

Shared structure facilitates working memory of multiple sequences

Qiaoli Huang , Huan Luo 

School of Psychological and Cognitive Sciences, Peking University • PKU-IDG/McGovern Institute for Brain Research, Peking University • Beijing Key Laboratory of Behavior and Mental Health, Peking University • Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Reviewed Preprint

v2 • June 19, 2024

Revised by authors

Reviewed Preprint

v1 • December 15, 2023

Abstract

Daily experiences often involve the processing of multiple sequences, yet storing them challenges the limited capacity of working memory (WM). To achieve efficient memory storage, relational structures shared by sequences would be leveraged to reorganize and compress information. Here, participants memorized a sequence of items with different colors and spatial locations and later reproduced the full color and location sequences one after another. Crucially, we manipulated the consistency between location and color sequence trajectories. First, sequences with consistent trajectories demonstrate improved memory performance and a trajectory correlation between reproduced color and location sequences. Second, sequences with consistent trajectories show neural reactivation of common trajectories, and display spontaneous replay of color sequences when recalling locations. Finally, neural reactivation correlates with WM behavior. Our findings suggest that shared common structure is leveraged for storage of multiple sequences through compressed encoding and neural replay, together facilitating efficient information organization in WM.

eLife assessment

This **valuable** study uses a novel experimental design to elegantly demonstrate how we exploit stimulus structure to overcome working memory capacity limits. The presented behavioural and neural evidence are **solid** and in line with the proposed information compression mechanism. This study will be of interest to cognitive neuroscientists studying structure learning and memory.

<https://doi.org/10.7554/eLife.93158.2.sa1>

Introduction

A well-known feature of working memory (WM) is its limited capacity (Baddeley, 2000 ; Cowan, 2001 ) , which constrains the amount of information that can be temporarily retained for future behavior. Meanwhile, in daily experiences, memorized items do not exist independently but are always part of a common framework or share the same structure, which could be leveraged to compress information and overcome the WM capacity challenge (Brady et al., 2011 ). For

example, while shopping at a supermarket, numerous items could be grouped into a few categories, such as drinks, vegetables, fruits, and meats, to facilitate memory. Other types of abstract associations such as relational regularities and structure schema could also mediate WM organization (Al Roumi et al., 2021 [↗](#); Gathercole & Baddeley, 2014 [↗](#); Mathy & Feldman, 2012 [↗](#)). Computational modeling also suggests that higher-order structures (e.g., summaries and relative relations) would reduce memory uncertainty by constraining individual-item representations (Brady & Tenenbaum, 2013 [↗](#); Ding et al., 2017 [↗](#)).

Cognitive maps, as one type of spatial schema (Farzanfar et al., 2023 [↗](#); Gilboa & Marlatte, 2017 [↗](#)), provide a general structure framework for organizing information in different tasks and across various domains (Tolman, 1948 [↗](#); Whittington et al., 2020 [↗](#)). They were first identified as representations of physical maps during navigation, but recently have been shown to also support other higher-level processes, such as conceptual knowledge, reasoning, planning, and decision-making (Behrens et al., 2018 [↗](#); Bellmund et al., 2018 [↗](#); O'keefe & Nadel, 1978 [↗](#)). Two major neural signatures of cognitive maps, grid-like code (Constantinescu et al., 2016 [↗](#); Doeller et al., 2010 [↗](#); Hafting et al., 2005 [↗](#); Park et al., 2021 [↗](#)) and neural replay in the hippocampal-entorhinal system (D. J. Foster & Wilson, 2006 [↗](#); Liu et al., 2019 [↗](#); Liu, Mattar, et al., 2021 [↗](#); Schuck & Niv, 2019 [↗](#); Skaggs & McNaughton, 1996 [↗](#)), are identified in both spatial and non-spatial tasks. Neural replay, the rapid item-by-item reactivation in a forward or backward direction, is posited to not only repeat past experience, but also reflect an internal model of the world (Kurth-Nelson et al., 2023 [↗](#); Ólafsdóttir et al., 2018 [↗](#)).

