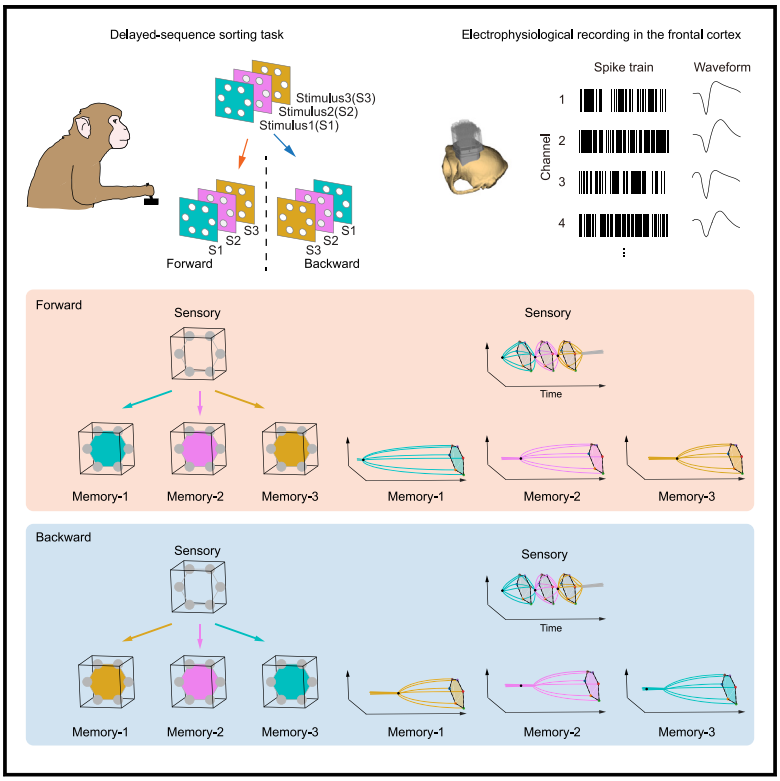
Neuron

Article

Flexible control of sequence working memory in the macaque frontal cortex

Graphical abstract



Highlights

- Separate low-dimensional sensory and memory subspaces are found in the frontal cortex
- Spatial items enter a shared sensory subspace and then move into memory subspaces
- Neural dynamics in forward and backward tasks reflect flexible control
- Neural activities in memory subspaces predict monkeys' behavior

Authors

Jingwen Chen, Cong Zhang, Peiyao Hu, Bin Min, Liping Wang

Correspondence

minbin@lglab.ac.cn (B.M.), liping.wang@ion.ac.cn (L.W.)

In brief

Chen et al. discovered the separable and generalizable subspaces for sensory and sequence working memory in macaque frontal neural states. The neural activities and their dynamics within these subspaces reflected the flexible control of sequence working memory in various sequence sorting tasks, predicting monkeys' behavior in single trials.

Chen et al., 2024, Neuron *112*, 1–13
October 23, 2024
https://doi.org/10.1016/j.neuron.2024.07.024

CellPress

Neuron

CellPress

Article

Flexible control of sequence working memory in the macaque frontal cortex

Jingwen Chen,[1,4] Cong Zhang,[1,4] Peiyao Hu,[1,3,4] Bin Min,[2,*] and Liping Wang[1,5,*]
[1]Institute of Neuroscience, Key Laboratory of Brain Cognition and Brain-Inspired Intelligence Technology, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China
[2]Lingang Laboratory, Shanghai 200031, China
[3]Present address: Department of Psychology, New York University, New York, NY 10003, USA
[4]These authors contributed equally
[5]Lead contact
*Correspondence: minbin@lglab.ac.cn (B.M.), liping.wang@ion.ac.cn (L.W.)
https://doi.org/10.1016/j.neuron.2024.07.024

SUMMARY

To memorize a sequence, one must serially bind each item to its rank order. How the brain controls a given input to bind its associated order in sequence working memory (SWM) remains unexplored. Here, we investigated the neural representations underlying SWM control using electrophysiological recordings in the frontal cortex of macaque monkeys performing forward and backward SWM tasks. Separate and generalizable low-dimensional subspaces for sensory and memory information were found within the same frontal circuitry, and SWM control was reflected in these neural subspaces' organized dynamics. Each item at each rank was sequentially entered into a common sensory subspace and, depending on forward or backward task requirement, flexibly and timely sent into rank-selective SWM subspaces. Neural activity in these SWM subspaces faithfully predicted the recalled item and order information in single error trials. Thus, compositional neural population codes with well-orchestrated dynamics in frontal cortex support the flexible control of SWM.

INTRODUCTION

Imagine you are filling in your date of birth on a form. You sequentially recall your birth year, month, and day one by one and flexibly reorder this sequence based on the requirements of the form, which could be yy/mm/dd or dd/mm/yy. In this process, the brain has to assign a particular item (here, numbers) to a particular role (year, month, or day). However, whether and how the brain performs these assignments in a sequence remains a mystery, and little evidence yet shows whether animals possess such ability.[1–3]

To encode a sequence, the brain needs to solve the control problem by flexibly and sequentially updating the flow of information into working memory (WM), avoiding interference, and robustly maintaining them in sequence working memory (SWM).[4,5] One possible solution is a WM "gate," which assumes useful information is updated into appropriate WM spaces and interfering information is kept out.[4] Thus, hypothetically, through this gate, the input items can selectively bind to correct ordinal positions in SWM while the remembered items are stored in different WM subspaces to avoid interference.

Similar control mechanisms have been previously documented in the research at both single and population neural levels within the motor or premotor cortices, accompanying one or sequential voluntary movements.[6–11] For example, single neurons during the preparation period showed selective activities to internally generated sequential movements.[6,7] At the population level, it has been demonstrated that sequential reaches were prepared and executed in orthogonal subspaces to achieve the correct preparatory state for the next reach while the current reach was still underway.[9] Besides the motor cortex, the frontal cortex is specialized relative to the sensory cortex to control the flow of information, in particular in WM.[4,12] Early connectionist models have demonstrated that the prefrontal cortex could utilize gating to minimize interference during cognitive control,[13,14] and basal ganglia could control the gating of prefrontal representations.[15,16] For example, independent subspaces were found in the prefrontal cortex to separately represent WM content and its motor preparation,[17] and multiple-item information in WM could be transferred using orthogonal subspaces through selective attention.[12] Furthermore, on one hand, it has been proposed that WM is broadly distributed across sensory and prefrontal regions,[18–20] and the flexible maintaining or controlling of an item in WM could rely on the reciprocal connections between the prefrontal areas and the sensory cortex,[21] on the other hand, the dynamic process of the gating system (e.g., context-dependent selection) could also unfold within the prefrontal cortex without interacting with sensory responses.[22]

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

Nevertheless, how the SWM control is represented in the frontal cortex has not been experimentally explored.

Using two-photon calcium imaging of monkeys, our team recently demonstrated that prefrontal neural states in SWM during the late delay period could be decomposed into low-dimensional disentangled rank subspaces, each storing the corresponding item's visuospatial information.[23] Because of the poor temporal resolution of calcium signals, the neural dynamics of SWM control (e.g., gating, binding, and maintenance) have yet to be investigated. In this study, we trained three macaque monkeys to perform novel visuospatial delayed-sequence-sorting tasks. In each trial, after viewing a sequence of locations and memorizing it during a short delay, the animal was prompted to reproduce the sequence by consecutive touches in either the presented order (forward) or the reverse order (backward). We used a 157-channel, high-throughput electrophysiological recording system to record thousands of neurons in the frontal cortex.

RESULTS

Paradigm and behavior

Three macaque monkeys were trained to learn delayed-sequence-sorting tasks (Figure 1A). On each trial, spatial sequences with 1, 2, or 3 items were visually presented during the sample period, while the monkey had to fixate on the dot at the center of the screen. Each sequence item was drawn (without replacement) from one of the six spatial locations of a ring (or hexagon). Monkeys had to memorize the sequences for a 1.15–1.55 s delay with fixation and reproduce the sequences by making sequential touches to the appropriate locations on the screen. In different sessions, monkeys were required to touch the locations in either the presented order (forward) or the reverse order (backward) (Figure 1A; see STAR Methods). As the sequence length was randomized for each trial, monkeys could not anticipate the number of items before a trial. A water reward was given to monkeys after the correct completion of the sequence.

Overall, the three monkeys performed the tasks well: the mean percent correct rate at each rank of length-1, -2, and -3 sequences was significantly higher than chance (Figure 1B, all $p << 0.001$, two-tailed t test). When making errors, monkeys were more likely to choose an item spatially close to the target (Figures 1C and S1A–S1C) and confuse an order with the neighboring order (Figures 1D and S1A–S1C). The behavioral performance (details in Figure S1) was broadly consistent with the previous findings in monkeys.[1,23,24]

Recordings and single-neuron responses

A 157-channel microdrive electrode array[25] was implanted in the frontal areas of the three monkeys (Figures 1E and S1D), mainly covering the prefrontal cortex and the premotor cortex. In total, 6,790 neurons were recorded when monkeys performed the two tasks (monkey O: 3,829 neurons, monkey G: 1,890 neurons, and monkey L: 1,071 neurons; Table S1).

Neural responses exhibited diverse and mixed task-related activities. Two types of neural activities were immediately apparent. We first found neurons showing stimulus-evoked activities during the sample period. This type of neuron transiently encoded spatial items upon each stimulus onset, with identical item selectivity across ordinal ranks (Figure 1F, the neuron tuned to item 3 at all three ranks). We also discovered neurons that showed sustained activity during the delay period. Such neurons exhibited conjunctive coding for rank and item after the appearance of the stimulus and maintained the information throughout the delay (Figures 1G–1I). For example, there were sustained neurons selective to item 1 at rank 1 (Figure 1G), items 1 and 6 at rank 2 (Figure 1H), and item 6 at rank 3 (Figure 1I), respectively. Crucially, the stimulus-evoked and WM-sustained activities were not always present in separate sets of neurons. Instead, these two kinds of activities seemed deeply entwined at the single-neuron level (see the proportion of mixed coding neurons in three monkeys in Figures S1E and S1F). For instance, the example delay-selective neurons in Figures 1H and 1I also showed selectivity during the sample period. Since we only collected neural data with length-2 trials for monkey L, the further analysis with length-3 sequences mainly included data from monkeys O and G.

Dynamics in frontal neural subspaces in the forward task

In line with our prior two-photon calcium imaging findings,[23] we hypothesized that the task variables (i.e., sensory and WM) could be separately represented in low-dimensional subspaces in the frontal cortex's neuronal state space. Specifically, we postulated that each rank WM is represented in a distinct low-dimensional subspace. In addition, before binding with ordinal rank, we expected that spatial items presented during sample periods might be transiently encoded in a shared low-dimensional sensory subspace. In these subspaces, the evolution of the task representations throughout the entire trial could be tracked, allowing us to examine the dynamic control process underlying the SWM.

