RESEARCH ARTICLE

NEUROSCIENCE

Geometry of sequence working memory in macaque prefrontal cortex

Yang Xie[1]†, Peiyao Hu[1]†, Junru Li[1], Jingwen Chen[1], Weibin Song[2], Xiao-Jing Wang[3], Tianming Yang[1], Stanislas Dehaene[4,5], Shiming Tang[2,6]*, Bin Min[7]*, Liping Wang[1]*

How the brain stores a sequence in memory remains largely unknown. We investigated the neural code underlying sequence working memory using two-photon calcium imaging to record thousands of neurons in the prefrontal cortex of macaque monkeys memorizing and then reproducing a sequence of locations after a delay. We discovered a regular geometrical organization: The high-dimensional neural state space during the delay could be decomposed into a sum of low-dimensional subspaces, each storing the spatial location at a given ordinal rank, which could be generalized to novel sequences and explain monkey behavior. The rank subspaces were distributed across large overlapping neural groups, and the integration of ordinal and spatial information occurred at the collective level rather than within single neurons. Thus, a simple representational geometry underlies sequence working memory.

Episodic experiences in the real or mental world are, by their nature, a succession of events. The ability to remember the ordinal succession of items in a sequence is crucial for various higher-level cognitive functions, including language, episodic memory, and spatial navigation (*1*). However, how a sequence is represented and stored in memory remains largely unknown. There could be two ways of encoding sequences. First, there may be a repertoire of representations for every sequence encountered—in other words, a separate representation for each sequence. Alternatively, the representation could be factorized, for instance with distinct memory slots for items at different ordinal ranks or by separating the temporal structure from the content items (*2*, *3*).

Such factorized representation is also referred to as a form of disentangling (*4*). The hypothesis posits that our brain benefits from representing the underlying structure of the world in a disentangled manner because changing the properties in one part of the structure would leave the representation of other parts intact. Thus, disentangling temporal structures from particular events may lead to faster generalization and novel inferences (*3*, *5*, *6*). However, whether and how the neural representations encode abstract temporal structures in sequence working memory (SWM) remains unclear.

At the single-neuron level, it is often proposed that our brain binds information from multiple domains through multiplicative gain modulation (*7*, *8*). One popular hypothesis for SWM is that abstract information about ordinal number could be conjoined with item-specific sensory information through a gain-field mechanism, such that individual prefrontal neurons would be tuned to the product of those two variables (*2*). Alternatively, the neural codes for sequences may be distributed across a large neural population and bound using matrix or tensor products (*9*). Recent studies have suggested that abstract information may be represented in high-dimensional neural state space (*10*, *11*). The trajectories within subregions (neural manifolds) of this space can instantiate the hidden organizing structures that underlie, for example, motor movements in motor areas (*12*) or time and abstract knowledge in the hippocampus and prefrontal cortex (*13*–*16*).

To investigate the neural representations of SWM at both the single-neuron and population levels, we asked the following questions: (i) whether low-dimensional manifolds underlie the disentangled representation of temporal structure in SWM, (ii) how neurons integrate neural representations of temporal order and sensory items in SWM, (iii) how single neurons are organized anatomically and functionally to contribute to these manifolds, and (iv) whether we can provide a unified mathematical description of those computations at the single-neuron and population levels. To address these questions, we trained two monkeys to perform a visuospatial delayed sequence-reproduction task and used two-photon calcium imaging to record neurons in the lateral prefrontal cortex (LPFC).

Paradigm and behavior

Two macaque monkeys were trained on a delayed spatial sequence-reproduction task (*17*). On each trial, during the sample period, sequences of two or three spatial locations were visually presented while the monkey fixated on a dot at the center of the screen. Each sequence item was drawn without replacement from one of six spatial locations on a ring. Monkeys had to memorize the sequence over a delay of 2.5 to 4 s and then reproduce it by making sequential saccades to the appropriate locations on screen (Fig. 1A).

Overall, the two monkeys performed the task well: At each rank, the mean percent correct rate was significantly higher than chance (Fig. 1B; all P values <<0.001, two-tailed t test) without any significant spatial bias (see fig. S1 for detailed task performance). Recall accuracy decreased with sequence length (Fig. 1B). Both monkeys showed an advantage for items presented at the start of the sequence (the primacy effect). No recency effect was observed. When an item was recalled at an incorrect serial position, its recall spatial location was likely to lie near the original location (Fig. 1C, left), and its recall order was likely to have been swapped with the neighboring orders. Such transposition errors increased with increasing order (Fig. 1C, right).

Hypothesis: Disentangled representation of SWM

The factorized model posits that the brain finds the natural decomposition of sequences comprising two generative factors: ordinal information (ranks 1 to 3) and spatial location (six items). Thus, we tested whether the vector space representing the sequence in memory would be a concatenation of multiple independent rank representations, each embedding a representation of the corresponding spatial item. For instance, to represent the sequence [5 2 4], item 5 is bound to rank 1, item 2 is bound to rank 2, and so on.

To measure the neural state, we injected GCaMP6s virus into the LPFCs of the two monkeys to enable two-photon calcium imaging of the LPFC (Fig. 1, D to F, and fig. S2) [monkey 1, 3609 neurons from 20 fields of view (FOVs); monkey 2, 1716 neurons from 13 FOVs]. We focused on neural activity during the late delay period (1 s before the "fixation-off" go signal) while the monkeys maintained length-2 or -3 spatial sequences in memory.

Neurons exhibiting a conjunctive preference for rank and location were immediately apparent (Fig. 1G; see the proportion of conjunctive neurons in different FOVs in fig. S2). Such neurons responded selectively to

[1]Institute of Neuroscience, Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China. [2]Peking University School of Life Sciences and Peking-Tsinghua Center for Life Sciences, Beijing 100871, China. [3]Center for Neural Science, New York University, New York, NY 10003, USA. [4]Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay, NeuroSpin Center, 91191 Gif/Yvette, France. [5]Collège de France, Universite Paris Sciences Lettres, 75005 Paris, France. [6]IDG/McGovern Institute for Brain Research at Peking University, Beijing 100871, China. [7]Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai 200031, China.
*Corresponding author. Email: liping.wang@ion.ac.cn (L.W.); bin.min@bsbii.cn (B.M.); tangshm@pku.edu.cn (S.T.)
†These authors contributed equally to this work.