Accordingly, many higher-level processes could be described as mental explorations of a sequence of states within an abstract map, similar to tracing a route on a physical map. As a result, cognitive map can serve as a common reference for aligning different features or domains. In line with the view, conjoined cognitive maps have been recently revealed in the rodent hippocampus (i.e., physical space and abstract task variables) (Nieh et al., 2021 [↗](#)), and alignment of different feature maps speeds learning performance (Aho et al., 2022 [↗](#)). Moreover, both spatial and conceptual distances are relied on to generalize when searching for correlated rewards in value-guided learning, supporting a cognitive-map dependent computational mechanism (Wu et al., 2020 [↗](#)). Based on these findings, we hypothesize that cognitive maps might be employed to reorganize memory information across domains to overcome capacity bottlenecks in WM.

Here we examined whether cognitive maps shared by multiple feature domains would be naturally leveraged to make efficient use of limited WM capacity. To address this question, participants were asked to memorize a color sequence presented at a list of spatial locations and later reproduce both the color and location sequences on two rings, one after another. In other words, subjects need to retain in WM two sequences of features, i.e., color and location, both of which could be characterized as sequence trajectories along their respective rings on a spatial map. Crucially, we manipulate the consistency between color and location sequence trajectories on the rings. Specifically, for the aligned condition, the color and location sequence share a common spatial trajectory, i.e., separated by the same distance between successive items between maps (**Figure 1B** [↗](#)), whereas for the misaligned condition, they have distinct relative trajectories (**Figure 1C** [↗](#)). We hypothesize that humans would spontaneously combine the structure shared by the two sequences to facilitate memory formation, even though it is unsupervised and non-mandatory.

To preview the results, we provide converging behavioral and neural evidence for spontaneously leveraging common structures to facilitate WM. Behaviorally, sequences with consistent color-location trajectories (aligned condition) show enhanced memory precision and a significant correlation between reproduced color and location sequence trajectories. Neurally, aligned location-color sequences demonstrate reactivation of shared trajectory during both encoding and