To evaluate our hypothesis of rank-shared sensory subspace and rank-specific WM subspaces, we extended the static linear decoding approach in our prior SWM study[23] to a method that addressed the intricacies of evolving dynamics. In this method, the weights of each subspace were first estimated using an initial decoding template (e.g., decoding rank-1 item in the delay period for rank-1 WM subspace, or decoding items at all ranks in their corresponding sample period for rank-shared sensory subspace), and then refined using an optimization procedure that features adaptive decoding time windows and decoding targets (see STAR Methods). For instance, for length-2 sequences, the identification of sensory subspace (subspace-1 shown in Figure 2A) involved a joint effort to maximize the decoding accuracy of both rank-1 and rank-2 items within a time window after their onset. By contrast, the identification of rank-1 WM subspace (subspace-2 in Figure 2A) relied on the decoding accuracy of rank-1 items during the delay period. The temporal dynamics of task variables are then unveiled through the rise and fall of information content in the corresponding subspace quantified by the decoding accuracy.

We first analyzed length-3 sequences in the forward task and identified four subspaces: one sensory and three rank-WM subspaces (Figure 2B; see Figure S2A for the explained variance for each subspace and Figure S2B for the contribution of each time point to the identified rank-WM subspaces). In the sensory

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

CellPress

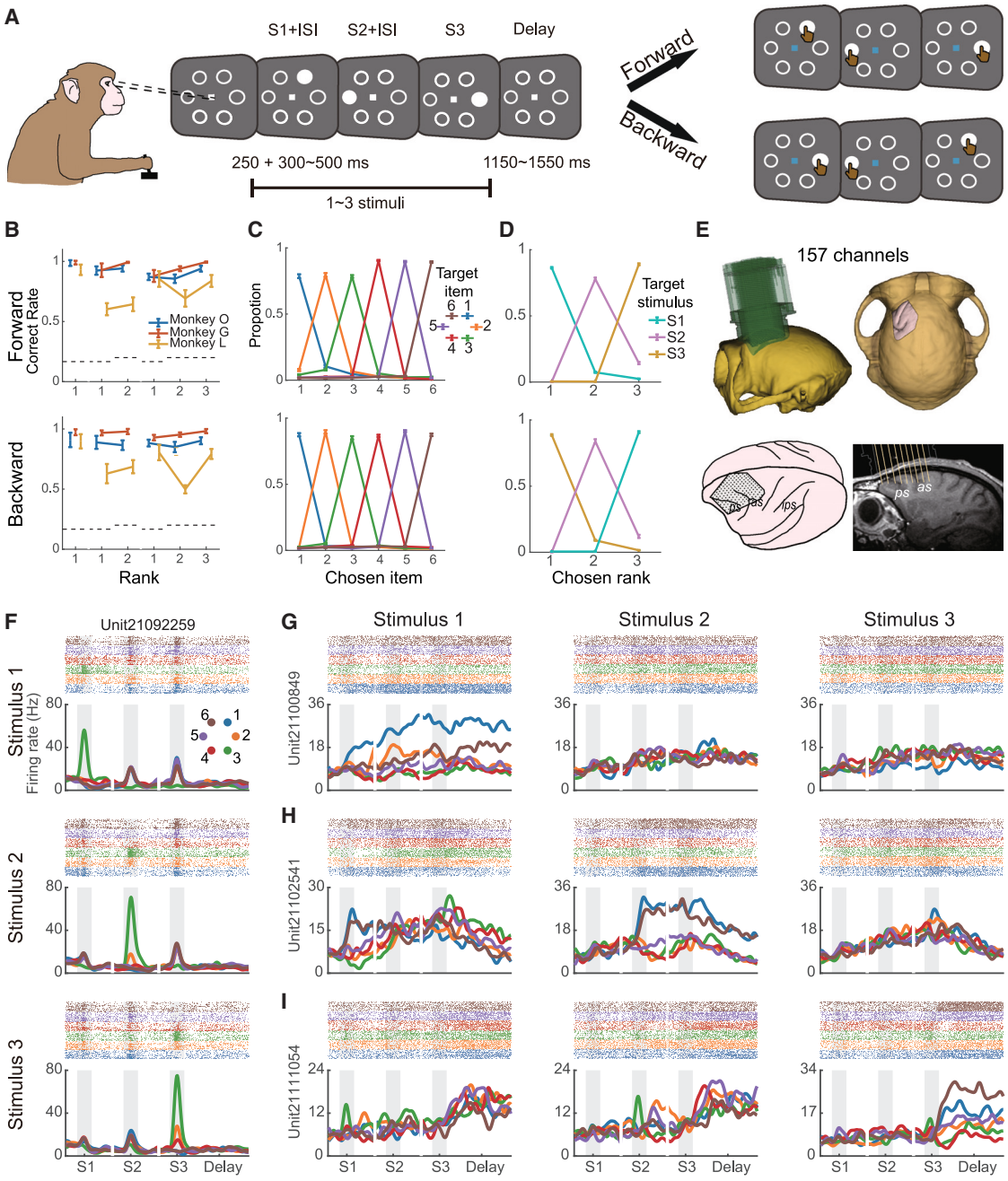


Figure 1. Delayed-sequence-sorting task: design, behavior, and neuronal recordings

(A) Task structure. On each trial, a sequence containing one, two, or three consecutive spatial locations (randomly chosen without replacement) was displayed on the screen. Monkeys were instructed to memorize the sequence and reproduce it either in the presented order (forward) or in the reverse order (backward), in separate sessions. A trial was terminated immediately when monkeys made an error response (see STAR Methods).

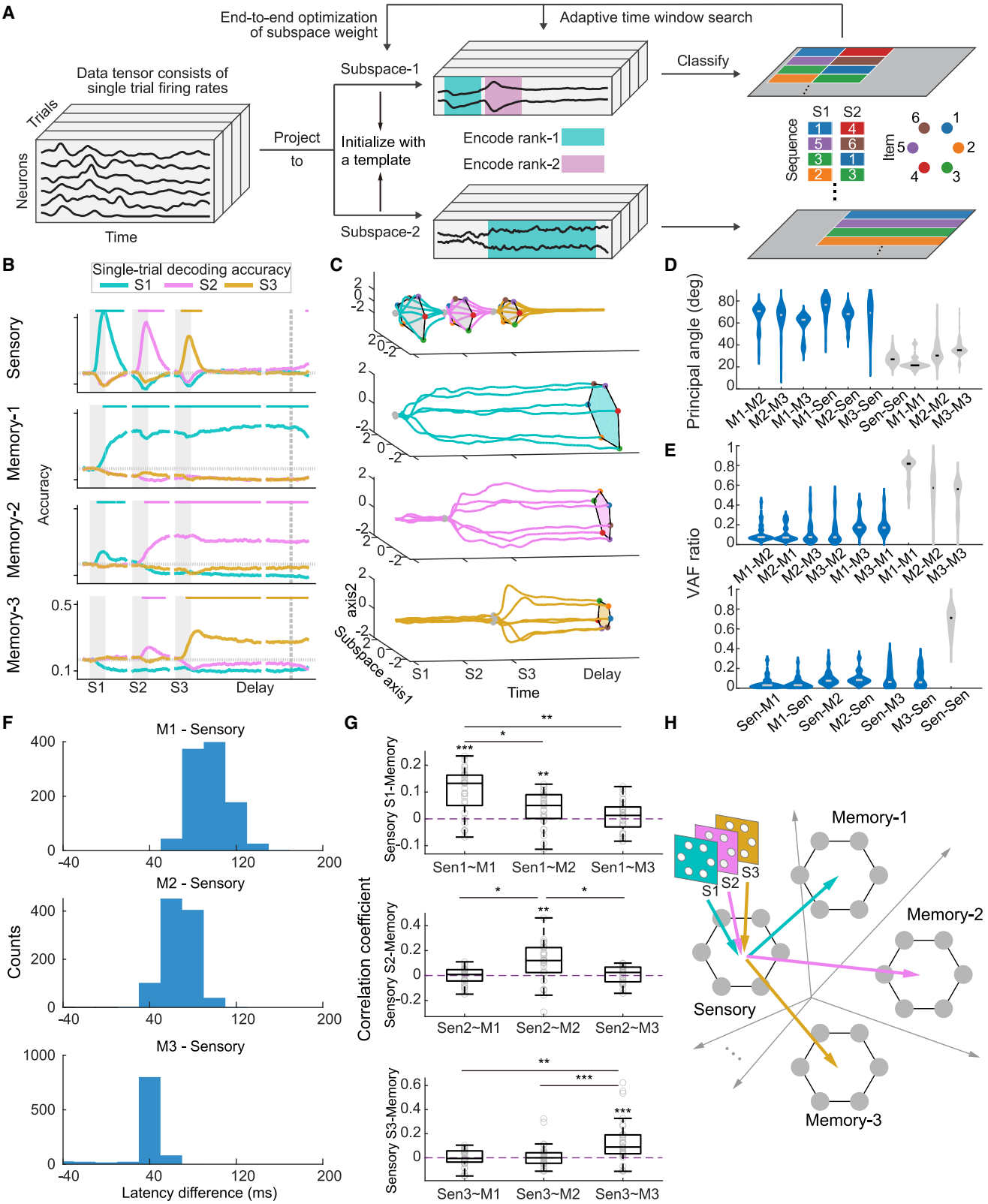
(B–D) Behavioral performance. (B) Conditional correct rate as a function of ordinal rank for length-1, length-2, and length-3 trials (error bar: SD across sessions; dashed lines: chance level); all $p < 0.001$, two-tailed t test. (C and D) Spatial location and ordinal error patterns for only length-3 trials averaged across three monkeys (error bar: SD; see Figure S1 for more details).

(E) Recording sites. A 157-channel microdrive electrode array was implanted over the frontal lobe. Segmented models of the skull and the microdrive (up left), models of the skull after craniotomy and the brain (upright), schematic of electrode locations (bottom left), and MRI image of monkey O (bottom right).

(F–I) Example of neurons with stimulus-related (F) and memory-related activity (G–I). In each panel, raster plots (up) and mean firing rates (bottom) were color coded according to the spatial locations (shown in the first panel of F) at a given ordinal rank (labeled as stimulus 1/2/3). Vertical gray bars: stimulus presentation window. Note that (F) and (G) were recorded in the forward sessions, while (H) and (I) were recorded in the backward sessions.

See also Figure S1 and Table S1.

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024



(legend on next page)

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

subspace, upon each stimulus onset, the decoding accuracy of corresponding stimulus information gradually rose, peaked at the stimulus offset, and then fell back to the chance level after about 300 ms (Figure 2B, first row). To better visualize this dynamic process, we depicted the trajectory of neural activities within this subspace (Figure 2C, first row): the neural representation of spatial item information at each rank formed a well-organized ring structure during the corresponding sample period, and the rise and fall of this ring structure in this sensory subspace underlay the transient dynamics of decoding accuracy shown in Figure 2B.