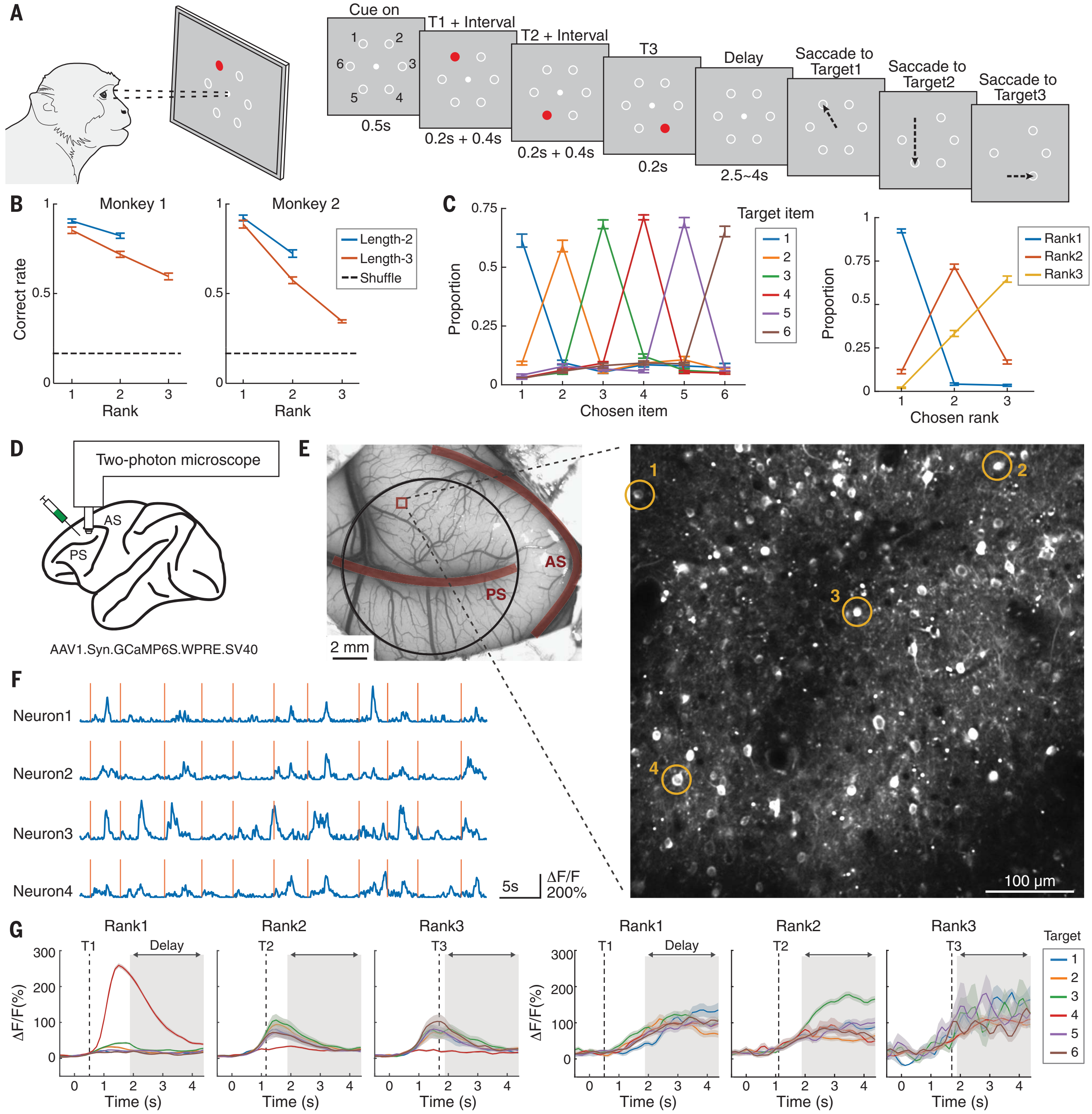

Fig. 1. Two-photon calcium imaging of macaque LPFCs during a delayed sequence-reproduction task. (**A**) Task structure (*17*). T1, target 1. (**B**) Behavioral performance averaged across imaging sessions. Performance is shown as a function of ordinal rank in length-2 sequences (blue) and length-3 sequences (red). Error bars represent SEMs. (**C**) Spatial location (left) and ordinal rank (right) error patterns, averaged across the two monkeys. Location error pattern is shown as a function of spatial location, averaged across different ranks. Rank error pattern is shown as a function of ordinal rank, averaged across different spatial locations. Error bars represent SEMs. (**D**) Illustration of two-photon calcium imaging of monkey LPFCs (*17*). AS, arcuate sulcus; PS, principal sulcus. (**E**) An example FOV, indicated by the red square (left), and its enlargement (right). (**F**) Normalized calcium traces of four example neurons [yellow circles in (E)]. ΔF/F, normalized fluorescent intensity. (**G**) Two example neurons exhibiting the property of conjunctive coding for spatial location and ordinal rank. Traces were aligned to cue onset and pooled across trials according to spatial location and ordinal rank.

particular spatial locations on the ring but also to their ordinal rank in the sequence (first, second, or third). We quantified the influence of spatial item and ordinal rank on the neural responses of single neurons during the late delay period using linear regression, incorporating spatial item and ordinal rank as variables (6 items × 3 ranks = 18 combinations) to fit the calcium signals of individual neurons. We used the regression coefficients to measure each neuron's selectivity to either item or rank variable (*17*).

Disentangled representation of SWM by the LPFC neural population

To examine whether the high-dimensional state of LPFC neuron activity reflected a disentangled representation of SWM, we first obtained vector representations in population states for the 18 location-rank combinations by concatenating the regression coefficients of all neurons from correct trials of monkey 1 (see fig. S3 for monkey 2). We then divided these 18 vectors into three groups along rank and, for each rank, performed a principal components analysis to obtain the axes that captured the major response variance resulting from item changes (fig. S3).

The analysis yielded a highly reliable state-space portrait that captured both the relationship among different rank subspaces and the geometry of spatial representations within each rank subspace (*17*). First, for length-3 sequences, we found three two-dimensional (2D) subspaces, one for each rank (Fig. 2A). Those subspaces were oriented in a near-orthogonal manner in neural state space, as evident by the large principal angles between them (Fig. 2B). To further quantify the degree of alignment across different rank subspaces, for any two ranks—e.g., rank 1 and rank 2—we calculated the variance accounted for (VAF) ratio by projecting the data from the rank-1 subspace to the rank-2 subspace and computing the remaining data variance after the projection. If the two rank subspaces are near orthogonal, the projection from one subspace will capture little of the data variance of the other subspace, which results in a low VAF ratio. The result showed low VAF ratios for all cross-subspace pairs. As a control, if the ranks were shuffled while holding item location constant, the orthogonality of the subspaces was lost (fig. S4). Furthermore, neurophysiology reflected behavior: VAF ratios for within-subspace trial pairs, measuring the stability of rank subspace estimation, were high on correct trials (Fig. 2C) but were low for misremembered locations, where rank-2 and -3 subspaces became difficult to estimate (fig. S5).