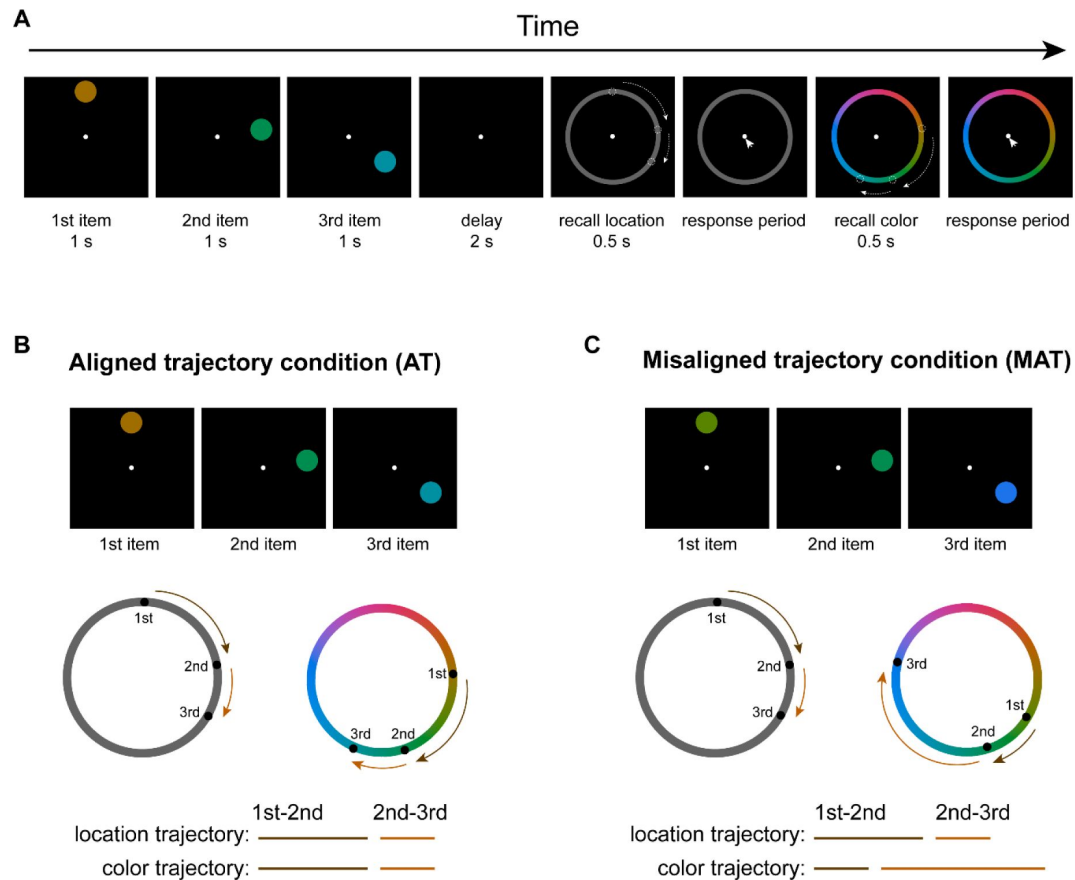


Figure 1.

Experimental paradigm

(A) Participants were presented with a sequence of disks of different colors and at different locations. They were asked to memorize both the location and color of the sequence and later reproduce the full location and color sequences one after another by clicking the corresponding positions on the respective report rings. During the “recall location” phase, a grey ring appeared and participants prepared for subsequent location recall without motor movement, to ensure memory decoding without motor interventions. During the following “response period”, subjects serially selected memorized spatial locations on a “location ring”. Next, a color ring appeared (“recall color”) for subjects to be ready for subsequent color recall. They then clicked the remembered colors on a “color ring” (“response period”). **(B)** Aligned trajectory condition (AT) wherein the trajectory distances between consecutive items (i.e., 1st to 2nd, 2nd to 3rd) of location and color sequences are identical, although the two sequences occupy different locations within their respective rings. **(C)** Misaligned condition (MAT), wherein the trajectory distances between consecutive items are different for location and color sequences.

retention periods and interestingly, spontaneous replay of color sequences when recalling location sequences. Together, shared common structures enable the storage of multiple sequences in WM through compressed encoding and neural replay.

Results

Experimental procedure and behavior performance

Thirty-three human participants performed a visual sequence WM task while their brain activities were recorded using EEG. As shown in **Figure 1A**, at the beginning of each trial, three disks with different spatial locations and colors were sequentially presented. Participants were required to concurrently remember their locations and colors as well as their orders, i.e., one location sequence and one color sequence. After a 2-second memory delay, a grey ring (location ring) was presented to instruct participants to prepare for subsequent location sequence recall without making motor responses (**Figure 1A**, “recall location” period). This is to ensure memory signals are decoded without explicit motor interventions. Next, a cursor appeared at the center of the screen (“response period”), and participants clicked the three spatial locations on a “location ring” in their correct order. Upon completion of location recall, participants were instructed to prepare for color sequence recall (“recall color”), and they clicked three locations on the color ring based on the color sequence (“response period”). One key aspect was manipulating the consistency between location and color trajectories so that the two sequences share or do not share a common structure in a cognitive map. Specifically, in the aligned trajectory condition (AT condition), despite the location and color sequences occupying different positions within their respective rings, their trajectory distances (between 1st and 2nd items and between 2nd and 3rd) were the same (**Figure 1B**). In other words, by rotating certain angles, the three points in the two rings can exactly match, and the rotated angles varied from trial to trial, which allowed us to separately decode location sequence and color sequence in the following analysis. In contrast, the location and color sequences in the misaligned trajectory condition (MAT condition) differed both in positions and trajectory distances within rings (**Figure 1C**).

