent items were put into the same chunk. The model assumed that chunking processing would improve the order precision (λ). The precision of the temporal order for each target was determined by the chunk size (the number of targets in a chunk). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. n.s., Nonsignificant.

The AICs and BICs averaged across $100 \times 3 = 300$ validations are shown in Table 1.

Data availability

The datasets and the software code that support the findings of this study are available from the corresponding author on reasonable request.

Results

Subjects (40 adults, 154 children, and 41 sessions from two macaque monkeys; for details, see Materials and Methods) were engaged in a sequence reproduction task. On each trial, spatial sequences with a length of 3, 4, 5, or 6 elements (adults: length-4, length-5, and length-6; children: length-4; monkeys: length-3 and length-4) were visually presented. Each element of the sequence was drawn (without replacement) from one of the six

spatial locations of a hexagon. Participants had to reproduce the sequences by successively touching the appropriate location on the screen (Fig. 1A; for details see Materials and Methods). Feedback (reward) was given to monkeys after correct completion of each sequence.

Behavioral benchmarks of sequence memory

We first identified some behavioral benchmarks of sequence memory (Oberauer et al., 2018) in the sequence reproduction task in the adults, children, and macaque monkeys. Given that all groups had very high accuracy for the length-3 sequences, and that there was a limit of memory capacity in children and monkeys, we mainly focused on length-4 sequences. Length-5 and length-6 sequences were only used in adults.

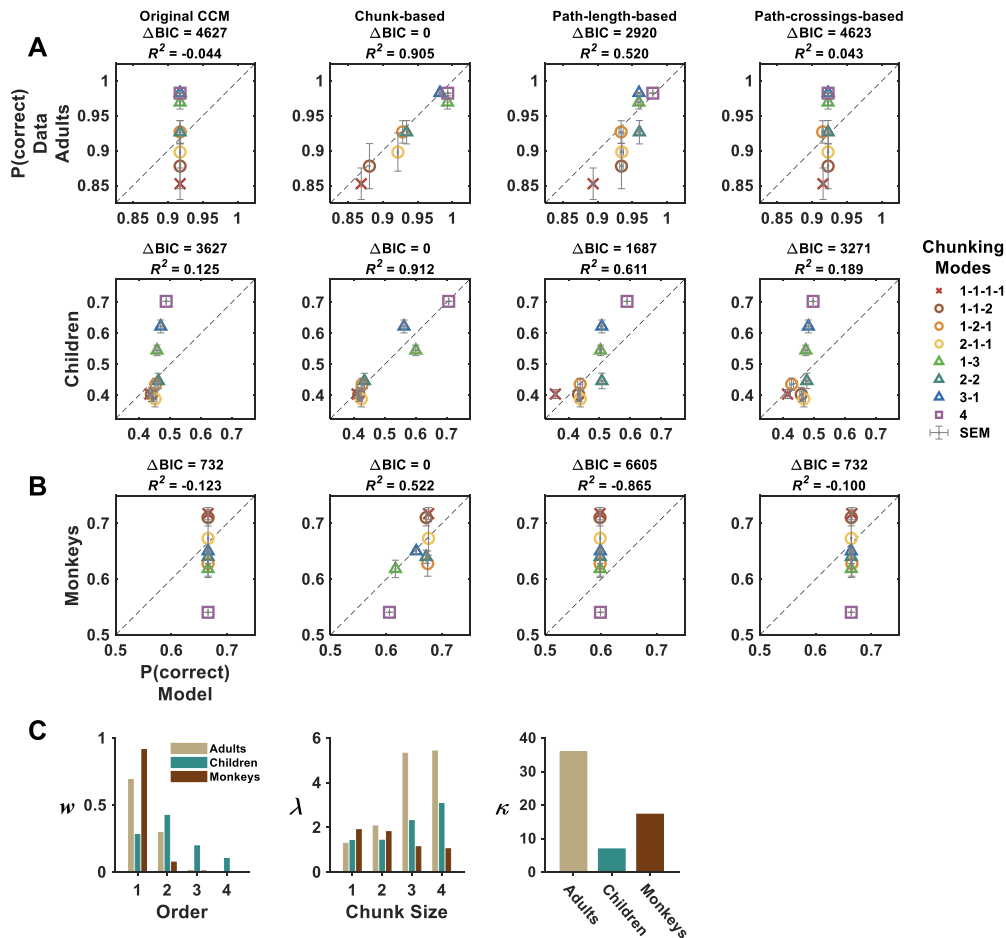


Figure 5. Chunk-based conjunctive coding model and model comparisons. **A**, In both children and adults, the chunk-based conjunctive coding model outperformed the other models, including the conjunctive coding model, the path length-based model, and the path-crossings model. Error bars indicate SEM across patterns. **B**, The chunk-based model in monkeys showed the opposite prediction profile that chunking modes containing larger chunking size displayed worse behavioral performance. **C**, The weight (w) of each ordinal position, the precision of temporal order (λ), and the spatial location (κ) of each chunk size fitted from the chunk-based conjunctive coding model in the three groups. All of the above results used the medians of the best fitting parameters from 1000 bootstrap resamples.