Next, we explored the neural encoding of location within each rank subspace. At each rank, we found a common geometric ring structure, reminiscent of the ring shape of the spatial items presented to the monkeys (Fig. 2A). This ring structure was not observed during the baseline period when the visual stimuli were not yet present (Fig. 2A; dots located around [0, 0]). The size of the ring in each subspace, reflecting the encoding strength of location information, decreased with ordinal rank. The ring size became smaller, and ring structure was nearly undetectable on error trials (fig. S5).

We then quantified how well a simple mathematical model with three rank subspaces, each relying on the same 2D spatial code with a distinct modulation factor, could approximate the full 18-variable regression model at the collective level (Fig. 2D and eqs. S1 and S2). There was a high similarity between the data and the model at each rank (score 0.95 for rank 1, 0.99 for rank 2, and 0.98 for rank 3) (*17*), which supports the hypothesis of a factorized representation of spatial items at the collective level (Fig. 2E) with an additive combination of vectorial representations of location at each ordinal rank. A similar geometry was observed in the second monkey for length-2 sequences and, to a lesser extent, for length-3 sequences (fig. S3).

Disentangled rank subspaces at the single-trial level in individual FOVs

The above state-space analysis pools neurons recorded from different FOVs and averages their responses over trials. We investigated

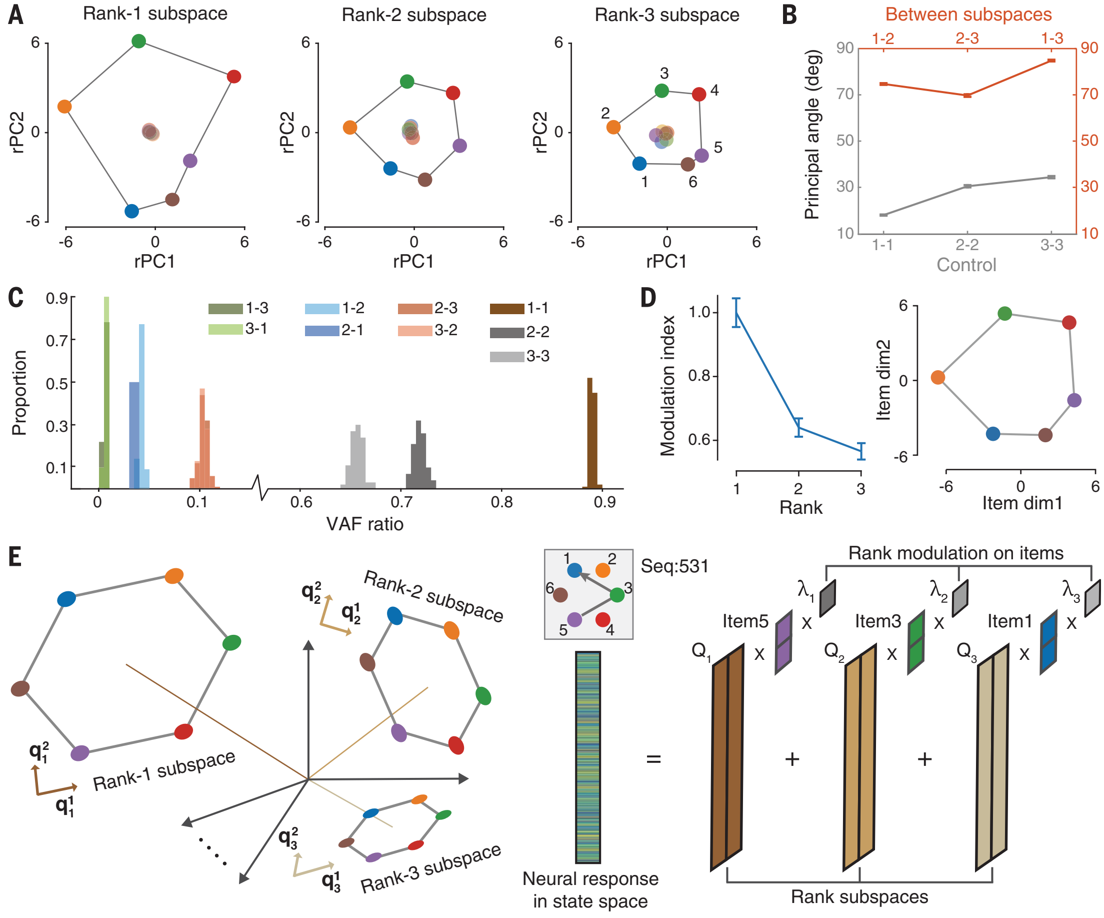

Fig. 2. Disentangled neural state space representation of SWM. (**A**) The population response for a given rank-location combination projected to the corresponding rank subspace for monkey 1. Responses were obtained through linear regression of averaged late delay activity (1 s before fixation off). Locations are color coded. The center points were data at the beginning of the sample period. rPC, rotated principal component. (**B**) The principal angle between different rank subspaces (red). As a control, we randomly split trials in half to obtain two separate estimations of each rank subspace and computed their principal angle (gray). deg, degree. (**C**) The VAF ratio with respect to different rank subspace pairs. As a control, we randomly split trials in half to obtain two separate estimations of each rank subspace and computed their VAF ratio. (**D**) Gain modulation approximation of the projected value at difference rank subspaces. These collective variables can be well approximated by a gain modulation model parameterized by a shared spatial layout (right) and a rank modulation vector (left). dim, dimension. (**E**) A graphical summary of SWM representation in neural state space. Three 2D rank subspaces are oriented in a nearly orthogonal manner in neural state space (left). The neural representation of a sequence can be decomposed into a sum of component items in rank subspaces (right). q and Q, axes of subspace; λ, rank modulation index.

whether the disentangled representation of sequence memory could be validated at the single-trial level. We tested whether the rank subspaces were abstract enough to generalize to different datasets, including untrained, different-length, or error sequences.