Aligned color-location trajectory improves memory performance

We first estimated memory precision for color and location sequences by calculating the reciprocal of circular standard deviation of response error (circular difference between reported location (color) and correct location (color)) across trials ($1/\sigma$) (Bays et al., 2009). As shown in **Figure 2A**, a 2-way repeated ANOVA (alignment (AT vs. MAT) $\times$ task (location vs. color)) revealed significant main effects for alignment ($F_{(1,32)} = 4.279$, $p = 0.047$, $\eta^2 = 0.118$) and task ($F_{(1,32)} = 139.382$, $p < 0.001$, $\eta^2 = 0.813$), but nonsignificant interaction effect ($F = 0.618$, $p = 0.438$, $\eta_p^2 = 0.019$). Specifically, AT condition had better memory performance than MAT condition, supporting our hypothesis that shared structure facilitates memory of multiple sequences. Moreover, location memory performed better than color memory. Further comparison revealed that the aligned condition mainly enhanced color memory (paired-t test, $t_{(32)} = 2.446$, $p = 0.020$, Cohen's $d = 0.426$) but not location (paired-t test, $t_{(32)} = 1.538$, $p = 0.134$, Cohen's $d = 0.268$). Better location vs. color memory performance indicates that alignment operation is less effective in improving memory (i.e., location sequence) that is already very robust (Wu et al., 2020). In terms of serial position in sequence, color sequences demonstrated better memory performance under AT versus MAT conditions, especially for the 2nd and 3rd items (paired-t test, 1st: $t_{(32)} = -0.315$, $p = 0.755$, Cohen's $d = 0.055$; 2nd: $t_{(32)} = 4.069$, $p < 0.001$, Cohen's $d = 0.709$; 3rd: $t_{(32)} = 2.583$, $p = 0.015$, Cohen's $d = 0.450$) (**Figure 2B**). Meanwhile, the location sequences exhibited similar performance for all positions (paired-t test, 1st: $t_{(32)} = 0.972$, $p = 0.338$, Cohen's $d = 0.169$; 2nd: $t_{(32)} = 1.245$, $p = 0.222$, Cohen's $d = 0.216$; 3rd: $t_{(32)} = 1.290$, $p = 0.206$, Cohen's $d = 0.225$) (**Figure 2C**). Overall, behavioral findings demonstrate an improvement in WM performance with a common trajectory across feature domains, and indicate that the aligned trajectories (1st-2nd, 2nd-3rd) may be applied to reduce the memory uncertainty for the 2nd and 3rd colors.