There were several commonalities among the three groups. The sequence accuracy of adults and monkeys showed a typical length effect, whereby an increased sequence load resulted in a decreased recall accuracy (Fig. 1B). There is an advantage for items presented at the start of the sequence (the primacy effect) and at the end of the sequence (the recency effect); thus, plotting recall accuracy by serial position typically results in a “bow-shaped” curve [effect of ordinal position: adults: length-4 sequences (Friedman test), $\chi^2(3) = 46.268$, $p < 0.001$, Kendall’s $W = 0.386$; length-5 (Friedman test), $\chi^2(4) = 64.675$, $p < 0.001$, Kendall’s $W = 0.404$; length-6 (Friedman test), $\chi^2(5) = 49.270$, $p < 0.001$, Kendall’s $W = 0.246$; children: length-4 (Friedman test), $\chi^2(3) = 110.282$, $p < 0.001$, Kendall’s $W = 0.270$; monkeys: length-3 (Friedman test), $\chi^2(2) = 20.000$, $p < 0.001$, Kendall’s $W = 1$; length-4 (Friedman test), $\chi^2(3) = 102.422$, $p < 0.001$, Kendall’s $W = 0.898$]. Almost all three groups displayed this profile for behavioral results (Fig. 1B): the primacy effect was found in all the groups (planned pairwise comparisons with Bonferroni’s correction, first vs second item: adults: length-4, $p = 0.021$, Cohen’s $d = 0.094$; length-5, $p < 0.001$, Cohen’s $d = 0.327$; length-6, $p < 0.001$, Cohen’s $d = 0.236$; children: length-4, $p < 0.001$, Cohen’s $d = 0.234$; monkeys: length-3, $p = 0.004$, Cohen’s $d = 0.571$; length-4, $p < 0.001$, Cohen’s $d = 0.974$), but, interestingly, the recency

Table 1. Model comparison after cross-validations

Model	Adults			Children			Monkeys		
	k	ΔAIC	ΔBIC	k	ΔAIC	ΔBIC	k	ΔAIC	ΔBIC
Conjunctive coding	7	2888	2866	7	2692	2670	7	584	562
Chunk based	10	0	0	10	0	0	10	0	0
Path length based	9	1025	1018	9	1200	1192	9	1493	1486
Path crossings based	8	2911	2896	8	2422	2407	8	279	264

The AICs and BICs shown were averaged across $100 \times 3 = 300$ validations. k , Number of parameters.

effect was almost absent in monkeys (planned pairwise comparisons with Bonferroni’s correction: monkeys: length-3, second greater than third item, $p = 0.006$, Cohen’s $d = 0.866$; length-4, third greater than fourth item, $p = 0.002$, Cohen’s $d = 0.482$). Furthermore, when an item was recalled at an incorrect serial position, its recall spatial location was likely to lie near its original position, and its recall order was more likely to swap with its neighbor orders, which is called a transposition gradient. We found that the error distributions in all three groups displayed transposition gradients for both temporal order (Fig. 1C) and spatial location (Fig. 1D).

Extraction of relational structures in humans, but not macaque monkeys

Sequences can be encoded not just by their spatial locations but also by their relational structures between locations. We next

examined whether monkeys and humans were sensitive to such relations. In the task, each sequence item could be at one of six spatial locations, resulting in a large number of combinations. For length-4 sequences, a total of 360 sequences was included, given that each location was only sampled once. Based on the sequential geometrical relationships among the items, the sequences can be categorized into 30 patterns (Fig. 2A,B). For example, the sequences “1234,” “2345,” “6543,” and “2165” share the same relational structure—repeat a one-step movement three times—which was termed pattern 1. Visualization of the spatial structures of the 30 patterns demonstrated different spatial organizations and complexities of these geometrical relationships (Fig. 2B).

We then asked whether the three groups could spontaneously extract these spatial patterns and use this information to inform sequence encoding (e.g., using the relational structures between locations to encode sequences in a more succinct form; Amalric et al., 2017; Wang et al., 2019; Al Roumi et al., 2021). Note that during either training in monkeys or behavioral testing in humans, there was no explicit instruction to use such spatial patterns. Therefore, if the subject indeed spontaneously learned these structures, we could expect to observe a similar task performance for sequences that shared the same relational pattern, and a substantial performance difference between sequences with distinct relational patterns. The results showed a double dissociation between humans and monkeys. In adults and children, there were no significant differences in accuracy among the 12 sequences within each pattern [Fig. 2C; 30 patterns, corrected for multiple comparisons; adults (Kruskal–Wallis test): p values > 0.270 , $\eta^2 < 0.138$; children (Kruskal–Wallis test): p values > 0.087 , $\eta^2 < 0.051$], but there was a significant difference in accuracy among the patterns (Fig. 2D; Friedman test; adults: $\chi^2(29) = 86.043$, $p < 0.001$, Kendall's $W = 0.247$; children: $\chi^2(29) = 145.504$, $p < 0.001$, Kendall's W). By contrast, in monkeys, there was no difference among the patterns (Fig. 2D; Friedman test; M1: $\chi^2(29) = 16.917$, $p = 0.964$, Kendall's W ; M2: $\chi^2(29) = 46.036$, $p = 0.023$, Kendall's W), and significant differences in accuracy among sequences within patterns (Fig. 2C; Kruskal–Wallis test; M1: p values < 0.035 , $\eta^2 > 0.070$, 23 of 30 patterns; M2: p values < 0.027 , $\eta^2 > 0.141$, 29 of 30). As the trial number was different in each sequence pattern across the three groups, we additionally compared their effect sizes. The result supported the difference in the within-pattern effects between humans and monkeys [95% CI of the effect size (η^2): adults, [0.006, 0.043]; children, [0.002, 0.0135]; M1, [0.095, 0.145]; M2, [0.211, 0.290]]. When further examining the sequence differences within patterns in monkeys, we found that they were mainly because of the biases of spatial location but not sequence direction (clockwise and counterclockwise; spatial location of the starting point, Friedman test; M1: $\chi^2(5) = 50.186$, $p < 0.001$, Kendall's $W = 0.335$; M2: $\chi^2(5) = 51.733$, $p < 0.001$, Kendall's $W = 0.345$. Sequence direction, Friedman test; M1: $\chi^2(1) = 0.006$, $p = 0.940$, Kendall's $W < 0.001$; M2: $\chi^2(1) = 2.222$, $p = 0.136$, Kendall's $W = 0.012$). Such biases were not present in the performance of adults and children [spatial location of the starting point (Friedman test): adults: $\chi^2(5) = 6.332$, $p = 0.275$, Kendall's $W = 0.042$; children: $\chi^2(5) = 7.257$, $p = 0.202$, Kendall's $W = 0.048$; sequence direction (Friedman test): adults: $\chi^2(1) = 0.659$, $p = 0.417$, Kendall's $W = 0.004$; children: $\chi^2(1) = 0.360$, $p = 0.549$, Kendall's $W = 0.012$]. We then removed the patterns that contained the biased locations and recalculated the within-pattern and between-pattern differences in both monkeys. The result confirmed the double dissociation between humans and monkeys [difference between the patterns (Friedman test): adults: $\chi^2(16) = 54.849$, $p < 0.001$, Kendall's $W = 0.286$; children:

$\chi^2(16) = 99.903$, $p < 0.001$, Kendall's $W = 0.520$; M1: $\chi^2(16) = 9.633$, $p = 0.885$, Kendall's $W = 0.050$; M2: $\chi^2(16) = 25.888$, $p = 0.056$, Kendall's $W = 0.135$; difference within the patterns (Kruskal–Wallis test); adults: p values > 0.153 , $\eta^2 < 0.138$; children: p values > 0.050 , $\eta^2 < 0.051$ (not significant in 10 of 17 patterns); M1: p values < 0.028 , $\eta^2 > 0.047$ (significant in 14 of 17 patterns); M2: p values < 0.015 , $\eta^2 > 0.096$ (significant in 15 of 17 patterns)], as shown in Figure 2, C and D.

However, we should notice that the comparison between humans and monkeys was based on pooling human participants and monkey behavioral sessions. To test whether the within-pattern effect could also be found on a participant-by-participant basis, we additionally recruited six human adults, who were asked to perform 3600 trials within 10 d (see Materials and Methods). We found that the lack of within-pattern difference was highly consistent in individual human participants [Fig. 2E; 30 patterns in each participant, corrected for multiple comparisons (Kruskal–Wallis test): p values > 0.334 , $\eta^2 < 0.133$; except in one pattern in one participant: p values = 0.027, $\eta^2 = 0.214$].

Did human adults and children implement a similar strategy or language to detect the complexities of the 30 patterns? We plotted the behavioral performance of the three groups in sequences of all 30 patterns in descending order of accuracy in children (i.e., highest to lowest; Fig. 2F, dark cyan curve). The performance of adults showed a trend similar to that of children (Fig. 2F, khaki curve), but the performance of the monkeys was entirely different from that of humans (Fig. 2F, brown curve). The statistical analysis confirmed a significant positive correlation in sequence performance across the 30 patterns between adults and children (Fig. 2G; Spearman's $\rho_{(28)} = 0.829$, $p < 0.001$), but not between humans and monkeys (Fig. 2H: adults vs monkeys: $\rho_{(28)} = -0.177$, $p = 0.349$; Fig. 2I: children vs monkeys: $\rho_{(28)} = -0.099$, $p = 0.601$). These results indicate that while adults and children adopted a similar internal language of extracting relational structures during spatial sequence processing, macaque monkeys might lack the ability to spontaneously detect the geometrical structures and use them to compress the sequences in memory.

Fitting data to the conjunctive coding model

As a first attempt to model the performance of the three groups of subjects, including the positional accuracy and transposition gradients in both spatial and ordinal dimensions, we adopted the conjunctive coding model (Botvinick and Watanabe, 2007; Oberauer and Lin, 2017; Fig. 3A; Materials and Methods). The assumption was that the representational code of spatial sequences is a conjunction of approximate codes for the spatial items (e.g., six locations on the hexagon) and their corresponding ordinal positions (e.g., first, second, third, and fourth). This model allowed us to accurately describe representations of individual spatial locations as a scaled von Mises distribution, which is a normal distribution that is appropriate for spatial locations (Eq. 2; Materials and Methods). The six spatial locations were assumed to share a similar distribution in the model. For the ordinal representation, we made no prior assumptions of a compressive code, according to which ordinal tuning curves would broaden with increasing order (Botvinick and Watanabe, 2007). Instead, we described representations of ordinal information using the scaled Laplace distribution (Brown et al., 2007; Eq. 1; Materials and Methods). Finally, we assumed that ordinal information is integrated with spatial information through multiplicative gain modulation, resulting in a conjunctive representation of the sequence in memory (Eq. 3; Materials and Methods). During

the sequence reproduction task, the retrieval probability of each item was conditional, given that each location was sampled only once, without replacement, within a sequence (Eq. 4; Materials and Methods).

The results of model fitting in the three groups replicated the sequence reproduction benchmarks shown in Figure 1. The positional accuracy of the model displayed the same “bow-shaped” curve in humans (Fig. 3B). This pattern of performance (primacy and recency effects) stems from interference effects because the probability of exchanging items with near neighbors is lower at the start and end of the sequence. More importantly, the model can reproduce not only the behavioral profile of correct trials, but also the distribution of error responses, by showing the same profile of location and rank transposition gradients as the behavior results in Figure 3, C and D. Items in nearby ordinal or spatial locations are represented more similarly than items at more distant positions, which makes it relatively easy for the model to confuse the locations of closely spaced items in both ordinal and spatial manners.

Although we initially set the ordinal representation as the scaled Laplace distribution, it is worth noting that the fitting results demonstrated a compressive ordinal code in all the three groups (Fig. 3E). That is, the ordinal tuning curves broadened with increasing order. Such a compressive profile in the encoding matrix was reflected by the pattern of the assigning weight (w) of each order; the weights decreased with increasing order (Fig. 3E). The code profile was consistent with previous electrophysiological work in monkeys by Nieder and Miller (2003) and Nieder et al. (2006), which showed that parietal neurons represent count information using a compressive code that is reflected by more broadly tuned receptive fields for larger numbers. Thus, the primacy effect and the increasing of the transposition error along ranks derive, additionally, from the higher precision of orders at the beginning of the sequence, which is driven by the compressive ordinal code of the model.