Single-trial decoding methods were used to decode item locations at ranks 1, 2, and 3, respectively (Fig. 3A) (*17*). Neurons in almost all (30 of 33) FOVs contained item information at rank 1, and neurons in 21 of 33 (64%) FOVs contained item information at other ranks (rank 2 or rank 3) (tables S1 and S2). Figure 3B shows the decoding results for the six items at rank 1, rank 2, and rank 3 from an example FOV located in the dorsolateral prefrontal cortex (monkey 1, FOV4; fig. S2A). At each rank, the corresponding item could be decoded at above-chance levels during the sample, delay, and reproduction periods. During the delay period, the code for the item was stable, with the decoder performing well even when the training and testing times differed. However, the code during the delay period did not generalize to the sample and reproduction periods, which indicates dynamic changes in the neural code. Similar decoding profiles were found in other FOVs in both monkeys (fig. S6, A and B). By examining decoder error patterns during the late delay period (Fig. 3C), we found that most errors were confusions with the neighboring spatial items.

We next visualized the dynamics of the neural code for location by projecting the population activities at each time bin of a trial to the three decoder-based rank subspaces, which were obtained using neural responses during the late delay period (*17*). The six locations were well separated in the rank subspaces, and, crucially, the ring structure was preserved for all ranks (Fig. 3D). We also investigated the relationships between the three rank subspaces by examining the cross-rank decoding performance and calculating their cross-subspace VAF ratios. The results confirmed the findings from the state-space analysis and showed the minimal cross-rank decoding performance (fig. S6C) and little overlap between the three rank subspaces (fig. S6D), which supports the disentangled representation of sequence memory at the single-trial level.

If sequences are disentangled into a rank-location encoding, the neural subspaces of ordinal rank should generalize to other untrained sequences. We tested this idea using three generalization analyses. First, we used leave-one-sequence-out cross-validation to confirm that the rank subspaces revealed in Fig. 3B remained stable for left-out sequences that were not used during decoder training. The neural subspaces of ordinal ranks (ranks 1 and 2) correctly and stably separated the six spatial items in the left-out sequences during the delay period (Fig. 3E). Second, we tested whether the rank subspaces transferred to sequences of a different length. The rank-1 and -2 subspaces trained on length-2 sequences successfully generalized to length-3 sequences (Fig. 3F) and vice versa (Fig. 3G). Finally, according to the definition of disentangled representation, rank subspaces are independent and could therefore independently fail. We thus tested whether the decoders, trained on correct trials, generalized to error trials that had a correct response at a given rank. For example, when the response to the sequence [1 3 6] is [2 3 5], the code for rank 2 could be expected to transfer between the correct and error trials, despite the errors at rank 1 and rank 3. Figure 3H shows such successful generalization. However, because of the heterogeneous nature of the LPFC, not all FOVs passed these generalization tests (see fig. S7 and tables S3 to S5 for all FOVs).

The geometry of SWM explains sequence behavior

Although SWM relies on disentangled representations, the rank subspaces are not perfectly orthogonal. We therefore asked whether the detailed characteristics of these representations could explain classic sequence-reproduction behaviors, such as the primacy and length effects and the transposition gradient shown in Fig. 1 and fig. S1. We first looked at the relationship between ordinal ranks. The VAF ratios between ranks demonstrate a graded and compressive code (Fig. 2C) (*18*). First, the neural overlap between ranks increased with rank: The VAF ratio between rank 2 and rank 3 was larger than that between rank 1 and rank 2. Second, the overlap was larger for neighboring ranks: VAF ratios between neighboring ranks (rank 1 versus rank 2 and rank 2 versus rank 3) were larger than VAF ratios between distant ranks (rank 1 versus rank 3).

We propose that such compressive coding in the rank dimension is one of the hallmarks of sequence representation in working memory and can explain the monkeys' behavior during sequence recall. First, the larger overlap between adjacent rank subspaces promotes the confusion of locations at consecutive ordinal ranks, leading to the ordinal transposition gradient (Fig. 1C, right) whereby most recall order errors are swaps with the neighboring ranks. Furthermore, the increasing number of transposition errors with rank could potentially arise from the smaller overlap of orders at the beginning of the sequence, resulting in high precision of item information at this stage. Finally, the ring structure in each rank subspace may also explain the frequent confusion of nearby locations (Fig. 1C, left).

Distributed single-neuron basis of rank subspaces

What is the implementation of rank subspaces at the level of single neurons? Does a single neuron contribute to multiple rank subspaces, and, if so, does it exhibit the same preferred locations across different ranks? For each neuron, we projected the unit vector along its axis onto the different rank subspaces (Fig. 4A). The geometric relationship between a single neuron axis and rank-r subspace was characterized by A_r and φ_r, where A_r measures the degree of alignment between single neuron axis and rank-r subspace and φ_r specifies the spatial item preference of a single neuron in rank-r subspace. We could then ask what proportion of neurons contribute to each subspace, whether single neurons align with multiple subspaces, and, if so, whether they have the same preferred location φ_r at different ranks.

The normalized participation ratio (PR) evaluates the fraction of neurons contributing to each subspace (*17*). A value close to 1 indicates that the corresponding rank subspace is distributed across the entire recorded population, whereas a value close to 0 indicates that it is localized to just a few neurons. Around 38% of neurons contributed to rank 1 (34% for rank 2 and 32% for rank 3; Fig. 4B), which suggests that rank memory is broadly distributed in the LPFC population. The three rank subspaces recruited both overlapping and disjoint neurons (fig. S8A).

Next, for neurons contributing to at least two rank subspaces (see materials and methods for neuron selection criteria), we asked whether their preferred spatial location was the same at different ranks. The difference in preferred location φ_r was broadly distributed for all rank pairs and substantially removed from a distribution concentrated around 0 (Fig. 4C and fig. S8B). Thus, the angle φ_r varied with rank for many neurons. Figure 4D shows two example neurons, one exhibiting identical spatial tuning but different amplitudes across the three ranks (classical gain modulation; Fig. 4D, left) and the other showing a shift of spatial tuning across the three ranks (tuning to item 6 at rank 1, items 4 to 5 at rank 2, and items 3 to 4 at rank 3; Fig. 4D, right). The angle φ_r provided a good summary of the neuron's spatial preference at each rank because the angular difference between ranks predicted the difference in spatial location preference (Fig. 4E; see the angle estimation in fig. S9 and the tuning curves for the 35 neurons that contribute most to each rank subspace in fig. S10). Similar findings were obtained from monkey 2 (figs. S8, S9, and S11).