The remaining three rank-WM subspaces, denoted as memory-1, -2, and -3 (M1, M2, and M3), respectively, showed decoding patterns distinct from the one exhibited by the sensory subspace (Figure 2B). For instance, in the memory-1 subspace, upon onset of the rank-1 stimulus, the decoding accuracy of rank-1 item information gradually rose and remained at a high level throughout the delay. In line with this finding, a persistent rank-1 ring structure was present throughout the whole delay period in this subspace (Figure 2C). Similar decoding patterns and state-space trajectories were found for both memory-2 and -3 subspaces (Figures 2B and 2C). As a control, these analyses were also performed when including more dimensions of the subspaces (e.g., 5 dimensions), with similar outcomes as in the original two-dimension case (Figures S2L–S2O). We also examined the geometric relationship between these four subspaces by computing both the principal angle and the variance accounted for (VAF) ratio[23] (see STAR Methods). For any subspace pair, compared with the controls (different estimations for the same subspace), we found considerably larger principal angles (Figure 2D) and smaller VAF ratios (Figure 2E). Similar results held for the second monkey (Figures S2D–S2F, S2H, and S2I). Thus, the results confirmed that the four subspaces were near orthogonal, and the sensory and WM representations were relatively separated in frontal state space at the population level.

To examine the neural basis of these subspaces at the level of single neurons, we then projected the unit vector along single-neuron axis onto different subspaces. The derived square scalar projections were then used to quantify the contribution of single neurons to different subspaces[23] (see STAR Methods). Our analysis showed that around 34% of neurons contributed to the sensory subspace, 30% for memory-1, 30% for memory-2, and 26% for memory-3 subspace (Figures S2C and S2G for monkey G), suggesting that these subspaces are broadly distributed in the frontal population.

Next, we investigated the functional relationship between the sensory and rank-WM subspaces during the control process. We first compared the latency of the activities in the sensory and rank-WM subspaces and found that for each ordinal rank, the emergence of item information in the sensory subspace was significantly earlier than in the corresponding rank-WM subspace (Figures 2F and S2J; see STAR Methods), suggesting that item information first entered the sensory subspace and was subsequently routed to the associated rank-WM subspace. If this sequential relationship holds, we expect to observe a trial-by-trial correlation of neural activity between sensory and rank-WM subspaces. We used the single-trial decoding probability to quantify item information at the single-trial level. We found that for each rank, items during the sample period in the sensory subspace and the delay period in the corresponding rank-WM subspace showed a significant correlation (Figures 2G and S2K). As a control, for item information at other ordinal ranks (e.g., rank-1 item in sensory and rank-2 item in memory-2), there was no or minimal correlation (Figures 2G and S2K).

Thus, the results showed that the sensory and rank WMs could be represented in separate low-dimensional subspaces within the same frontal neuronal population,[21] even though they are

Figure 2. Dynamics of neural population responses in frontal subspaces in the forward task

(A) To examine if the hypothetic rank-shared sensory subspace (e.g., subspace 1) and rank-specific WM subspaces (e.g., subspace 2) exist in the neural data, we extended the static decoding approach in our prior work[23] to a dynamic method featuring adaptive decoding time windows and decoding targets (see STAR Methods), capable of addressing the intricacies of evolving dynamics in these low-dimensional subspaces. Four examples length-2 sequences were plotted (note that for length-2 trials, there are 30 sequences in total).

(B) Time courses of single-trial decoding accuracy for stimuli in different ranks (labeled at the bottom) in the four identified subspaces with length-3 trials for monkey O. 2-fold cross-validation was used for each session, and the average accuracy across all trials in all sessions was shown. Horizontal bars in each panel indicate above-chance decoding performance ($p < 0.001$, pixel-based permutation test, see STAR Methods).

(C) Neural activity trajectories in the identified subspaces with length-3 trials for monkey O. Unlike (B), we just showed trajectories of stimuli in certain ranks during certain periods. For the sensory subspace, the trajectory is colored according to the ordinal position during the corresponding sample period, with the delay period following the color of rank 3. For memory subspaces, the trajectory is colored according to the dominant ordinal rank (rank-1, -2, and -3 for memory subspace-1, -2, and -3, respectively).

(D and E) Principal angles (D) and VAF ratios (E) between the identified subspaces with length-3 trials for monkey O across sessions (see STAR Methods). As a control, the principal angle and VAF ratio between 2-fold estimations for the same subspace across sessions were computed, and they are significantly different from the actual principal angles and VAF ratios (see STAR Methods).

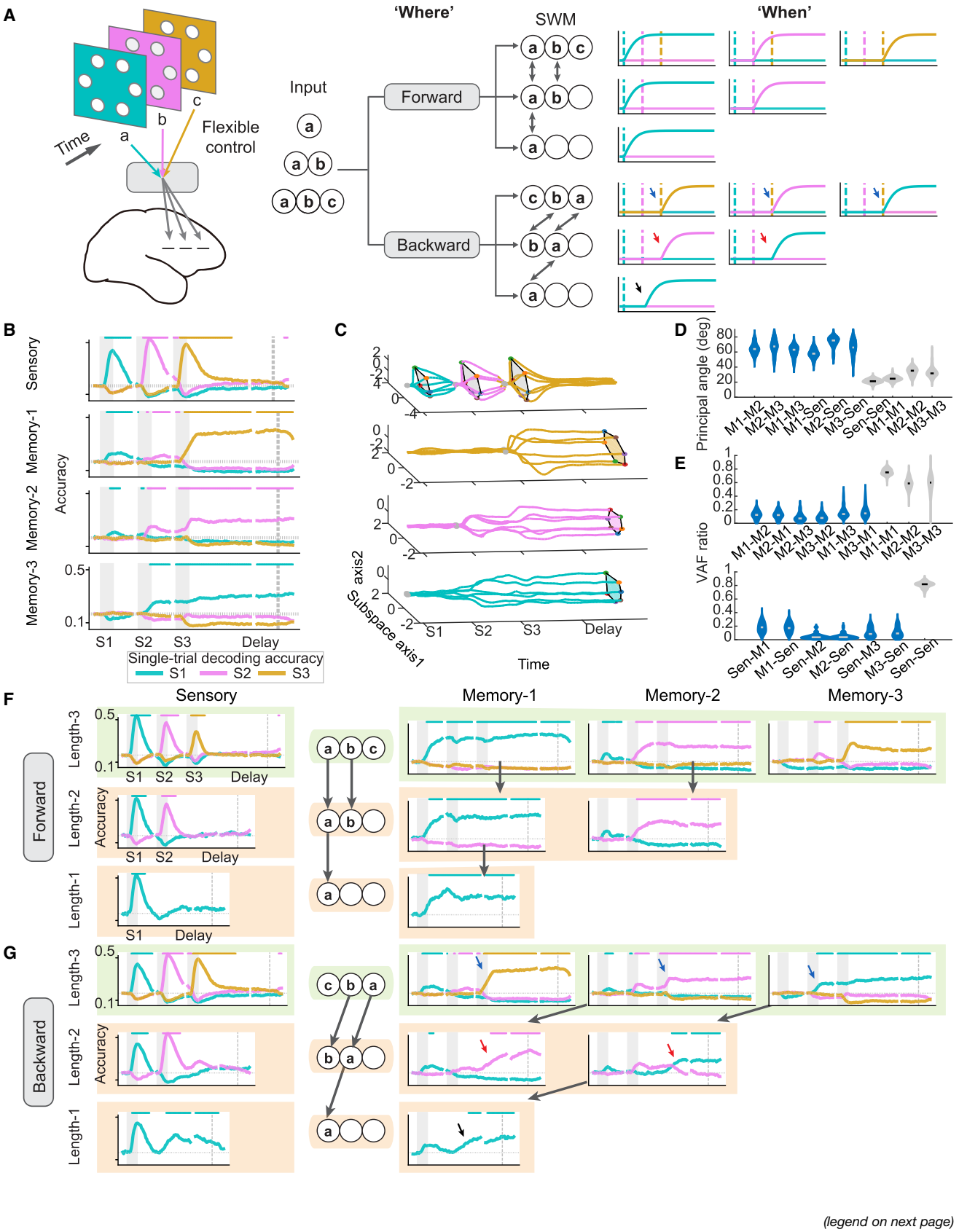
(F) Histograms showing the distribution of latency difference for three sensory-WM subspace pairs. For each sensory-WM subspace pair, the distribution of latency difference is generated through a bootstrap procedure (n = 1,000 bootstrap resamples of neurons; see STAR Methods). All latency differences were significantly greater than 0 ($p < 0.001$, Wilcoxon signed-rank test).

(G) Single-trial correlation of item information for sensory-WM subspace pairs. Item information is quantified by averaging the decoding probability during the sample (delay) period for sensory (WM) subspace (see STAR Methods). Each gray circle showed the correlation coefficient in a single session. All the paired correlation coefficients (S1 in sensory vs. S1 in memory-1, etc.) were significantly larger than 0, while all the unpaired correlation coefficients (S1 in sensory vs. S2 in memory-2 or S3 in memory-3, etc.) were not significantly larger than 0 except the S1 in sensory vs. S2 in memory-2 pair. The paired correlation coefficients were also significantly larger than the unpaired correlation coefficients. Wilcoxon signed-rank test $^{***}p < 0.001$, $^{**}p < 0.01$, and $^{*}p < 0.05$.

(H) Schematic of the control process in the frontal cortex in the forward task.

See also Figure S2.

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024



(legend on next page)

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

intensively mixed at the single-neuron level. The control of SWM in the forward task was reflected in the subspaces' neural dynamics. Specifically, each item was sequentially entered into a common (sensory) subspace and bound with its associated rank-WM subspaces for WM maintenance within the same frontal population (Figure 2H, schema).

Frontal neural states and flexible control in the backward task

Identical sequence inputs can be reordered, leading to different behavioral outputs depending on the "sorting algorithm"—e.g., forward or backward. In the present experiment, monkeys can sort sequences (Figure 1B), implying the existence of flexible control mechanisms within the frontal population that could select appropriate items into their sorted subspaces in SWM. We found that each rank's information was rapidly routed into its associated rank-WM subspace in the forward task. We then asked whether the backward control was also implemented in the frontal neurons. We proposed that similar neural subspaces (e.g., sensory and rank-WM subspaces) would be recruited and that their dynamics could reflect such flexible control.

In the backward task, the length-1, -2, and -3 sequences were randomized such that for each trial, the selection of items into the rank-WM subspaces in SWM depended on sequence length. Thus, to reverse a sequence, we predicted the WM system should control: (1) *which* rank subspace is the item gated into ("where-to-gate")? (2) When is the item gated into WM ("when-to-gate") (see the hypothesis in Figure 3A)? Consider a trial where items "a," "b," and "c" have been sequentially presented—in the forward task, the "a" ("b," "c") is directly gated into memory-1 (2, 3) subspace. In the backward task, the sequence "a b c" ("a b" for length-2) has to be reversed and stored as "c b a" ("b a" for length-2) in the delay. Thus, item "b" would be loaded into the memory-2 subspace (i.e., the ordinal position 2 in "c ***b*** a") in length-3 trials and into the memory-1 subspace (i.e., the ordinal position 1 in "***b*** a") in length-2 trials (*where-to-gate*). In addition, unlike the forward task, where the gating takes place immediately after the presentation of each item, the gating in the backward task can only be determined by the sequence length (e.g., the existence of the following items) (*when-to-gate*). As monkeys learned that the longest sequence in the block was three, the absence of item "b" or "c" and the presence of item "c" indicated the finish of the sequence, which could serve as a signal for gating. Therefore, we hypothesized that the gating time of item "a" would depend upon the presentation of items "b" and "c," and the gating time of item "b" would depend upon the existence of item "c." It is worth noting that, alternatively, monkeys could still memorize the sequence as "a b c" in the delay of backward task and reverse it later during the response period. In this case, the gating subspaces in the backward task would be identical to the forward task.