Despite these similarities in behavioral benchmarks, there were several notable differences among the three groups. First, the overall performance of children (mean \pm SD; $45.01 \pm 21.65\%$) and monkeys ($64.38 \pm 16.69\%$) was much lower than that of adults ($91.24 \pm 7.24\%$; Fig. 3H; Kruskal–Wallis test: $\chi^2(2) = 102.6$, $p < 0.001$; pairwise Wilcoxon rank-sum test with Bonferroni’s correction: adults vs children, $p < 0.001$; adults vs monkeys, $p < 0.001$; children vs monkeys, $p < 0.001$). To exclude the possibility that the poor performance of monkeys and children was because of a lower level of understanding of the task procedure, we examined their performance of length-3 sequences using the same task. We found all three groups of subjects demonstrated very high performance (adults, $99.22 \pm 0.86\%$; children, $72.18 \pm 22.38\%$; monkeys, $84.45 \pm 10.11\%$).

To identify the mechanism underlying the inferior sequence-processing ability in children and monkeys, we examined between-group differences by comparing the precision of spatial location (κ) and temporal order (λ), and the assigned weight on each temporal order (w) in the model. We found that the precision of the temporal order (λ) of children and monkeys is significantly lower than that of human adults, and there were no significant differences between children and monkeys. Meanwhile, children’s precision of spatial location (κ) was significantly lower than that in human adults and monkeys, and there were no significant differences between adults and monkeys [Fig. 3F,G; random permutation tests ($N = 1000$), λ : adults vs children: $\hat{p} = 0$, 99.9% CI: [0, 0.0076]; adults vs monkeys: $\hat{p} = 0.001$, 99.9% CI, [0, 0.01]; children vs monkeys: $\hat{p} = 0.874$, 99.9% CI, [0.8361, 0.9061]; κ : adults vs

children: $\hat{p} = 0$, 99.9% CI, [0, 0.0076]; adults vs monkeys, $\hat{p} = 0.257$, 99.9% CI, [0.213, 0.3047]; children vs monkeys: $\hat{p} = 0$, 99.9% CI, [0, 0.0076]). Even with extensive long-term training (>2 years), the precision of temporal order in monkeys only reached the same level as that of children, who were completely naive to the spatial sequences.

Furthermore, the curve of assigned weights (w) along with the ordinal ranks in monkeys was much steeper than that seen in adults and children (Fig. 3E). This may suggest that, compared with humans, monkeys reallocated most resources to the first item (almost 100%) and much less to the other items. This profile of weight assigning in w that is small enough for monkeys, and the background noise (η ; for details, see Materials and Methods) becomes important and cannot be ignored. Therefore, multiple factors, including the interference effect, small w , and the background noise, caused the dramatically decreased recall accuracy along with the ordinal position and the absence of recency effect in monkeys (Fig. 1B).

Chunking as an internal algorithm for sequence compression

Although the conjunctive coding model can account for the positional accuracy and transpositional gradients in both spatial and ordinal dimensions, the model failed to explain the variance of the performance between the sequence patterns (Fig. 3I). What is the internal format used by humans to compress spatial sequence processing and memory? What algorithm can explain the observed variations in working memory for the 30 sequence patterns? Previously, we showed that human adults and preschoolers can quickly grasp a “geometrical language” endowed with simple primitives of symmetries and rotations, and combinatorial rules in an eight-item spatial sequence, and that they use this internal language to predict the next item of a sequence (Amalric et al., 2017).

To identify potential primitives or rules for the length-4 sequences, we first examined the RTs for each item during the sequence production of the three groups. There was a similar pattern in RTs averaged over all sequences between human adults (Friedman test: $\chi^2(3) = 56.550$, $p < 0.001$, Kendall’s $W = 0.471$; planned pairwise comparisons with Bonferroni’s correction: first vs second item: $p < 0.001$, Cohen’s $d = 1.465$; second vs third item: $p = 0.062$, Cohen’s $d = 0.118$; third vs fourth item: $p < 0.001$, Cohen’s $d = 0.267$) and children (Friedman test: $\chi^2(3) = 15.062$, $p = 0.001$, Kendall’s $W = 0.037$; planned pairwise comparisons with Bonferroni’s correction: first vs second item: $p = 0.160$, Cohen’s $d = 0.276$; second vs third item: $p = 0.589$, Cohen’s $d = 0.062$; third vs fourth item: $p < 0.001$, Cohen’s $d = 0.191$), whereby there were shorter RTs for each subsequent item in a sequence, previously referred as a “collective search” (Fig. 4A; Ohshiba, 1997; Conway and Christiansen, 2001), which may indicate that humans use an internal forward model to compress items within a sequence into an integrated chunk or unit. Conversely, the RTs of monkeys show a different trend, with similar RTs for the first two items and then longer RTs for each subsequent item (Friedman test: $\chi^2(3) = 41.053$, $p < 0.001$, Kendall’s $W = 0.360$; pairwise comparisons with Bonferroni’s correction: first vs second item: $p > 0.999$, Cohen’s $d = 0.318$; second vs third item: $p = 0.002$, Cohen’s $d = 0.206$; third vs fourth item: $p < 0.001$, Cohen’s $d = 0.863$), which indicates that they might have used a different strategy of “serial search” in working memory (Fig. 4A). That is, monkeys retrieved the first item, touched it on the screen, then retrieved the next item, touched it on the screen, and so on.