These results reject a simple model where gain modulation occurs at the level of single neurons, with each neuron having a fixed spatial tuning curve modulated by a different scalar at each rank (2). Rather, gain modulation is a collective phenomenon that occurs at the neural population level and is best described by matrix rather than scalar multiplication (Fig. 2E and eq. S2). As a result, the memorized content at each ordinal rank is sent in a different direction of neural hyperspace, and the underlying single-neuron tuning curves, characterized by φ_r, may deviate greatly from a simple gain modulation profile and exhibit a tuning shift with rank.

Anatomical organization of the compositional code in the LPFC

Two-photon imaging provided us with the opportunity to examine the spatial anatomical organization of the neural codes and follow it longitudinally across days. We calculated a spatial clustering index for neurons contributing to each rank subspace (Fig. 5A), as assessed by their alignment A_r, and compared it with shuffled distributions obtained by randomly permuting the positions of all neurons (17). The rank code showed no significant anatomical clustering at any spatial scale ranging from ~10 to ~500 μm (Fig. 5B). We examined whether neurons with similar

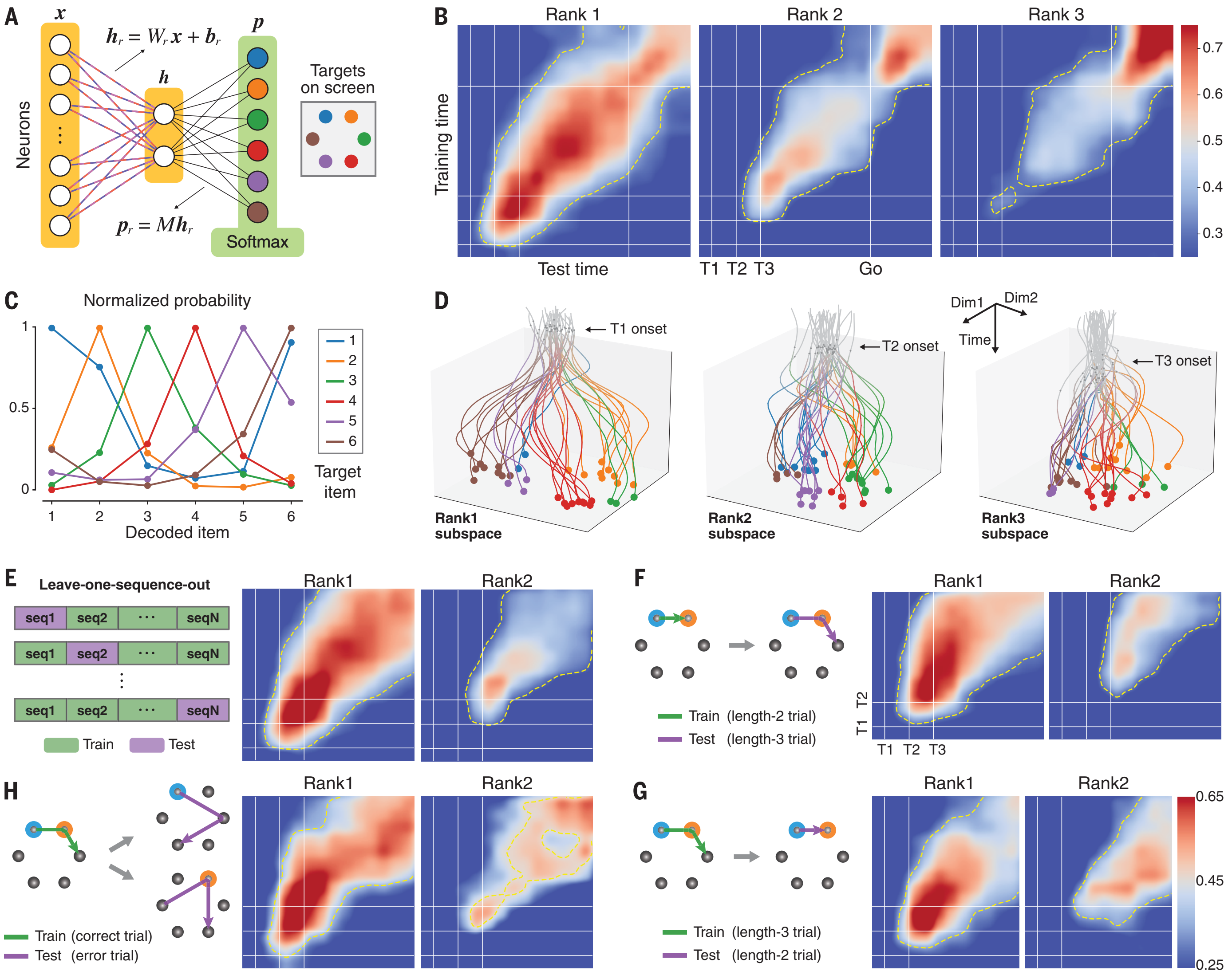

Fig. 3. Single-trial decoding analysis and compositional generalization test of rank subspace. (**A**) The architecture of the decoder (17). Neural activity (x) was linearly projected (W, weight; b, bias) into a 2D hidden state (h), which was classified against target matrix (M) to obtain softmaxed scores (p) for all items. Rank-specific variables are indicated by a subscripted r. (**B**) Cross-temporal decoding accuracy for spatial locations of each rank in length-3 sequences in an example FOV from monkey 1. Only correct trials were used, and a leave-one-trial-out cross-validation protocol was used. The yellow contours enclose areas of strong decoding performance ($P < 0.001$, extreme pixel-based permutation test). T1, T2, and T3 indicate the onset of the first, second, and third targets, respectively. Go indicates fixation point off. Colormap: 0.25 to 0.75. (**C**) The normalized distribution of decoded locations averaged in the last three time windows of the delay period and across ranks. (**D**) Sequence trajectories in three decoding-based rank subspaces evolving across time until the end of delay. Each trajectory was obtained by averaging trials from the same sequence and was colored according to the location of the corresponding rank. (**E**) Cross-temporal decoding accuracy obtained by leave-one-sequence-out protocol. The correct trials of length-3 sequences were split into test and training sets. The test set contained all the trials for one particular sequence, whereas the training set consisted of the remaining trials for other sequences. Contours, $P < 0.005$ (extreme pixel-based test). Colormap: 0.25 to 0.65 [same for (E) to (H)]. (**F**) Cross-length decoding. Decoders trained with trials of the length-3 sequence were tested in trials of length-2 sequence. All the data used for training and testing were correct trials. (**G**) Cross-length decoding from length-2 sequences to length-3 sequences [similar to (F)]. (**H**) Decoding location match in error trials. Decoder trained with all the length-3 correct trials was tested in error trials with the correct response at the rank where the decoder was trained.

location tuning were anatomically located closer to each other (Fig. 5C). The location code displayed significant clusters at a scale of <150 μm for ranks 1 and 2 (Fig. 5D). A similar anatomical pattern was obtained from other FOVs in both monkeys (fig. S12).