With the hypothesis above, using the same subspace identification method for the forward task, we examined the neural code of length-3 SWM in the backward task. As predicted, we also identified four subspaces: one sensory and three rank-WM subspaces (Figure 3B; see Figure S3A for the explained variance for each subspace). Similar to the forward task, the sensory subspace transiently encoded the item information after stimulus presentation (Figures 3B and 3C, first row), and the three rank-WM subspaces stably encoded three rank-WMs throughout the delay, respectively (Figures 3B and 3C, last three rows)—these subspaces were near orthogonal (Figures 3D and 3E) and highly distributed across the recorded neural population (Figures S3B and S4A). A similar result was held for the monkey G (Figures S3C–S3G and S4B).

To examine the flexibility of SWM control, we used generalization tests to examine how the control of where-to-gate is accomplished in the forward and backward tasks. As we predicted, in the forward task, each subspace (e.g., rmemory-1) identified with length-3 sequences could be successfully generalized to the same rank-WM subspace (e.g., memory-1) in both length-1 and -2 sequences (Figure 3F) and vice versa (Figure S3I). However, in the backward task, the rank-WM subspaces could also be well generalized across variable-length sequences but in a length-dependent and reversed manner (Figures 3G and S3I). For example, successful generalizations could be found between subspaces of rank-1 (length-1), rank-2 (length-2), and rank-3 (length-3) memories. This result is consistent with our aforementioned where-to-gate hypothesis but incompatible with the possibility that monkeys memorized the past sequence as "a b c" in the backward task.

We next investigated the control of when-to-gate in both tasks. In contrast to the forward task, the three rank-WM subspaces in the backward task showed distinct temporal dynamics. Specifically, for the length-2 sequence ("a b") in the backward task, the items were stably controlled to fall into memory-2 and -1 subspaces, respectively, just after confirming the absence of the third item (red arrows, Figure 3G, middle row). For the length-3 sequence ("a b c"), similar to the length-2 sequence, the items were guided into their associated rank-WM ("memory-3, memory-2, memory-1") subspaces after the onset of the third item (blue arrows, Figure 3G, first row). The only exception was that items "a" and "b" in length-3 sequences

Figure 3. Flexible control process of subspaces in the forward and backward task

(A) Hypothetic illustration of "where" and "when" flexible control. Although the input items (a, b, and c) remained the same in the forward and backward tasks, they were gated into certain subspaces ("where") at certain time points ("when") according to the sorting algorithm and the length of the sequence.

(B–E) Time courses of decoding accuracy (B), state-space trajectories (C), principal angles (D), and VAF ratios (E) in the backward task for monkey O. The principal angles and VAF ratios are significantly different from the control. The annotations were similar to Figure 2.

(F) Cross-length generalization test in the forward task. Decoders were trained on length-3 correct trials, and accuracy was averaged across all trials in all sessions. Top: cross-validated accuracy for length-3 trials. Middle and bottom: generalization test using length-2 and -1 trials, respectively. Horizontal bars indicated time windows where the decoding performance was significantly higher than the chance level ($p < 0.001$, pixel-based permutation test).

(G) Same as (F), results for the backward task.

See also Figure S3.

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

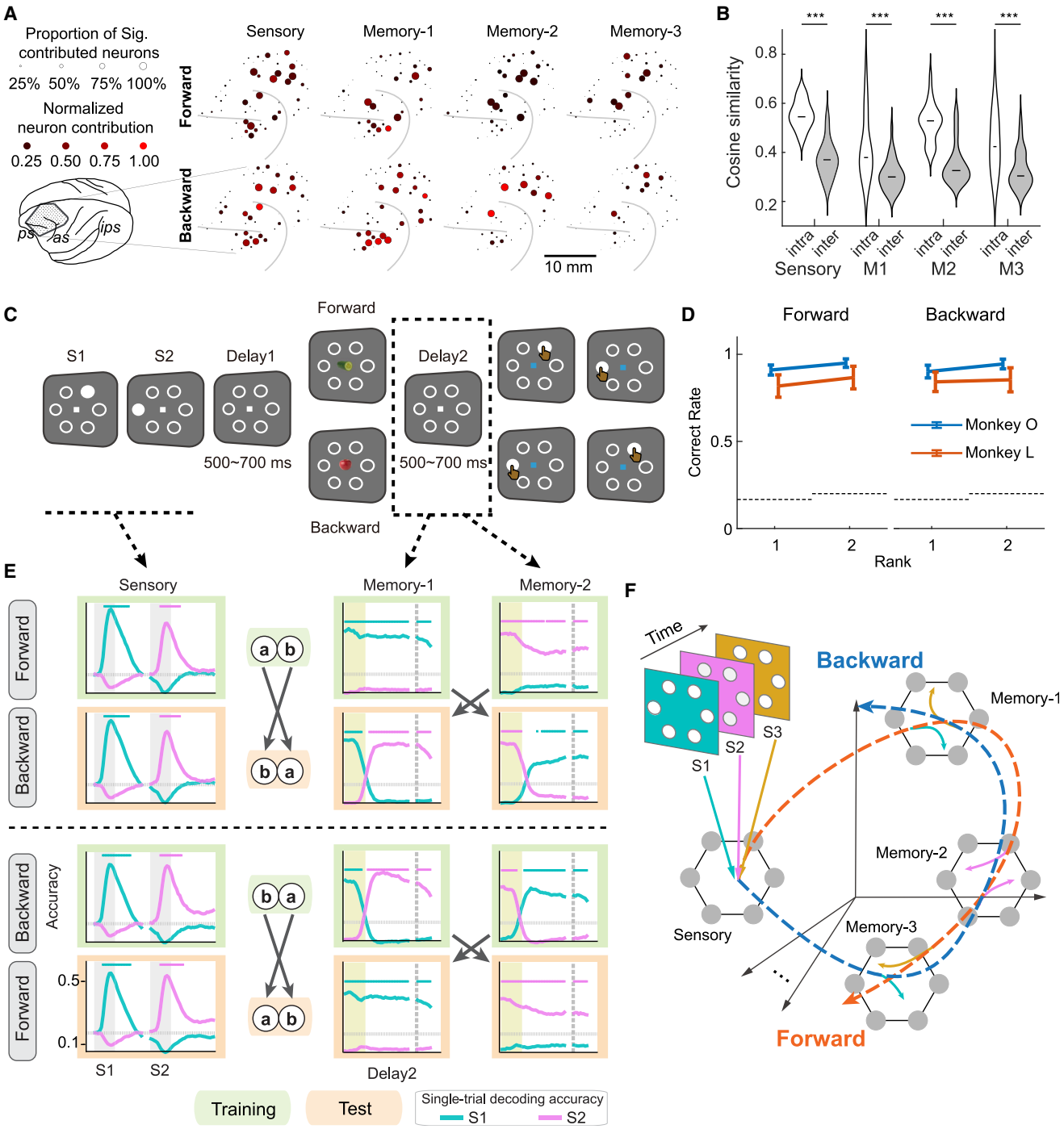


Figure 4. Abstract control for sequence sorting in frontal neural states

(A) Normalized neural contribution and proportion of significantly contributed neurons at the recording channel level (see STAR Methods) for both the forward (top) and backward (bottom) tasks were shown. Each circle represented one channel, its size corresponded to the proportion of neurons with a significant contribution, and its color was the normalized neural contribution.

(B) Cosine similarities of neural contribution (weighted by the proportion of significantly contributed neurons at the same channel) distribution for the intra-subspace pairs (opened curves, e.g., forward sensory vs. backward sensory, forward memory-1 vs. backward memory-1, etc.) as well as for inter-subspace pairs (filled curves, e.g., forward sensory vs. backward others, forward memory-1 vs. backward others, etc.). All the cosine similarities for the intra-subspace pairs were larger than those of the inter-subspace pairs. ***$p < 0.001$, t test.

(C) Forward and backward trials with rule cues. In these trials, we trained monkeys to perform the forward or backward task following a cue in the middle of the delay on a trial-by-trial basis and collected neural data using the same recording technique (see STAR Methods).

(legend continued on next page)

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

were loaded slightly earlier than expected (see discussion). A similar result was held for the second monkey (Figures S3H and S3J). We also noticed a gap period when the items disappeared from the sensory subspace and did not enter into the corresponding rank-WM subspace. Thus, after identifying the sensory and rank-WM subspaces in the backward task, we further searched for the extra subspaces and found some temporary subspaces that could temporarily hold the item information to potentially bridge the gap between sensory and rank-WM (Figures S3K and S3L). Therefore, these results supported our hypothesis, showing that sequence information flow in the frontal cortex could be flexibly and timely controlled, reflected by the compositional population codes with separate and generalizable low-dimensional sensory and rank-WM subspaces and their length-dependent neural dynamics.

Abstract control for sequence sorting in frontal neural states

Do the SWMs in the forward and backward tasks share the same neural subspaces? Furthermore, does the same control system coordinate the two sorting processes? As the neurons in the two tasks were recorded in separate blocks but with the same chronic recording electrodes, we then computed the distribution of neural contribution (DNC) for different subspaces in the forward and backward tasks at the level of recording channels (Figures 4A and S4C). Compared with DNC at the single-neuron level (Figures S4A and S4B), DNC at the channel level is obtained by averaging the contribution of single neurons within the same channel (see STAR Methods). We discovered a remarkable similarity in the neural substrates of SWM subspaces between the forward and backward tasks, evident by significantly higher similarities in DNCs between the intra-subspace pairs (e.g., forward sensory vs. backward sensory, forward memory-1 vs. backward memory-1, etc.) compared with the inter-subspace pairs (forward sensory vs. backward others, forward memory-1 vs. backward others, etc.) (Figures 4B and S4D).

To better confirm that the sensory and SWM subspaces are shared between the two sequence sorting tasks, we further trained two monkeys (O and L) to perform the forward or backward task on a trial-by-trial basis (Figure 4C). On each trial, in the middle of the delay, a visual cue told monkeys whether to reproduce the sequence forward or backward. Both monkeys performed the tasks well (Figure 4D). The same population of neurons (new datasets, 3,219 neurons for monkey O, 727 neurons for monkey L) was recorded for both tasks (see STAR Methods). For the SWM subspaces, we focused on the neural activity in the second delay period (Figure 4C), where the memory-1 and memory-2 subspaces were either maintained or reversed by the control system. If the two tasks share the same memory control system, the subspaces should generalize to each other based on the sorting cue (Figures 4E and S4E). As predicted, we found that the memory-1 subspace in the forward task could be generalized to the memory-2 subspace in the backward task, and the memory-2 (forward) could be transferred to the memory-1 subspace (backward).

Thus, we propose that the abstract and flexible control coordinates the neural dynamics of how item information is carried across sensory and rank-WM subspaces to enable general sequence sorting processes (Figure 4F, schema). It is worth noting that the details of the item information, such as the spatial locations, are represented by the activities within each corresponding subspace in our data, implying that this abstract control guides the general flow of information between neural subspaces, agnostic to the details of represented items.