As a further attempt to capture the two different search strategies used by humans and monkeys, we used a simple algorithm—spatial chunking—which was based on the gestalt principles of proximity and similarity, whereby only spatially and temporally adjacent items were chunked together. The 30 patterns were thus divided into eight groups according to the size of their consecutive chunks (Fig. 4B, right; e.g., “2-2,” two consecutive chunks of size 2, including patterns 4, 5, 8, 9, and 12). We then plotted the RTs of the eight modes individually (Fig. 4B, left). This revealed decreasing RTs for items within chunks (marked by gray zones) in both adults and children, but not in monkeys (Fig. 4B, left). This finding indicates that humans use a generalized strategy across different patterns that collectively chunk spatially and temporally close items within sequences, while monkeys may only learn to chunk in a subset of sequences but fail to generalize across patterns. To examine whether the performance of subjects was reflected by chunking, we defined the complexity of a sequence using the average chunk sizes for each pattern (i.e., the sequence 1234 has one length-4 chunk, and the sequence 1352 has four length-1 chunks), whereby a bigger chunking size within a sequence was considered to result in a lower sequence complexity and easier memory compression. We found that sequence reproduction accuracy and RTs in adults and children were well predicted by chunk size (Fig. 4C; adults: Spearman’s $\rho_{(28)} = -0.592$, $p < 0.001$; children: Spearman’s $\rho_{(28)} = -0.522$, $p = 0.003$; RT: adults: Spearman’s $\rho_{(28)} = 0.767$, $p < 0.001$; children: Spearman’s $\rho_{(28)} = 0.828$, $p < 0.001$). In contrast, the performance of monkeys was positively correlated with chunk size (Fig. 4C; Spearman’s $\rho_{(28)} = 0.539$, $p = 0.002$). That is, the sequence with the biggest chunk size (i.e., sequence 1234) was associated with the worst sequence production. This could indicate the presence of the interference effect in the conjunctive coding model; for monkeys, while the spatially and temporally close locations within a sequence were not efficiently integrated into chunks, these locations heavily interfered with each other, resulting in a high error rate of sequence reproduction for both spatial and temporal dimensions. As shown in Figure 5C, the precision of order (λ) decreased with increasing chunk size in monkeys, which agreed with the stronger interference between spatially and temporally close items in larger chunks.

We also examined whether the children who benefit more from a spatial chunking strategy had better results at school. The average scores of children’s mathematics and Chinese examinations ~2 months after test sessions were used as an index of examination performance. We divided sequences into two categories, depending on whether chunking strategies were involved in sequence reproduction. We found that, unlike the use of root memory in the sequence task (the group 1-1-1-1: Spearman’s $\rho_{(131)} = 0.172$, $p = 0.06$), the task performance of the sequences using the chunking strategy (other groups except Fig. 4B, group 1-1-1-1) was significantly correlated with children’s examination score (Spearman’s $\rho_{(131)} = 0.202$, $p = 0.025$; see Materials and Methods).

Finally, to explain the variance of task performance at the relational structure level, we added the component of pattern complexity (chunk size) to our basic conjunctive coding model by recalculating the precision (λ) of each temporal order based on the chunk sizes in a sequence (Eq. 5; Materials and Methods). The assumption was that chunking improves the precision of ordinal coding. We fitted the model to our behavioral data; while the conjunctive coding model could predict well the behavioral responses of both correct and incorrect responses (positional accuracy and transposition gradients) and explained the sequence

variance solely by the interference effect, the chunk-based conjunctive coding model explained significantly more variance at relational structure levels in human adults and children (Fig. 5A). Indeed, as predicted, while the distribution pattern of weights (w) on each ordinal did not change, the precision of temporal order predicted by the model increased along with the chunking size in both adults and children (Fig. 5C). In contrast, the chunk-based model in monkeys showed the opposite prediction (Fig. 5B), whereby the chunking modes with a larger chunking size was associated with a worse behavioral performance, which is consistent with the correlation analysis shown in Fig. 4C. Furthermore, we compared the efficacy of the chunk-based model with that of a simpler model by which the precision of temporal order was modulated according to spatial crossing or total sequence path; these two factors have been proposed as a measurement of spatial sequence complexity (De Lillo et al., 2016). The chunk-based model significantly outperformed the path-length or crossing-based models (see Materials and Methods; Eqs. 6, 7; Fig. 5A, Table 1, model comparison).

Discussion

The current study examined the computational mechanisms underlying sequence representation in adults, children, and macaque monkeys with a common sequence reproduction task, and used conjunctive coding models to assess the between-group differences in behavioral measures. We found the following (1) the precision of spatial location and of temporal order were the main factors contributing to the poor performance of sequence processing in children and monkeys; (2) even with long-term training, macaque monkeys demonstrated a strategic limitation of resource reallocation along the ordinal ranks; (3) compared with human subjects (adults and children), who used a common internal format for sequence representation, macaque monkeys lacked the ability to spontaneously detect spatial relational structures; and (4) while spatiotemporal interference could explain the behavior of correct and error responses, human behavior at structural level required the conjunctive coding using chunking as the internal algorithm. Our data thus provide a direct assessment of the relative contributions of development and evolution to sequence representation in humans, which could also have implications for uniquely human cognitive capacities (e.g., language processing).

Our observation of differences in temporal precision between human adults and children is consistent with those of previous studies that have proposed that the learning of neural representation of temporal order continues to develop over early and middle childhood (Lipton and Spelke, 2003; Loucks and Price, 2019). Our results also expand on prior reports by showing that both spatial and temporal accuracies were both low in monkeys, which was not because of a lack of behavioral training. In addition, our results indicate that monkeys reallocated almost all of their attentional resources to the first item, whereas humans use a more balanced reallocation strategy for each item. The intrinsic limit of temporal precision combined with this extreme strategy of resource reallocation in monkeys was one of the reasons explaining the between-species difference in cognitive capacity and inductive learning of retaining and updating sequential information in working memory.

Little work has examined how spatial sequences are encoded and retrieved in humans and animals, or whether and how a model can predict each item during the sequence reproduction.