We also examined whether the code in the population of neurons was stable across different recording days. For the same recording FOV, we trained the decoder using data from one recording day and tested it on data from a different day (Fig. 5E). The disentangled rank subspaces generalized well across days (Fig. 5F), indicating the long-term stability of the code embedded in the monkey's LPFC.

Discussion

Using two-photon calcium imaging in the LPFC of macaque monkeys performing a visuospatial sequence-reproduction task, we revealed the representational geometry of SWM in the LPFC neural state space. Sequence memory relied on a compositional neural code with separate disentangled low-dimensional rank subspaces for every rank, each of which was broadly distributed across the neural population. Rank and item variables were integrated through multiplicative gain modulation at the collective level, but not within single neurons. Furthermore, the rank subspaces were abstract and generalizable to novel and variable-length sequences.

Disentangled rank representation and gain modulation

How does the brain efficiently learn complex cognitive tasks such as delayed sequence reproduction? One important strategy is to split a complex whole into simpler parts that reflect the underlying structure of a task. In the present study, we explicitly searched for a neural representation with axes that aligned with the generative factors of the model—i.e., the ordinal ranks. We found that the LPFC neural population implements a decomposition into three subspaces that reflect the underlying structure of sequence memory—i.e., three spatial rings, one for each rank (Fig. 2E). The simple 3-by-2 geometrical structure that we observed reflected the 2D spatial content memorized at each rank. Although we showed generalization to left-out trials and sequences (Fig. 3), future research should examine generalization to untrained sequences and new item types (e.g., letters and numbers). If the LPFC neural population geometry is a ubiquitous feature of brain activity that extends beyond the spatial domain, we predict that orthogonal subspaces, one for each ordinal rank, should continue to be observed and may contribute to learning and inference in any task that relies on the temporal structure of ordinal knowledge.

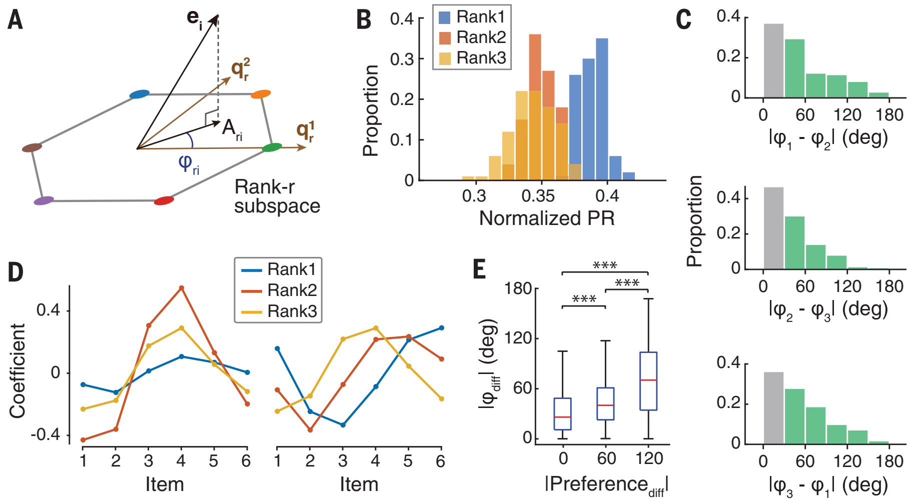

Fig. 4. Single-neuron basis of rank subspaces. (**A**) Illustration of how each single-neuron normalized response vector was projected onto the rank subspaces. (**B**) Quantification of the degree of localization in neural state space for different rank subspaces for monkey 1. The histogram shows the empirical distribution of normalized PR estimated by bootstrap. (**C**) Histograms of $φ_r$ difference for different rank pairs. A difference of <30° (gray bar) suggests the same preferred location at two ranks, whereas the larger differences (green bars) indicate different location preferences across ranks. (**D**) Two example neurons with tuning curves showing classical gain modulation (left) and preference shift (right), respectively. (**E**) The correspondence of $φ_r$ difference (based on $φ_r$ extracted from the geometric relationship between single-neuron axis and rank subspaces) and preference difference (based on preference of spatial location extracted from the raw regression coefficients).

Diverse cognitive functions, including coordinate transformation, multimodal integration, place anchoring, abstraction, and attention, are performed through a canonical neural computation of gain modulation (*7*, *8*, *19*–*21*). Accordingly, a previous model had proposed that sequences are encoded through the binding of ordinal and identity information in individual prefrontal neurons tuned to the product of these two variables (*2*). Our results are close to this gain modulation model but depart from it in a crucial way: In contrast to the predictions of a single-neuron gain modulation, our data suggest that neural gain modulation occurs only at the collective level and is therefore best described by matrix rather than scalar multiplication. This aspect of our results is compatible with models that show how recurrent neural networks can learn vectorial representation of sequences (or even sentences) by implicitly compiling them into a sum of filler-role bindings using tensor products (*9*). The present data suggest that LPFC neural states implement vector symbolic architectures and tensor-product representations for sequence memory—an idea with a vast number of applications to artificial neural networks.

Neural mechanisms of classical behavioral effects in serial recall from working memory

Serial recall from working memory is characterized by empirical findings from both behavioral and theoretical perspectives (*22*, *23*), including not only the quality and quantity of item information maintained but also item-order binding information (e.g., binding errors). Yet, there has previously been no neural evidence providing a mechanistic explanation for most of those behavioral observations. The neural code for SWM that we observed shows that, although the rank subspaces are nearly orthogonal, the sequence representation implements a graded and compressive encoding of rank information, with subspaces showing increasing overlap as rank increases. This neural response profile is consistent with previous findings from single-unit recordings in macaque prefrontal and parietal areas, in which ordinal numbers were represented with the characteristic signature of Weber's law (*18*). The code property we describe also provides insight into several influential models of sequence memory, such as the slot, resource, and interference model, and can thus explain many behavioral benchmarks of working memory for serial order, including the effects of list length and composition, the primacy effect, the temporal transposition errors, and potentially also working memory capacity (*24*).