Activities in SWM subspaces reflect the control behavior

Finally, if the identified neural dynamics of SWM indeed represented the control process, they should predict the monkeys' behavior—whether that behavior is correct or erroneous. By taking advantage of recording hundreds of neurons simultaneously (one session), we next performed the decoding analysis for the error trials at the single-trial level (see STAR Methods). In both tasks, monkeys mainly made two types of errors: order and item errors. We expected that the neural dynamics in the corresponding subspaces could faithfully predict the error information in single trials.

We first discovered that the decoding accuracy in the sensory subspace on error trials was similar to that of correct trials (backward: Figures 5A–5C), implying that the sensory information was propagated to and represented successfully in the frontal cortex even when monkeys made errors. We next focused on decoding performance in the rank-WM subspaces in backward trials, as there were more error trials. For the order errors, most of them were found in the swap of items between ranks-2 and -3 (69.9%, 900/1,288 trials), e.g., the SWM "c **b a**" was mistaken for "c **a b**." Remarkably, the decoding trajectory of neural activity showed a similar swap (Figure 5B), indicating that the control system gated item information into wrong rank-WM subspaces during the delay prior to producing the error. For example, item "b" was supposed to be in the memory-2 subspace in the correct trials (Figure 5A), but on error trials, this item was found in the memory-3 subspace, and correspondingly, item "a" was found in thememory-2 subspace (Figure 5B).

For item errors, most trials were found at rank 3 (94.0%, 856/911 trials), e.g., the SWM "c b **a**" was mistaken for "c b **x**," where "a" is the target item and "x" is another (incorrect) response location. Accordingly, the neural activity in the memory-3 subspace showed high decoding accuracy for the incorrect item ("x") (Figure 5C), indicating that the control system gated incorrect items into associated rank-WM subspaces. We further differentiated these location error trials using the distance between the target ("a") and the response ("x") (Figure 5D). We

(D) Correct rate as a function of ordinal rank (error bar: SD across sessions; dashed lines: chance level); all $p < 0.001$, two-tailed t test.
(E) Cross-sorting-algorithm generalization test for both sensory and rank-specific subspaces. Top: forward-to-backward generalization test. Bottom: backward-to-forward generalization test. Horizontal bars indicated time windows where the decoding performance was significantly higher than the chance level ($p < 0.01$, pixel-based permutation test).
(F) Schematic of the control process in the forward and backward task.
See also Figure S4.

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

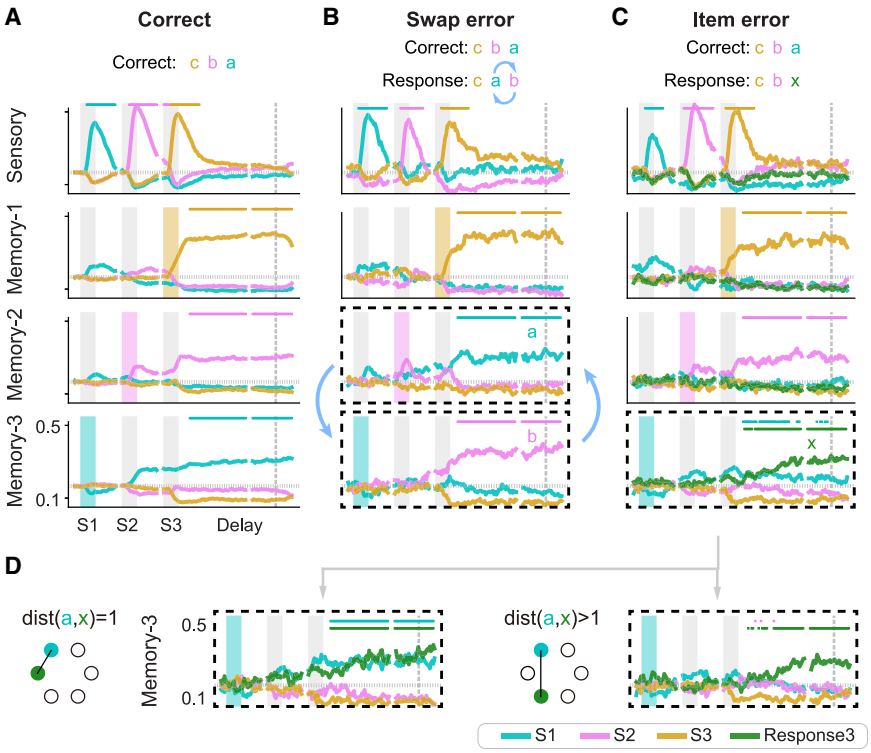


Figure 5. Error trials in the backward task

(A) Decoding accuracy in correct trials estimated the correct trial averaged across all trials in all sessions. The same decoders as in Figure 3B. Horizontal bars indicate time windows where the decoding performance is significantly higher than chance ($p < 0.001$, pixel-based permutation test). We only tested the significance during the sample period for sensory subspace and the delay period for memory subspaces ($p < 0.01$, pixel-based permutation test).

(B) Decoding accuracy in trials with ordinal swap error ($p < 0.01$, pixel-based permutation test).

(C) Decoding accuracy in trials with item error ($p < 0.01$, pixel-based permutation test).

(D) Subtypes of item error. Item error trials were further separated according to the distance between the correct and reported locations ($p < 0.01$, pixel-based permutation test).

See also Figure S5.

found that when the response item was the neighbor location of the target (e.g., distance (a,x) = 1), both the target and response items could be decoded throughout the delay (Figure 5D, left), suggesting an interference or competition between the neighbor items in SWM. Importantly, when the distance was long (e.g., distance (a,x) > 1), the target item was not gated into SWM subspaces, and the response item seemed to be conjured up at the late stage of the delay period (Figure 5D, right). This thus perfectly illustrated the control process in SWM, even when the controlled item was an internally generated guess. The decoding trajectories in rank-WM subspaces for the two monkeys in all error trials, including other ranks and lengths, and errors in the forward task are shown in Figure S5.

DISCUSSION

By recording thousands of neurons in the frontal cortex of macaque monkeys performing visuospatial sequence sorting (forward and backward) tasks, we revealed compositional population codes of SWM and their well-orchestrated neural dynamics in the frontal cortex, supporting the flexible control of SWM: item information first entered into a common subspace and subsequently, depending on the sorting algorithm, was flexibly bound with the corresponding rank-associated SWM subspace. The neural dynamics of item information—carried across sensory and rank-WMs subspaces across two tasks—reflected a general and abstract control system of SWM. Importantly, single-trial analyses using error trials showed that neural dynamics in SWM subspaces could accurately predict monkeys' error types (order/item), bridging the gap between the control process in the frontal cortex and sequence task behavior.

The frontal control for information binding in SWM

The distinct sequential neural dynamics of the identified sensory and rank-WM subspaces in different tasks indicate the presence of a control system that utilizes unique activity patterns of frontal neuronal subspaces to support flexible sequence-ordering behavior.[26,27] However, the control process may highly depend on the task structures that animals acquire during learning. For example, in the backward task, information in length-3 sequences was loaded into SWM subspaces slightly earlier than expected. The exact reason underlying this deviation remains to be determined. It may reflect the monkey's specific strategy when facing a challenging cognitive control problem. For example, because the proportion of the length-3 sequence was relatively higher than the sequence with other lengths, loading relevant information into the memory-3 subspace ahead may be able to alleviate the heavy cognitive control demand at the time point of rank-3 onset.

More importantly, we showed that the information flow is controlled to the proper subspace (*where*) at a specific moment (*when*) across tasks, irrespective of the incoming item (e.g., spatial locations). This fact suggests the existence of abstract contextual representations, which only guide the general flow of information between neural subspaces, could be agnostic to the detailed concrete item information. Such abstract control is essential, as it implies that the same control system could broadly apply to various cognitive structures with a progression through ordinal sets, where different stimuli, tasks, and memories are potentially embedded.

Where is the control circuitry in the brain? Previously, it has been assumed that sensory inputs are processed in sensory areas, converging onto the prefrontal cortex, and the reciprocal connection between sensory and prefrontal networks is what sustains and controls representations in WM.[21] That is, the sensory input is controlled before reaching prefrontal subspaces by

the nature of reciprocal connections. In this study, instead, we showed that the control could be implemented within the frontal cortex, represented by separate and generalizable low-dimensional sensory and rank-WM subspaces. Therefore, the sensory inputs, WMs, and their controls seem to be instantiated by the same frontal circuitry.[22] However, due to the lack of recordings from the sensory cortex, the present results do not exclude the simultaneous presence of the control representation within the reciprocal connections between brain regions (e.g., between visual and prefrontal areas and/or between prefrontal and basal ganglia regions[15,28,29]). Multiple control circuitries for SWM may coexist within the brain.

SWM vs. motor planning

Work memory serves as the buffer between past sensations and future actions, making it vital to understand not only how the brain encodes and retains sensory information but also how it plans for its upcoming use. In the current task, the identified activities in rank-WM subspaces can also be regarded as prospective sequential actions selected from memory. Future analyses and experiments need to explore the dissociation of sensory representations and action plans.[30,31] In the field of motor control, previous research has shown neural response patterns of many cortical areas regarding the transformation from sensory to single or sequential motor planning.[32–34] Previous single-neuron studies showed rich item-, rank-, or sequence-selective neurons.[7,18–20] Furthermore, the orthogonal representations were found in the last seminal work, e.g., between motor preparation and action,[8] between WM and motor preparations,[17] and between the memories in different task contexts.[12] However, there is still a lack of a complete framework of how the sequential control process is implemented in the frontal neural populations. The extrapolation from planning a single motor action to planning a series of motor actions is highly non-trivial. This is because when there are multiple planned motor actions or memories, each time a sensory input comes in, depending on the context or rule, the brain has to select one appropriate memory subspace among multiple subspaces, involving a series of complex gate-in (-off) computations.[4,5] For example, in the backward task, for sequence "a b," the action "b" has to be routed into the memory-1 subspace, requiring the memory-1 subspace to be in the open state while all other memory subspaces are in the closed state. However, for sequence "a b c," the same action "b" has to be routed into the memory-2 subspace, thereby requiring the circuit in a completely different state (i.e., memory-2 subspace in the open state while all other memory subspaces in the close state). This gate-in process, shown in this study, features rich and well-orchestrated dynamics involving a complex control process, perfectly aligned with a series of motor neuroscience works in which the preparatory null space is described to be "expansive." We thus postulate that, after this expansive transformation, the SWM, with the compositional representation, can be conveniently retrieved step-by-step through the neural dynamics, irrespective of the sequence length or trial types. In future analyses, the relationship between SWM (or sequence motor planning) and executive motor actions needs to be systematically investigated, particularly during the period of SWM retrieval.