Previous studies have investigated cross-species differences in pattern identification and found that humans use a more global perception. Specifically, humans have an advantage over monkeys in grouping visual information into global shapes (Fagot and Deruelle, 1997; Parron and Fagot, 2007; Spinozzi et al., 2009; Neiworth et al., 2014). In acquiring a nonlanguage grammatical structure, monkeys have weaker capability compared with humans (Fitch and Hauser, 2004; Saffran et al., 2008; Wang et al., 2015; Jiang et al., 2018). For example, monkeys can be trained to produce sequences with supragrammars, but the learning is much slower than for preschool children (Jiang et al., 2018). A recent study has shown that humans can use recursive hierarchical strategies in a nonlinguistic sequence generation task early in development, while monkeys did so only with additional exposure (Ferrigno et al., 2020). Despite these behavioral studies, none of them has examined the computational mechanisms underlying the group differences. At the structure level of spatial sequences, we showed that humans, but not monkeys, displayed significant differences in accuracy and reaction time between patterns, indicating that humans, but not monkeys, are able to spontaneously detect spatial regularities and encode the sequence in memory. The difference in pattern complexities was mainly because of the chunk strategy used in both adults and children. However, we did not tend to conclude that chunking was the only human-specific strategy, because the sequences used in the current study were too short and too simple to assess the possible use of other, even higher, levels of sequence encoding (Dehaene et al., 2015), and therefore, to test the predictions of other measures of sequence complexity such as language of thought (Fodor, 1975) and entropy (Kamae and Zamboni, 2002). In previous work, using a longer eight-item spatial sequence, we demonstrated that adults and preschoolers could spontaneously grasp a “geometrical language” endowed with several simple primitives of symmetry and rotation, as well as recursive combinatorial rules (Amalric et al., 2017). In the future, the present task may allow testing of this model in monkeys as well. One hypothetical suggestion from our comparative study is that monkeys only focus on the individual locations and fail to spontaneously learn any kind of spatial relational structures linking them (Fagot and Deruelle, 1997; Parron and Fagot, 2007; Spinozzi et al., 2009; Neiworth et al., 2014). Here, the failure to learn such regularities was not because of a lack of training, as the two monkeys were trained with hundreds of thousands of trials over >2 years. Behavioral analyses and the conjunctive coding model suggested that children outperformed monkeys in using global geometric structure and chunking to compress the sequence spontaneously, although on average, they showed a similarly poor sequence reproduction performance.

The difference in behavioral performance between humans (adults and children) and monkeys cannot be interpreted by other experimental accounts. For example, one may argue that humans are more familiar or have more prior experience with the geometrical layouts than monkeys, which may therefore have higher possibilities for grasping abstract patterns. This seems unlikely, as monkeys have been habituated with the spatial sequences with different patterns for years and many trials (>600) in every training day. Furthermore, previous behavioral studies have indicated that infants, without much prior experience, already possess a capacity to quickly grasp abstract sequence patterns in the first days of life (Dehaene-Lambertz et al., 2002). The other confounding issue could be memory capacity or attention level between humans and monkeys. This could be easily excluded, as children and monkeys may share similar working memory capacity (Cowan, 2001; Buschman et al., 2011;

Heyselaar et al., 2011; Lara and Wallis, 2012; Simmering, 2012), but their performance of learning abstract patterns was significantly different. Also, differences in the task design, such as intertarget delays (ITDs) or stimulus onset asynchronies (SOAs), were unlikely to account for our main observations. The two monkeys were tested with different SOAs but did not differ in their strategies. The presentation duration used in the present study (>250 ms) was also of the range (50–100 ms for a single item) in which performance was enhanced with increased presentation duration (Vogel et al., 2006; Bays et al., 2011). In addition, longer intertarget intervals could lead to better performance of memory tasks (Neath and Crowder, 1990, 1996; Guérard et al., 2010), while in the present study, monkeys were presented with a longer ITD but showed a worse memory performance than humans. Finally, the learning strategy may differ between groups, as the training of the monkey is involved in complicated procedures. It is worth noting that the current study tested the spontaneous learning of abstract pattern in both humans and monkeys. The task requirement, which is repeating sequences, is orthogonal to the learning of geometrical regularities within the sequence.

However, we cannot exclude that monkeys would eventually be able to learn relational structures and chunking as strategies to process spatial sequences, if given certain feedback using reinforcement learning algorithms and with intensive training, or that such ability to use chunking strategy is qualitative or quantitative (Minier et al., 2016; Heimbauer et al., 2018; Jiang et al., 2018; Rey et al., 2019; Tosatto et al., 2021). It also has been demonstrated that monkeys could use chunking in other domains (e.g., motor sequences; Fujii and Graybiel, 2003; Ramkumar et al., 2016). Yet, most of the behavioral studies showing that animals could learn abstract rules or structures also demonstrated a long-time and intensive training requirement for task learning (Fujii and Graybiel, 2003; Minier et al., 2016; Ramkumar et al., 2016; Heimbauer et al., 2018; Rey et al., 2019; Tosatto et al., 2021). Therefore, our comparative observations may suggest that the difference in sequence processing between humans and other animals may depend on both human-specific neural circuitries (e.g., temporal–frontal language neural network) and specific structure-sensitive learning algorithms, rather than the mere memory capacity. It seems that only humans could use these algorithms to represent the world in a non-task-specific way. However, monkeys may still rely heavily on the reward as a reinforcer, which requires too many samples for training. Future research should examine the neural mechanisms underlying spontaneous pattern learning to test whether these sequence-processing tasks involve a universal attention or working memory circuitry, including dorsal visuospatial network or human-unique language regions (Wang et al., 2019).

References

- Al Roumi F, Marti S, Wang L, Amalric M, Dehaene S (2021) Mental compression of spatial sequences in human working memory using numerical and geometrical primitives. *Neuron* 109:2627–2639.
- Amalric M, Wang L, Pica P, Figueira S, Sigman M, Dehaene S (2017) The language of geometry: fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLoS Comput Biol* 13:e1005273.
- Barone P, Joseph J-P (1989) Prefrontal cortex and spatial sequencing in macaque monkey. *Exp Brain Res* 78:43–54.
- Bays PM, Gorgoraptis N, Wee N, Marshall L, Husain M (2011) Temporal dynamics of encoding, storage, and reallocation of visual working memory. *J Vis* 11(10):6, 1–15.
- Botvinick M, Watanabe T (2007) From numerosity to ordinal rank: a gain-field model of serial order representation in cortical working memory. *J Neurosci* 27:8636–8642.