Converting temporally segregated sensory inputs into spatially overlapping sustained brain activity patterns

Previous mechanistic models of sequence memory have mostly focused on a temporal encoding of sequences supported either by synfire chains, neural oscillations, or rhythmic fluctuations in neuronal excitability (*5*, *25*–*27*). All of these models posit that the order of items in a sequence is represented by the timing of neural activity, either by locking to the relative phase of a lower-frequency oscillation (theta, 3 to 8 Hz) (*28*) or by the replay of a neural trajectory (*29*, *30*). However, it is also widely accepted that attractor states of sustained neural activity play a central role in working memory (*31*, *32*). Neurophysiological studies in nonhuman primates have found that this mechanism could apply to the memory for sequences because the sustained activity of prefrontal cortex neurons maintained both

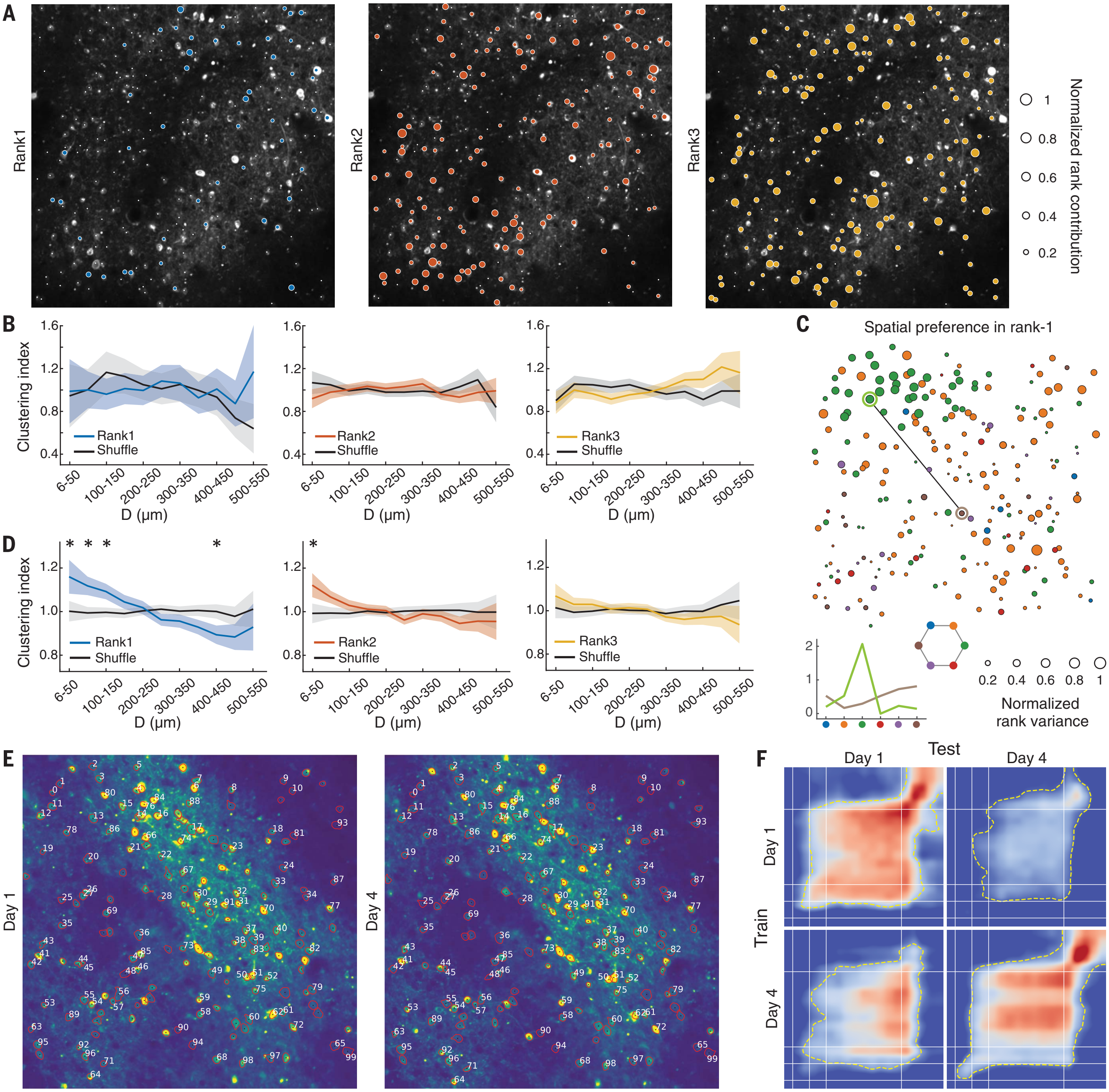

Fig. 5. Anatomical organization of the compositional code in the LPFC. (**A**) The organization of neural alignment with different rank subspaces in an example FOV from monkey 1. Normalized rank contributions (normalized A_r^2 for rank-*r* subspace) were overlaid with average calcium image. (**B**) The neurons with similar rank contributions show no spatial clustering in the example FOV. Neurons from the FOV in (A) with substantial rank contribution were collected to calculate the clustering index. Shaded areas represent 95% confidence intervals. (**C**) Pairwise analysis for one example neuron pair in rank-1 spatial preference map. Neurons were marked by colored circles, with color and size indicating spatial preference and normalized rank variance, respectively. (Bottom) Tuning curves of the circled neuron pair. (**D**) Functional clustering for the FOV in (C). The clustering index was based on the average Pearson correlation coefficient across neuron pairs within a particular cortical distance range (*17*). (**E**) Comparison of FOVs recorded in the same location across days. 171 regions of interest (ROIs) were identified on both days. The shared indexes of the first 99 ROIs were marked. (**F**) Cross-day decoding for rank-1 locations in correct, length-3 sequence. In off-diagonal panels, the decoder trained in one day was tested in another day by only using the overlapping neurons. In diagonal panels, decoders were trained with all neurons and tested with the leave-one-trial-out method. Colormap and contour are the same as in Fig. 3G.

item and order information during a working memory delay (*33*–*35*). Our data support such a sustained activity mechanism because the geometry of disentangled representations was stable during the delay period (Fig. 3B). Our results suggest that the brain transforms time into space by converting temporally segregated sensory inputs into spatially overlapping sustained brain activity patterns. They do not, however, exclude the simultaneous presence, at a finer time scale, of a temporal code involving phase synchrony or replay.