Flow control by interactions between hypothetical frontal subpopulations

An advancing frontier in cognitive neuroscience involves the identification of non-random mixed-selectivity subpopulations with the frontal cortex during flexible mental computations.[35–37] In an effort to better understand how the frontal circuits learn and achieve flexible SWM control, we made the following postulations: (1) multiple mixed-selectivity subpopulations may serve as the structural underpinning for the control of information flow in neural hyperspaces. (2) Items could be selected to corresponding rank-WM hyperspaces by multiplying their weight matrices, analogous to the permutation operation in vector symbolic architectures.[38,39] (3) This selection process could be enabled through a gain-controlled modulation mechanism, in which ordinal structure in the long term memory serves as the internal context (in contrast to the external context in previous works[22]) to tune the gain of different subpopulations. Presumably, the selection process is predetermined by the anatomical connectivity between different subpopulations, likely formed during the animal's task learning.[40] Nevertheless, this proposal necessitates further rigorous investigation in the broad domain of cognitive control and neurosymbolic computations, holding the potential for addressing the intricate challenge of information flow control across neural populations[41] or brain areas.[42]

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Behavior task
  - Large-scale recording system and surgeries
  - Electrophysiological recording
  - Single neuron selectivity
  - Subspace decomposition
  - Generalized linear regression, subspace identification and decoding
  - Geometric relationship between subspaces
  - Neuronal contributions to subspaces in the single neuron level
  - Neuronal contributions to subspaces in the recording channel level
  - Forward and backward trials with rule cues
  - Error trial analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical test of decoding performance in each subspace
  - Latency difference for sensory-memory subspace pairs
  - Correlation of decoding probabilities for sensory-memory subspace pairs
  - Statistical test of principal angles and VAF ratios

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2024.07.024.

ACKNOWLEDGMENTS

We thank Stanislas Dehaene for his critical comments on the manuscript. This work was supported by the STI2030-Major Project (2021ZD0204102), the National Science Fund for Distinguished Young Scholars (32225022), the CAS Project for Young Scientists in Basic Research (YSBR-071), the Shanghai Municipal Science and Technology Major Project 2021SHZDZX, and the Special Funds of the National Natural Science Foundation of China (grant no. T2341003) to L.W. and the STI2030-Major Project (2021ZD0204105) and the National Science Foundation of China 32271149 to B.M.

AUTHOR CONTRIBUTIONS

Conceptualization, L.W.; methodology, J.C., C.Z., P.H., B.M., and L.W.; investigation, J.C., C.Z., and L.W.; visualization, J.C., C.Z., and P.H.; writing – original draft, J.C., C.Z., P.H., B.M., and L.W.; writing – review and editing, J.C., C.Z., P.H., B.M., and L.W.; funding acquisition, B.M. and L.W.; supervision, L.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 10, 2023
Revised: April 10, 2024
Accepted: July 29, 2024
Published: August 22, 2024

REFERENCES

1. Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S., and Wang, L. (2018). Production of supra-regular spatial sequences by macaque monkeys. Curr. Biol. *28*, 1851–1859.e4. https://doi.org/10.1016/j.cub.2018.04.047.
2. Fitch, W.T. (2010). The Evolution of Language (Cambridge University Press). https://doi.org/10.1017/CBO9780511817779.
3. Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. Neuron *88*, 2–19. https://doi.org/10.1016/j.neuron.2015.09.019.
4. O'Reilly, R.C., Braver, T.S., and Cohen, J.D. (1999). A biologically based computational model of working memory. In Models of Working Memory: Mechanisms of Active Maintenance and Executive Control, A. Miyake and P. Shah, eds. (Cambridge University Press), pp. 375–411. https://doi.org/10.1017/CBO9781139174909.014.
5. Cohen, J.D., Braver, T.S., and O'Reilly, R.C. (1996). A computational approach to prefrontal cortex, cognitive control and schizophrenia: recent developments and current challenges. Philos. Trans. R. Soc. Lond. B Biol. Sci. *351*, 1515–1527. https://doi.org/10.1098/rstb.1996.0138.
6. Halsband, U., Matsuzaka, Y., and Tanji, J. (1994). Neuronal activity in the primate supplementary, pre-supplementary and premotor cortex during externally and internally instructed sequential movements. Neurosci. Res. *20*, 149–155. https://doi.org/10.1016/0168-0102(94)90032-9.
7. Shima, K., and Tanji, J. (2000). Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. J. Neurophysiol. *84*, 2148–2160. https://doi.org/10.1152/jn.2000.84.4.2148.
8. Elsayed, G.F., Lara, A.H., Kaufman, M.T., Churchland, M.M., and Cunningham, J.P. (2016). Reorganization between preparatory and movement population responses in motor cortex. Nat. Commun. *7*, 13239. https://doi.org/10.1038/ncomms13239.
9. Zimnik, A.J., and Churchland, M.M. (2021). Independent generation of sequence elements by motor cortex. Nat. Neurosci. *24*, 412–424. https://doi.org/10.1038/s41593-021-00798-5.
10. Churchland, M.M., and Shenoy, K.V. (2024). Preparatory activity and the expansive null-space. Nat. Rev. Neurosci. *25*, 213–236. https://doi.org/10.1038/s41583-024-00796-z.
11. Kaufman, M.T., Churchland, M.M., Ryu, S.I., and Shenoy, K.V. (2014). Cortical activity in the null space: permitting preparation without movement. Nat. Neurosci. *17*, 440–448. https://doi.org/10.1038/nn.3643.
12. Panichello, M.F., and Buschman, T.J. (2021). Shared mechanisms underlie the control of working memory and attention. Nature *592*, 601–605. https://doi.org/10.1038/s41586-021-03390-w.
13. Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167.
14. Cohen, J.D., Dunbar, K., and McClelland, J.L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. Psychol. Rev. *97*, 332–361. https://doi.org/10.1037/0033-295x.97.3.332.
15. O'Reilly, R.C., and Frank, M.J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. Neural Comput. *18*, 283–328. https://doi.org/10.1162/089976606775093909.
16. Pasupathy, A., and Miller, E.K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature *433*, 873–876. https://doi.org/10.1038/nature03287.
17. Tang, C., Herikstad, R., Parthasarathy, A., Libedinsky, C., and Yen, S.C. (2020). Minimally dependent activity subspaces for working memory and motor preparation in the lateral prefrontal cortex. eLife *9*, e58154. https://doi.org/10.7554/eLife.58154.
18. Christophel, T.B., Klink, P.C., Spitzer, B., Roelfsema, P.R., and Haynes, J.D. (2017). The distributed nature of working memory. Trends Cogn. Sci. *21*, 111–124. https://doi.org/10.1016/j.tics.2016.12.007.
19. Barone, P., and Joseph, J.P. (1989). Prefrontal cortex and spatial sequencing in macaque monkey. Exp. Brain Res. *78*, 447–464. https://doi.org/10.1007/BF00230234.
20. Ninokura, Y., Mushiake, H., and Tanji, J. (2003). Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. J. Neurophysiol. *89*, 2868–2873. https://doi.org/10.1152/jn.00647.2002.
21. Bouchacourt, F., and Buschman, T.J. (2019). A flexible model of working memory. Neuron *103*, 147–160.e8. https://doi.org/10.1016/j.neuron.2019.04.020.
22. Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84. https://doi.org/10.1038/nature12742.
23. Xie, Y., Hu, P., Li, J., Chen, J., Song, W., Wang, X.J., Yang, T., Dehaene, S., Tang, S., Min, B., et al. (2022). Geometry of sequence working memory in macaque prefrontal cortex. Science *375*, 632–639. https://doi.org/10.1126/science.abm0204.
24. Zhang, H., Zhen, Y., Yu, S., Long, T., Zhang, B., Jiang, X., Li, J., Fang, W., Sigman, M., Dehaene, S., et al. (2022). Working memory for spatial sequences: developmental and evolutionary factors in encoding ordinal and relational structures. J. Neurosci. *42*, 850–864. https://doi.org/10.1523/JNEUROSCI.0603-21.2021.
25. Dotson, N.M., Hoffman, S.J., Goodell, B., and Gray, C.M. (2017). A large-scale semi-chronic microdrive recording system for non-human Primates. Neuron *96*, 769–782.e2. https://doi.org/10.1016/j.neuron.2017.09.050.
26. MacDowell, C.J., Tafazoli, S., and Buschman, T.J. (2022). A Goldilocks theory of cognitive control: balancing precision and efficiency with low-dimensional control states. Curr. Opin. Neurobiol. *76*, 102606. https://doi.org/10.1016/j.conb.2022.102606.
27. Badre, D., Bhandari, A., Keglovits, H., and Kikumoto, A. (2021). The dimensionality of neural representations for control. Curr. Opin. Behav. Sci. *38*, 20–28. https://doi.org/10.1016/j.cobeha.2020.07.002.

28. Poulet, J.F.A., Fernandez, L.M.J., Crochet, S., and Petersen, C.C.H. (2012). Thalamic control of cortical states. Nat. Neurosci. *15*, 370–372. https://doi.org/10.1038/nn.3035.

29. Chatham, C.H., and Badre, D. (2015). Multiple gates on working memory. Curr. Opin. Behav. Sci. *1*, 23–31. https://doi.org/10.1016/j.cobeha.2014.08.001.

30. Takeda, K., and Funahashi, S. (2002). Prefrontal task-related activity representing visual cue location or saccade direction in spatial working memory tasks. J. Neurophysiol. *87*, 567–588. https://doi.org/10.1152/jn.00249.2001.

31. Jonikaitis, D., Noudoost, B., and Moore, T. (2023). Dissociating the contributions of frontal eye field activity to spatial working memory and motor preparation. J. Neurosci. *43*, 8681–8689. https://doi.org/10.1523/JNEUROSCI.1071-23.2023.

32. Funahashi, S., Bruce, C.J., and Goldman-Rakic, P.S. (1990). Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. J. Neurophysiol. *63*, 814–831. https://doi.org/10.1152/jn.1990.63.4.814.

33. Barash, S., Bracewell, R.M., Fogassi, L., Gnadt, J.W., and Andersen, R.A. (1991). Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. J. Neurophysiol. *66*, 1095–1108. https://doi.org/10.1152/jn.1991.66.3.1095.

34. Funahashi, S., Bruce, C.J., and Goldman-Rakic, P.S. (1991). Neuronal activity related to saccadic eye movements in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. *65*, 1464–1483. https://doi.org/10.1152/jn.1991.65.6.1464.

35. Hirokawa, J., Vaughan, A., Masset, P., Ott, T., and Kepecs, A. (2019). Frontal cortex neuron types categorically encode single decision variables. Nature *576*, 446–451. https://doi.org/10.1038/s41586-019-1816-9.

36. Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F., and Ostojic, S. (2022). The role of population structure in computations through neural dynamics. Nat. Neurosci. *25*, 783–794. https://doi.org/10.1038/s41593-022-01088-4.

37. Vyas, S., Golub, M.D., Sussillo, D., and Shenoy, K.V. (2020). Computation through neural population dynamics. Annu. Rev. Neurosci. *43*, 249–275. https://doi.org/10.1146/annurev-neuro-092619-094115.

38. Kanerva, P. (2009). Hyperdimensional computing: an introduction to computing in distributed representation with high-dimensional random vectors. Cogn. Comput. *1*, 139–159. https://doi.org/10.1007/s12559-009-9009-8.

39. Kleyko, D., Davies, M., Frady, E.P., Kanerva, P., Kent, S.J., Olshausen, B.A., Osipov, E., Rabaey, J.M., Rachkovskij, D.A., Rahimi, A., et al. (2022). Vector symbolic architectures as a computing framework for emerging hardware. Proc. IEEE Inst. Electr. Electron. Eng. *110*, 1538–1571. https://doi.org/10.1109/JPROC.2022.3209104.