- Botvinick MM, Wang J, Cowan E, Roy S, Bastianen C, Patrick Mayo J, Houk JC (2009) An analysis of immediate serial recall performance in a macaque. *Anim Cogn* 12:671–678.
- Brady TF, Konkle T, Alvarez GA (2009) Compression in visual working memory: using statistical regularities to form more efficient memory representations. *J Exp Psychol Gen* 138:487–502.
- Brown GDA, Neath I, Chater N (2007) A temporal ratio model of memory. *Psychol Rev* 114:539–576.
- Buschman TJ, Siegel M, Roy JE, Miller EK (2011) Neural substrates of cognitive capacity limitations. *Proc Natl Acad Sci U S A* 108:11252–11255.
- Chase WG, Ericsson KA (1982) Skill and working memory. *Psychol Learn Mem* 16:1–58.
- Conway CM, Christiansen MH (2001) Sequential learning in non-human primates. *Trends Cogn Sci* 5:539–546.
- Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 24:87–114.
- De Lillo C, Kirby M, Poole D (2016) Spatio-temporal structure, path characteristics, and perceptual grouping in immediate serial spatial recall. *Front Psychol* 7:1686.
- 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* 88:2–19.
- Dehaene-Lambert G, Dehaene S, Hertz-Pannier L (2002) Functional neuroimaging of speech perception in infants. *Science* 298:2013–2015.
- Endress AD, Nespore M, Mehler J (2009) Perceptual and memory constraints on language acquisition. *Trends Cogn Sci* 13:348–353.
- Ericsson K, Chase W, Faloon S (1980) Acquisition of a memory skill. *Science* 208:1181–1182.
- Fagot J, De Lillo C (2011) A comparative study of working memory: immediate serial spatial recall in baboons (*Papio papio*) and humans. *Neuropsychologia* 49:3870–3880.
- Fagot J, Deruelle C (1997) Processing of global and local visual information and hemispheric specialization in humans (*Homo sapiens*) and baboons (*Papio papio*). *J Exp Psychol Hum Percept Perform* 23:429–442.
- Farrell Pagulayan K, Busch RM, Medina KL, Bartok JA, Krikorian R (2006) Developmental normative data for the Corsi Block-Tapping task. *J Clin Exp Neuropsychol* 28:1043–1052.
- Feldman J (2000) Minimization of Boolean complexity in human concept learning. *Nature* 407:630–633.
- Ferrigno S, Cheyette SJ, Piantadosi ST, Cantlon JF (2020) Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians. *Sci Adv* 6:1–11.
- Fitch WT, Hauser MD (2004) Computational constraints on syntactic processing in a nonhuman primate. *Science* 303:377–380.
- Fodor JA (1975) *The language of thought*, Ed 1. Cambridge, MA: Harvard UP.
- Fujii N, Graybiel AM (2003) Representation of action sequence boundaries by macaque prefrontal cortical neurons. *Science* 301:1246–1249.
- Funahashi S, Inoue M, Kubota K (1997) Delay-period activity in the primate prefrontal cortex encoding multiple spatial positions and their order of presentation. *Behav Brain Res* 84:203–223.
- Gilchrist AL, Cowan N, Naveh-Benjamin M (2008) Working memory capacity for spoken sentences decreases with adult ageing: recall of fewer but not smaller chunks in older adults. *Memory* 16:773–787.
- Graybiel AM (1998) The basal ganglia and chunking of action repertoires. *Neurobiol Learn Mem* 70:119–136.
- Guérard K, Neath I, Surprenant AM, Tremblay S (2010) Distinctiveness in serial memory for spatial information. *Mem Cogn* 38:83–91.
- Hauser MD, Watumull J (2017) The Universal Generative Faculty: the source of our expressive power in language, mathematics, morality, and music. *J Neurolinguistics* 43:78–94.
- Heimbauer LA, Conway CM, Christiansen MH, Beran MJ, Owren MJ (2018) Visual artificial grammar learning by rhesus macaques (*Macaca mulatta*): exploring the role of grammar complexity and sequence length. *Anim Cogn* 21:267–284.
- Heyselaer E, Johnston K, Paré M (2011) A change detection approach to study visual working memory of the Macaque monkey. *J Vis* 11(3):11, 1–10.
- Inoue M, Mikami A (2006) Prefrontal activity during serial probe reproduction task: encoding, mnemonic, and retrieval processes. *J Neurophysiol* 95:1008–1041.
- Jiang X, Long T, Cao W, Li J, Dehaene S, Wang L (2018) Production of supra-regular spatial sequences by macaque monkeys. *Curr Biol* 28:1851–1859.e4.
- Kamae T, Zamboni L (2002) Sequence entropy and the maximal pattern complexity of infinite words. *Ergod Theory Dyn Syst* 22:1191–1199.
- Kermadi I, Joseph JP (1995) Activity in the caudate nucleus of monkey during spatial sequencing. *J Neurophysiol* 74:911–933.
- Kermadi I, Jurquet Y, Arzi M, Joseph JP (1993) Neural activity in the caudate nucleus of monkeys during spatial sequencing. *Exp Brain Res* 94:352–356.
- Lara AH, Wallis JD (2012) Capacity and precision in an animal model of visual short-term memory. *J Vis* 12(3):13, 1–12.
- Lipton JS, Spelke ES (2003) Origins of number sense: large-number discrimination in human infants. *Psychol Sci* 14:396–401.
- Loucks J, Price HL (2019) Memory for temporal order in action is slow developing, sensitive to deviant input, and supported by foundational cognitive processes. *Dev Psychol* 55:263–273.
- Luce RD (1959) *Individual choice behavior*. New York: Wiley.
- Marcus GF, Vijayan S, Bandi Rao S, Vishton PM (1999) Rule learning by seven-month-old infants. *Science* 283:77–80.
- Martin N, Gupta P (2004) Exploring the relationship between word processing and verbal short-term memory: evidence from associations and disassociations. *Cogn Neuropsychol* 21:213–228.
- McCormack T, Brown GDA, Vousden JI, Henson RNA (2000) Children's serial recall errors: implications for theories of short-term memory development. *J Exp Child Psychol* 76:222–252.
- Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 63:81–97.
- Minier L, Fagot J, Rey A (2016) The temporal dynamics of regularity extraction in non-human primates. *Cogn Sci* 40:1019–1030.
- Neath I, Crowder RG (1990) Schedules of Presentation and Temporal Distinctiveness in Human Memory. *J Exp Psychol Learn Mem Cogn* 16:316–327.
- Neath I, Crowder RG (1996) Distinctiveness and Very Short-term Serial Position Effects. *Memory* 4:225–242.
- Neiwirth JJ, Whillock KM, Kim SH, Greenberg JR, Jones KB, Patel AR, Steffel-Moore DL, Shaw AJ, Rupert DD, Gauer JL, Kudura AG (2014) Gestalt principle use in college students, children with autism, toddlers (*Homo sapiens*), and cotton top tamarins (*Saguinus oedipus*). *J Comp Psychol* 128:188–198.
- Nieder A, Miller EK (2003) Coding of cognitive magnitude. *Neuron* 37:149–157.
- Nieder A, Diester I, Tudusciuc O (2006) Temporal and spatial enumeration processes in the primate parietal cortex. *Science* 313:1431–1435.
- Ninokura Y, Mushiaki H, Tanji J (2003) Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *J Neurophysiol* 89:2868–2873.
- Ninokura Y, Mushiaki H, Tanji J (2004) Integration of temporal order and object information in the monkey lateral prefrontal cortex. *J Neurophysiol* 91:555–560.
- Nosofsky RM (1986) Attention, similarity, and the identification–categorization relationship. *J Exp Psychol Gen* 115:39–57.
- Oberauer K, Lin HY (2017) An interference model of visual working memory. *Psychol Rev* 124:21–59.
- Oberauer K, Lewandowsky S, Awh E, Brown GDA, Conway A, Cowan N, Donkin C, Farrell S, Hitch GJ, Hurlstone MJ, Ma WJ, Morey CC, Nee DE, Schweppe J, Vergauwe E, Ward G (2018) Benchmarks for models of short-term and working memory. *Psychol Bull* 144:885–958.
- Ohshiba N (1997) Memorization of serial items by Japanese monkeys, a chimpanzee, and humans. *Jpn Psychol Res* 39:236–252.
- Orsini A, Grossi D, Capitani E, Laiacina M, Papagno C, Vallar G (1987) Verbal and spatial immediate memory span: normative data from 1355 adults and 1112 children. *Ital J Neuro Sci* 8:537–548.
- Parron C, Fagot J (2007) Comparison of grouping abilities in humans (*Homo sapiens*) and baboons (*Papio papio*) with the Ebbinghaus illusion. *J Comp Psychol* 121:405–411.
- Pickering SJ, Gathercole SE, Peaker SM (1998) Verbal and visuospatial short-term memory in children: evidence for common and distinct mechanisms. *Mem Cognit* 26:1117–1130.