Seven decades ago, Karl Lashley (*36*) postulated that serial order is processed by creating and manipulating a spatial pattern of neural activity. He speculated that to control sequential actions, our brain needed to transform temporally segregated sensory experiences into a sustained spatial pattern of brain activity. In agreement with this early intuition, the simple geometrical organization of SWM that we uncovered may provide a fundamental neural mechanism to bridge our understanding of neural circuits and their computational functions.

REFERENCES AND NOTES

1. S. Dehaene, F. Meyniel, C. Wacongne, L. Wang, C. Pallier, *Neuron* **88**, 2–19 (2015).
2. M. Botvinick, T. Watanabe, *J. Neurosci.* **27**, 8636–8642 (2007).
3. Y. Liu, R. J. Dolan, Z. Kurth-Nelson, T. E. J. Behrens, *Cell* **178**, 640–652.e14 (2019).
4. Y. Bengio, A. Courville, P. Vincent, *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1789–1828 (2013).
5. G. Buzsáki, A. Draguhn, *Science* **304**, 1926–1929 (2004).
6. C. Baldassano *et al*., *Neuron* **95**, 709–721.e5 (2017).
7. R. A. Andersen, G. K. Essick, R. M. Siegel, *Science* **230**, 456–458 (1985).
8. A. Pouget, T. J. Sejnowski, *J. Cogn. Neurosci.* **9**, 222–237 (1997).
9. P. Smolensky, *Artif. Intell.* **46**, 159–216 (1990).
10. V. Mante, D. Sussillo, K. V. Shenoy, W. T. Newsome, *Nature* **503**, 78–84 (2013).
11. M. Rigotti *et al*., *Nature* **497**, 585–590 (2013).
12. M. M. Churchland *et al*., *Nature* **487**, 51–56 (2012).
13. S. Bernardi *et al*., *Cell* **183**, 954–967.e21 (2020).
14. E. H. Nieh *et al*., *Nature* **595**, 80–84 (2021).
15. M. F. Panichello, T. J. Buschman, *Nature* **592**, 601–605 (2021).
16. J. Wang, D. Narain, E. A. Hosseini, M. Jazayeri, *Nat. Neurosci.* **21**, 102–110 (2018).
17. Materials and methods are available as supplementary materials online.
18. A. Nieder, I. Diester, O. Tudusciuc, *Science* **313**, 1431–1435 (2006).
19. J. R. Duhamel, F. Bremmer, S. Ben Hamed, W. Graf, *Nature* **389**, 845–848 (1997).
20. R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, *Nature* **435**, 1102–1107 (2005).
21. J. H. Reynolds, D. J. Heeger, *Neuron* **61**, 168–185 (2009).
22. X. Jiang *et al*., *Curr. Biol.* **28**, 1851–1859.e4 (2018).
23. K. Oberauer *et al*., *Psychol. Bull.* **144**, 885–958 (2018).
24. M. J. Hurlstone, G. J. Hitch, A. D. Baddeley, *Psychol. Bull.* **140**, 339–373 (2014).
25. J. E. Lisman, O. Jensen, *Neuron* **77**, 1002–1016 (2013).
26. X. J. Wang, *Physiol. Rev.* **90**, 1195–1268 (2010).
27. M. Abeles, G. Hayon, D. Lehmann, *J. Comput. Neurosci.* **17**, 179–201 (2004).
28. M. Siegel, M. R. Warden, E. K. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21341–21346 (2009).
29. M. A. Wilson, B. L. McNaughton, *Science* **265**, 676–679 (1994).
30. K. Diba, G. Buzsáki, *Nat. Neurosci.* **10**, 1241–1242 (2007).
31. S. Funahashi, C. J. Bruce, P. S. Goldman-Rakic, *J. Neurophysiol.* **61**, 331–349 (1989).
32. A. Compte, N. Brunel, P. S. Goldman-Rakic, X. J. Wang, *Cereb. Cortex* **10**, 910–923 (2000).
33. P. Barone, J. P. Joseph, *Exp. Brain Res.* **78**, 447–464 (1989).
34. S. Funahashi, M. Inoue, K. Kubota, *Behav. Brain Res.* **84**, 203–223 (1997).
35. Y. Ninokura, H. Mushiake, J. Tanji, *J. Neurophysiol.* **89**, 2868–2873 (2003).
36. K. S. Lashley, *The Problem of Serial Order in Behavior* (Bobbs-Merrill, 1951).
37. Y. Xie *et al*., Data repository for 'Geometry of Sequence Working Memory in Macaque Prefrontal Cortex', *Zenodo* (2021); https://doi.org/10.5281/zenodo.5739376.
38. P. Hu, Y. Xie, Code for 'Geometry of Sequence Working Memory in Macaque Prefrontal Cortex,' Zenodo (2021); https://doi.org/10.5281/zenodo.5746184.

ACKNOWLEDGMENTS

We thank W. Fang for data analysis, Y. Li and H. Jiang for animal surgery, and Peking University Laboratory Animal Center for animal care. **Funding:** This study was supported by National Science and Technology Innovation 2030 Major Program 2021ZD0204102 (L.W. and B.M.), Shanghai Municipal Science and Technology Major Project 2021SHZDZX and 2018SHZDZX05 (L.W.), Strategic Priority Research Programs XDB32070201 (L.W.), Natural Science Foundation of China 11901557 (B.M.), and Natural Science Foundation of China 31730109 and U1909205 (S.T.). **Author contributions:** Conceptualization: Y.X., P.H., X.-J.W., T.Y., S.D., S.T., B.M., and L.W. Methodology: Y.X., P.H., J.L., X.-J.W., T.Y., S.D., B.M., and L.W. Investigation: Y.X., W.S., J.C., and S.T. Visualization: Y.X. and P.H. Funding acquisition: S.T. and L.W. Project administration: S.T. and L.W. Supervision: S.T., B.M., and L.W. Writing – original draft: B.M., S.D., and L.W. Writing – review and editing: Y.X., P.H., X.-J.W., T.Y., B.M., S.D., and L.W. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data and code used in this study are available on Zenodo (*37*, *38*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm0204
Materials and Methods
Figs. S1 to S12
Tables S1 to S5
References (*39*–*46*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

24 August 2021; accepted 12 January 2022
10.1126/science.abm0204