40. Gurnani, H., and Cayco Gajic, N.A. (2023). Signatures of task learning in neural representations. Curr. Opin. Neurobiol. *83*, 102759. https://doi.org/10.1016/j.conb.2023.102759.

41. Semedo, J.D., Zandvakili, A., Machens, C.K., Yu, B.M., and Kohn, A. (2019). Cortical areas interact through a communication subspace. Neuron *102*, 249–259.e4. https://doi.org/10.1016/j.neuron.2019.01.026.

42. Siegel, M., Buschman, T.J., and Miller, E.K. (2015). Cortical information flow during flexible sensorimotor decisions. Science *348*, 1352–1355. https://doi.org/10.1126/science.aab0551.

43. Gallego, J.A., Perich, M.G., Naufel, S.N., Ethier, C., Solla, S.A., and Miller, L.E. (2018). Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. Nat. Commun. *9*, 4233. https://doi.org/10.1038/s41467-018-06560-z.

44. Gao, P., Trautmann, E., Yu, B., Santhanam, G., Ryu, S., Shenoy, K., and Ganguli, S. (2017). A theory of multineuronal dimensionality, dynamics and measurement. Preprint at bioRxiv. https://doi.org/10.1101/214262.

45. Cohen, M.X. (2014). Analyzing Neural Time Series Data: Theory and Practice (The MIT Press). https://doi.org/10.7551/mitpress/9609.001.0001.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Eye tracking camera	SR Research	EyeLink 1000
Headpost and Microdrive electrodes array	Gray Matter Research	https://www.graymatter-research.com/
Neural recording data acquisition system	Plexon OmniPlex system	https://plexon.com/
Offline sorter	Plexon OmniPlex system	https://plexon.com/
Deposited data		
Neural data	This paper	https://doi.org/10.5281/zenodo.13119704
Experimental models: Organisms/strains		
Macaca mulatta monkeys	Institute of Neuroscience, Key Laboratory of Brain Cognition and Brain-Inspired Intelligence Technology, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences	N/A
Software and algorithms		
MATLAB	Mathworks	https://uk.mathworks.com/products/matlab.html
Python version 3.9.13	N/A	https://www.python.org/
Pytorch 1.12.1	N/A	https://pytorch.org/
IronClust	Flatiron Institute	https://github.com/flatironinstitute/ironclust
Psychtoolbox	N/A	http://psychtoolbox.org/
Codes	This paper	https://doi.org/10.5281/zenodo.13119667

RESOURCE AVAILABILITY

Lead contact

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

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data have been deposited at Zenodo and are publicly available as of the date of publication. DOIs are listed in the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Three male rhesus monkeys (Macaca mulatta, 7∼8 years old, 7∼10 kg) were used in our experiments. All experimental procedures and surgeries followed the requirements of the ethical committee of the Institute of Neuroscience, Chinese Academy of Sciences.

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

METHOD DETAILS

Behavior task

Monkeys were seated in a primate chair facing a 21.5" LED monitor. Eye positions were tracked by an infrared eye-tracking system (Eyelink, SR Research Ltd) at a sampling rate of at least 250 Hz.

Monkeys initiated a trial by pulling a level and fixating a yellow square at the center of the screen. Once fixated, the square would turn white, and six circles (2° in diameter, 11° away from the center) would appear after 100 ms. After monkeys gazed at the white square for another 500 ms, one, two, or three stimuli (flashing white dots chosen from six circles) were displayed sequentially during the sample period. Each stimulus was displayed for 250 ms, with a random interval of 300 ms to 500 ms in between. After a delay period (1150∼1500 ms), the white square at the center turned blue ("go" cue), freeing monkeys from fixation and the pulling of the level so that monkeys could make the report by touching the locations on the screen during the reproduction period. A trial would be terminated immediately when one of the following conditions was met: (1) monkeys released the level before the go cue appeared; (2) monkeys looked outside the fixation window (2.7°) before the go cue. For the first neural dataset, in different blocks, monkeys were required to memorize and touch the locations in either the presented order (*forward* task) or the reverse order (*backward* task) of the sequence shown during the sample period. When monkeys touched correctly, the touched location would flash in white color. Once an error location was reported, the touched location flashed in blue, and the trial ended. A drop of water or juice was given when the monkeys correctly reported the whole sequence. The forward and backward tasks were recorded in separate sessions (days). The spatial sequence and the sequence length were pseudo-randomly selected on each trial. For monkeys O and G, length-1, -2, and -3 sequences were used during the electrophysiological recordings. For monkey L, only length-2 sequences were applied.

Large-scale recording system and surgeries

We made simultaneous neural recordings in multiple brain regions by implanting the semi-chronic microdrive recording system (Gray Matter Research, USA)[25] in the left hemispheres of three monkeys. A 157-channel microdrive electrode array (LS-157, tungsten electrodes, AlphaOmega, ∼1 MΩ, 1.5 mm inter-electrode spacing) was used in the frontal cortex. Before surgeries, we first defined the recording chambers' locations and head holders' locations. The recording chamber was localized to cover the prefrontal and premotor cortices as large as possible, and the head holder was localized at the right hemisphere and fixed to the skull with bone cement and ceramic screws. Based on computed tomography and magnetic resonance imaging for the head of each monkey, the chamber was form-fitted to the surface of the skull, and the microdrive was form-fitted to the inside brain surface.

After the implantation of the head holder, the following surgeries, including chamber implantation, craniotomy, and microdrive implantation, were performed. During chamber implantation, the chamber was anchored to the skull with bone cement and screws (titanium), similar to the implantation of the head holder. To seal the junction between the chamber and the skull, the adhesive resin cement (3M Inc., USA) was applied with Single Bond Universal Adhesive (3M Inc., USA). About one week after the chamber implantation, the fluid in the chamber was swabbed and cultured to ensure sterility. Only the skull within the chamber was removed, and the boundary of the craniotomy was trimmed to keep aligned with the internal wall of the chamber. A dummy plug (3∼5 mm longer than the depth of the chamber) was used to check the alignment, and a sterility test was also performed after about one week. Before the implantation of the microdrive, the dura within the chamber was removed and trimmed with micro scissors and forceps. Then the microdrive was implanted and mounted to the chamber with screws. A dura gel was filled between the bottom of the microdrive and the cortices as protection.

Two days after the implantation of the microdrive, we began to advance the electrodes. Firstly, the electrodes were lowered fast until their impedance (impedance tester, nanoZ) fell in the normal range due to the electric insulative dura gel. Then, we moved the electrodes slowly, checked the spike waveform online, and stopped when the well-isolated spike waveform was picked up. All the electrodes were lowered into the surface of the cortices within one week in case dura hyperplasia would block the movement of the electrodes. During the recording sessions, the depths of the electrodes were adjusted appropriately to maximize the amount of the simultaneously recorded units.

Electrophysiological recording

We recorded the broadband electrophysiological signals at 40 kHz through a neural recording data acquisition system (OmniPlex, Plexon Inc.). Eye position signals and event markers were transported to the recording system in real time and stored with electrophysiological signals. The high-frequency signals (Butterworth filter, 300∼8000 Hz) were automatically sorted offline with Ironclust and manually curated with an offline sorter (Plexon Inc.). Single and multi-units were combined for the data analysis, and units with mean firing rates less than 1.2 Hz were kicked out.

Single neuron selectivity

To evaluate the tuning properties of single neurons, we first defined four distinct periods using length-3 trials: the onset of each stimulus extending to 50 ms after the corresponding stimulus offset (each spanning 300 ms, repeated for three stimulus periods, referred to as the sample period), as well as 300 ms window before the go cue (referred to as the delay period). Then, we computed the mean firing rate for each neuron for each period and conducted an ANOVA test accordingly. Neurons displaying selectivity to the present spatial location during the sample period (ANOVA, $p < 0.01$ for monkey O; $p < 0.001$ for monkey G; $p < 0.05$ for monkey L) were

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

defined as stimulus-related, while neurons exhibiting selectivity during the delay period (ANOVA, $p < 0.01$ for monkey O; $p < 0.001$ for monkey G; $p < 0.05$ for monkey L) were identified as memory-related. Note that stimulus- and memory-related properties are not mutually exclusive; a neuron can possess both stimulus- and memory-related tuning characteristics. We shuffled item labels for each trial 1000 times and performed the ANOVA test to get the distribution of shuffled proportion for each group of neurons. Shuffled proportions for different groups of neurons were calculated by the unique criteria for the corresponding group. All the proportions were significantly higher than the shuffle level ($p<0.001$, t-test).

Subspace decomposition

To evaluate our hypothesis of rank-shared sensory subspace and rank-specific WM subspaces, we extended the static linear decoding approach in our prior SWM study to a method that addressed the intricacies of evolving dynamics. At its core is an adaptive scheme to determine the time windows within which the representation of task relevant variables appear in each subspace. The time windows and target variables were first estimated by pseudo-population data aggregated from multiple sessions, and then fixed to train decoders in single sessions. The final results reported in the paper were based on these single-session decoders.

To better introduce this dynamic method, we first recapitulated the static linear decoding approach for a single time point.[23] Single-trial neural activity ($\boldsymbol{x}$) at time t was linearly projected (W, weight; b, bias) into a two-dimensional subspace

$$\boldsymbol{h} = W\boldsymbol{x} + b.$$

The hidden state ($\boldsymbol{h}$) was classified against the target matrix (M) to obtain softmaxed scores ($\boldsymbol{p}$) for all items

$$\boldsymbol{p} = \mathrm{softmax}(M\boldsymbol{h}).$$

At time t, the loss for decoding the rank-r item was defined as

$$loss_{\Theta}(r,t) = \frac{1}{n}\sum_{i}^{n}\mathrm{CrossEntropyLoss}(\boldsymbol{p}(t,i),\delta_{i_r}) + \mathrm{regularization},$$

where $\Theta = \{W, b, M\}$ is the learnable model parameter set, n is the total trial number, i_r is the index for rank-r item in the ith trial, and regularization refers to penalization on L2 norm of M and normalization of W. The model parameters were then trained through gradient descent to minimize this loss.

When taking into account the temporal dynamics, the subspaces should represent task variables for an extended period of time. The problem then became when and what task variables are represented. Therefore, the loss for identifying the subspace in the dynamic setting can be expressed as the temporal-weighted sum of the loss at different times:

$$Loss = \frac{1}{T}\sum_{t}^{T} w(t)\cdot loss_{\Theta}(\hat{r}(t),t),$$

where $w(t)$ is the temporal weighting factor, and $\hat{r}(t)$ is the target rank for decoding at time t. The temporal weighting factor can be viewed as a more general form of time windows, where the latter only describes whether $w(t) > 0$.

In the initialization stage, $w(t)$ was set as a constant in the delay period for rank-WM subspaces, and in the sample period for shared sensory subspace, $\hat{r}(t)$ was set to the corresponding rank that we want to decode. In the refinement stage, their values were evaluated adaptively. Specifically, the optimization process operated through an iterative procedure: given $w(t)$ and $\hat{r}(t)$, the model parameters Θ were updated by minimizing the loss function through gradient descent; based on the derived model parameters Θ, the decoding rank $\hat{r}(t)$ and temporal weighting factor $w(t)$ were updated in the following way.