- Planton S, van Kerkoerle T, Abbih L, Maheu M, Meyniel F, Sigman M, Wang L, Figueira S, Romano S, Dehaene S (2021) A theory of memory for binary sequences: evidence for a mental compression algorithm in humans. *PLoS Comput Biol* 17:e1008598.
- Ramkumar P, Acuna DE, Berniker M, Grafton ST, Turner RS, Kording KP (2016) Chunking as the result of an efficiency computation trade-off. *Nat Commun* 7:12176.
- Rey A, Minier L, Malassis R, Bogaerts L, Fagot J (2019) Regularity extraction across species: associative learning mechanisms shared by human and non-human primates. *Top Cogn Sci* 11:573–586.
- Saffran J, Hauser M, Seibel R, Kapfhamer J, Tsao F, Cushman F (2008) Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* 107:479–500.
- Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–1928.
- Shepard RN (2012) Universal Law of Generalization for Psychological Science Primacy of Generalization Apparent Noninvariance of Generalization. *Adv Sci* 237:1317–1323.
- Shepard RN (1987) Toward a universal law of generalization for psychological science. *Science* 237:1317–1323.
- Simmering VR (2012) The development of visual working memory capacity during early childhood. *J Exp Child Psychol* 111:695–707.
- Spinuzzi G, De Lillo C, Truppa V, Castorina G (2009) The relative use of proximity, shape similarity, and orientation as visual perceptual grouping cues in Tufted Capuchin monkeys (*Cebus apella*) and humans (*Homo sapiens*). *J Comp Psychol* 123:56–68.
- Terrace HS, Mcgonigle B (1994) Memory and representation of serial order by children, monkeys, and pigeons. *Curr Dir Psychol Sci* 3:180–185.
- Tosatto L, Fagot J, Nemeth D, Rey A (2021) The evolution of chunks in sequence learning. *bioRxiv* 430894. doi: 10.1101/2021.02.12.430894.
- Vogel EK, Woodman GF, Luck SJ (2006) The time course of consolidation in visual working memory. *J Exp Psychol Hum Percept Perform* 32:1436–1451.
- Wang L, Uhrig L, Jarraya B, Dehaene S (2015) Representation of numerical and sequential patterns in macaque and human brains. *Curr Biol* 25:1966–1974.
- Wang L, Amalric M, Fang W, Jiang X, Pallier C, Figueira S, Sigman M, Dehaene S (2019) Representation of spatial sequences using nested rules in human prefrontal cortex. *Neuroimage* 186:245–255.