First, we calculated the mean decoding probability at time t for each rank:

$$p(r,t) = \frac{1}{n}\sum_{i}^{n} p_{i_r}(t,i)$$

where $p_{i_r}(t,i)$ is the i_r-th element in the softmax scores $\boldsymbol{p}(t,i)$. Then the target rank for decoding at moment t was defined as

$$\hat{r}(t) = \mathrm{argmax}_r\, p(r,t),$$

which can be interpreted as the most decodable rank at the moment t in the subspace. Based on the derived decoding rank $\hat{r}(t)$, the unnormalized temporal weighting factor was defined as

$$w_0(t) = p(\hat{r}(t),t) - p_0,$$

where p_0 is the baseline probability. Then, the temporal weighting factor $w(t)$ was obtained through normalization:

$$w(t) = \frac{w_0(t)}{\sum_{t}^{T} w_0(t)}.$$

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

The iterative procedure evolved until convergence, culminating in the final projecting weights for the low-dimensional subspaces, along with the final optimizing time windows and corresponding decoding ranks. During the training, we monitored whether the decoded content within the preset time window matched the expectation. If it does not match, $w(t)$ and $\widehat{r}(t)$ will be reset to their initial values. As an additional adaptive mechanism to ensure the stability of training, the baseline probability p_0 is initialized as chance level (1/6 for our task) and gradually increased to a maximum value proportional to the max decoding probability.

The task subspaces were sequentially identified one by one. After identifying one subspace, the final *w(t)* values were stored and used in estimating subsequent subspaces. Specifically, in identifying the *m*th subspace, we subtracted the projection of neural activity in previous subspaces in the period when $w(t) > 0$.

$$\boldsymbol{x}_{n+1}(t) = \boldsymbol{x}_n(t) - W_n W_n^T \boldsymbol{x}_n(t), \text{for } n = 1 \text{ to } m - 1$$

where $\boldsymbol{x}_1$ is the original de-meaned neural activity. Once an appropriate number of subspaces are identified, there is an additional final step to simultaneously optimize the subspaces while updating the *w(t)* of all the subspaces. Again, the projections of neural activity in other subspaces were subtracted when calculating the loss in one subspace.

To make a robust estimation, $w(t)$ and $\widehat{r}(t)$were estimated using pseudo-population data sampled from multiple sessions, following the aforementioned procedure. Then, they were fixed to train decoders for each session. Note that in this stage, the decoder for each subspace was trained independently on the same neural activity data without subtracting projections from other subspaces. These single-session decoders were used to calculate the single-trial decoding accuracy.

Generalized linear regression, subspace identification and decoding

We fit a Poisson GLM model to each neuron, which helped us derive regression coefficients for 18 variables (3 stimuli × 6 items) for each 50 ms time bin, without overlap. We then reduced the dimensionality by PCA to identify the subspaces. For each memory subspace, we averaged the regression coefficients corresponding to a stimulus over the late delay period (100 ms after the last stimulus offset to the go cue). For the sensory subspace, we selected coefficients related to specific stimuli during their presentation period (from stimulus onset to stimulus offset). The averaged regression coefficients were used to perform PCA. Then the neural activities were projected onto those identified subspaces and SVM decoding was performed. Two-fold cross-validation was used to get the decoding accuracy.

Geometric relationship between subspaces

To quantify the geometric relationship between the decomposed subspaces, we examined the alignment of two subspaces by computing the principal angle (PA) and variance accounted for (VAF) ratio.[43]

Given subspaces *a* and *b* as well as the associated two-dimensional bases W_a and W_b, to compute PA between them, we performed the singular value decomposition onto $W_b^T W_a$ such that

$$W_b^T W_a = P_b C P_a^T,$$

where P_a and P_b are 2×2 orthogonal matrices and C is a 2×2 diagonal matrix whose elements are the ranked cosines of the PAs θ_1 and θ_2:

$$C = diag(\cos(\theta_1), \cos(\theta_2)).$$

Note that the principal angles are ordered from smallest to largest, thus we only reported the first PA θ_1.

The VAF ratio for subspace pair (a, b) was defined as

$$VAF_{ab} = \frac{Var(W_a W_a^T G_b)}{Var(G_b)},$$

where $G_b = W_b W_b^T X$, and X ($N \times K$) is the activities of N neurons in K trials. As for the temporal aspect of neural activity, we calculated the variance for each time point and summed them all.

Neuronal contributions to subspaces in the single neuron level

To quantify the single neuronal contribution to each subspace, we projected the unit vectors along the axis of each single neuron onto the corresponding sensory or rank-WM subspaces and defined the square scalar projections as the neuronal contribution to the corresponding subspace.[23,44] For simplicity, denote the contribution of the *i*th neuron to subspace *a* as A_{ai}^2. The normalized participation ratio

$$NPR_a = \frac{1}{N} \frac{\left(\sum_{i=1}^{N} A_{ai}^2\right)^2}{\sum_{i=1}^{N} A_{ai}^4}$$

was used to quantify the proportion of significantly contributed neurons for subspace *a*.[23] For example, an NPR value of 0.4 indicated the top forty percent of neurons in the session significantly contributed to the subspace.

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

Neuronal contributions to subspaces in the recording channel level

Since the neural recordings for the forward and backward tasks were made separately, it is impossible to directly compare the neural substrates of subspaces across the two tasks. However, since recording channels were invariant across the two tasks, we made efforts to compare the neuronal contribution distribution at the recording channel level. For each recording channel from each pseudo-single day (by pooling several consecutive days), we calculated the proportion of significantly contributed neurons and the averaged neuronal contribution (see Figures 4A and S4A), and then multiplied these two quantities to obtain the proportion-weighted neuronal contribution in the recording channel level. Based on this metric, the cosine similarities across the shared channels between the forward and backward tasks for any subspace pair were calculated. Statistical analysis (two-sample t-test without assuming equal variances) between the intra- and inter-subspace pairs was made. For monkey O, intra vs inter for sensory: p=0, for memory-1: $p=1.01\times10^{-4}$, for memory-2: $p=8.98\times10^{-13}$, for memory-3: $p=2.41\times10^{-6}$; for monkey G, intra vs inter for sensory: $p=3.22\times10^{-7}$, for memory-1: p=0, for memory-2: $p=3.33\times10^{-16}$, for memory-3: p=0.

Forward and backward trials with rule cues

To test whether the rank-WM subspaces were shared across the forward and backward tasks, two monkeys (monkey O and monkey L) were further trained to perform the two tasks trial-by-trial. In the middle of the delay period (500∼700 ms after the second stimulus), fruit imagery (cucumber for the forward task and apple for the backward task) was flashed for 250 ms to signal the trial type. The "go" cue then appeared after a random delay (500∼700 ms), as illustrated in Figure 4C. The remaining task parameters and the recording specifics were the same as the block trials. Note that, in this task, our analysis focused exclusively on neural activities during the sample period and the second delay.

Error trial analysis

We applied the single-session decoders trained with correct-trial data to error trials. As shown in Figures 1C and 1D, there were two main error types, including ordinal and item errors. Together with the fact that monkeys made few errors in length-1 and -2 trials, we mainly focused on these two error types for length-3 trials. Then, for each error type, the decoding accuracies shown in Figures 5 and S5 were obtained by averaging across trials from all sessions. For the ordinal error from monkey O, we mainly showed the swap error between rank-2 and -3 due to its predominancy in trials (Figure 5B). For the item error from monkey O, aside from the overall response-3 error result due to its predominancy in trials, we further divided response-3 errors into different subtypes according to the distance between the error response and the right target, trying to gain a better understanding of the control process underlying the item error (Figures 5C and 5D). We also included the error profile for other conditions and for monkey G in Figure S5.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical test of decoding performance in each subspace

To determine the statistical significance of single trial decoding performance in each subspace, we used an extreme pixel-based permutation test to report the p-values.[45] We trained surrogate decoders using the same $w(t)$ values but with labels shuffled across trials. Note that the labels are shuffled at the sequence level rather than the item level. For decoding results in each subspace, we only used the maximal accuracy value across all the time points. This ensures that the multiple comparisons are taken into account in statistical testing. Two-fold cross-validation was used to get the decoding accuracy. We repeated this process 1000 times to obtain a null distribution. The 99.9 (or 99) percentile of this distribution was used to threshold the decoding accuracies in testing conditions that correspond to $p<0.001$ ($p<0.01$).

Latency difference for sensory-memory subspace pairs

To assess whether item information sequentially enters into the sensory and rank-WM subspace in the forward task, we performed latency difference significance analysis for each sensory-memory subspace pair. To generate a distribution of sensory-memory latency difference, we used the bootstrap procedure with replacement 1,000 times. For each time, we employed the aforementioned subspace decomposition method to identify both sensory and rank-WM subspaces. Then, for each subspace, we chose the first time point during which the decoding accuracy exceeds 1/3 and then used this time point to calculate the latency difference between sensory and rank-WM subspaces. Wilcoxon signed-rank tests were used to test whether latency differences were significant.

Correlation of decoding probabilities for sensory-memory subspace pairs

To further quantify the correlation of information contents for each sensory-memory subspace pair in the forward task, we computed the single trial correlation of decoding probabilities for each sensory-memory subspace pair. For each subspace, the decoding probability of a rank (r) in a single trial (n) was acquired by averaging the decoding probabilities during the period (T) in which the corresponding rank information was significantly encoded:

$$p(r,n) = \frac{1}{T}\sum_{t}^{T} p(r,t)$$

Please cite this article in press as: Chen et al., Flexible control of sequence working memory in the macaque frontal cortex, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.07.024

where $\boldsymbol{p}(r,t)$ is the decoding probability of rank r at time t. We used Spearman correlation to calculate the correlation coefficients between the decoding probability in the sensory subspace and the memory subspaces for each session:

$$\rho = \frac{\Sigma_i(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\Sigma_i(x_i - \overline{x})^2 \Sigma_i(y_i - \overline{y})^2}}$$

where x is the rank of sensory subspace decoding probability and y is the rank of rank-WM subspace decoding probability. For example, x is the rank of the decoding probability of S1 in sensory, and y is the decoding probability of S1 in memory-1 (paired); x is the rank of the decoding probability of S1 in sensory, and y is the rank of the decoding probability of S2 in memory-2 (unpaired). The correlation coefficients from all days were then put together, and the Wilcoxon signed-rank test was used to determine (1) whether the paired coefficients were significantly larger than 0 and (2) whether they were different from each other (Bonferroni corrected).

Statistical test of principal angles and VAF ratios

To make the statistical tests for principal angles and VAF ratios, we first computed the PA and VAF ratio between two different estimations (from the two-fold cross-validation) for the same subspace as a control. Then we made the statistical tests between the control subspace pairs (e.g., Sensory-Sensory, M1-M1) and all the subspace pairs involved the control subspace pairs (e.g., M1-Sensory, M2-Sensory, M3-Sensory for Sensory-Sensory). We defined the difference as significant when the 0.05^{th} percentile of the distributions across sessions was not overlapped. Bonferroni corrections were applied for the multiple comparisons.