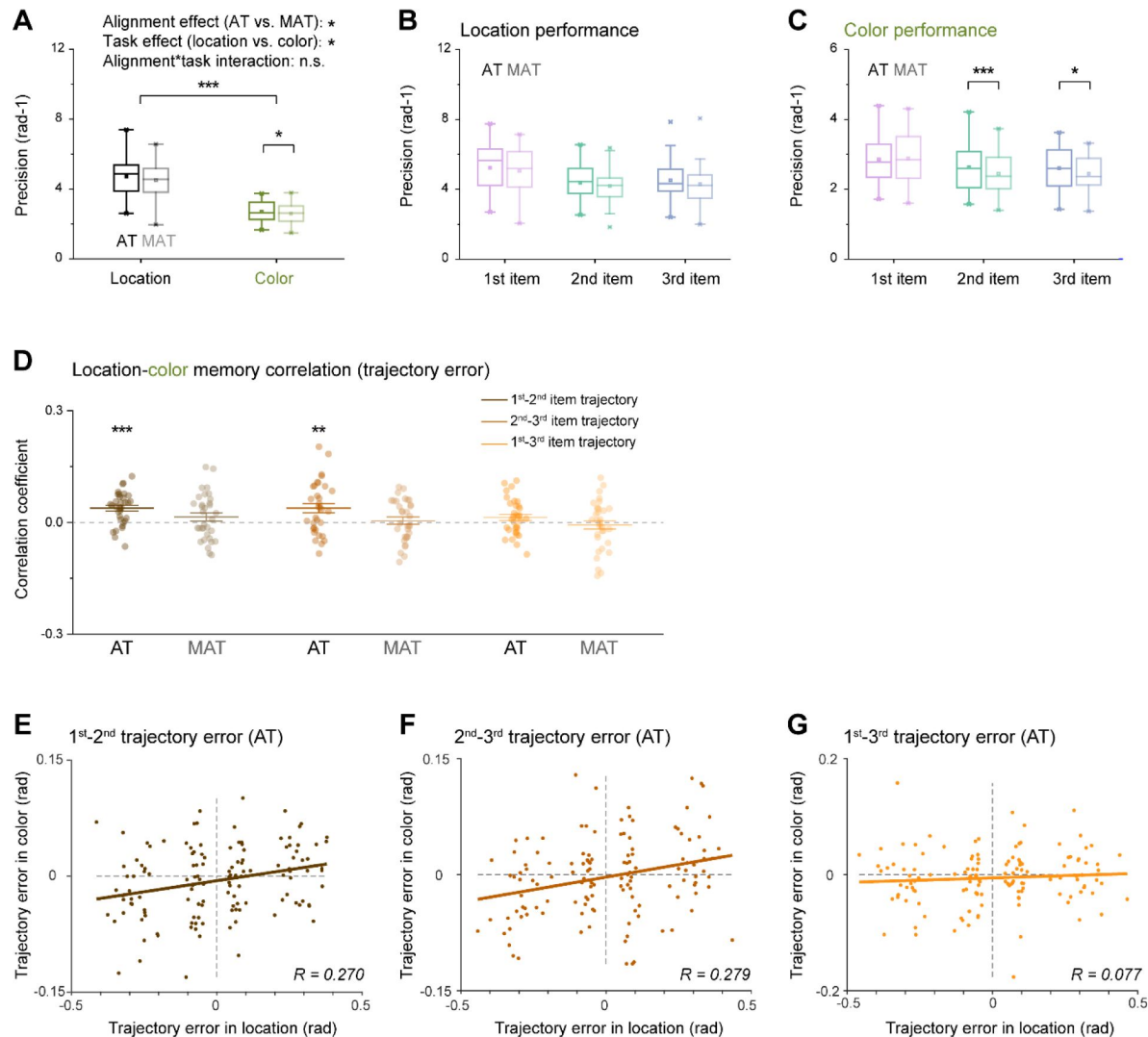


Figure 2.

Behavioral performance

(A) Memory precision performance of location (black) and color (green) sequences for AT (dark color) and MAT (light color) conditions. Horizontal line in the boxplots denotes the median; box outlines denote the 25th and 75th percentiles; whiskers denote $1.5 \times$ the interquartile range. Extreme values are denoted by crosses. (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$). **(B)** Memory precision of 1st (purple), 2nd (turquoise) and 3rd (blue) items of location sequence, for AT (dark color) and MAT (light color) conditions. **(C)** Memory precision of 1st (purple), 2nd (turquoise) and 3rd (blue) items of color sequence, for AT (dark color) and MAT (light color) conditions. **(D)** Grand average (mean $\pm$ SEM) correlation coefficients of recalled trajectory error between location and color sequences, for 1st-to-2nd trajectory (brown), 2nd-to-3rd trajectory (brickred), and 1st-to-3rd trajectory (orange), under AT (dark color) and MAT (light color) conditions. Dots indicate individual participant. **(E)** Scatterplot of 1st-to-2nd trajectory memory error for location sequence (X-axis) and Color sequence (Y-axis) under AT condition. Note that the trajectory error of all trials within each subject was divided into 4 bins according to the location trajectory error, resulting in 33 (subject number) $\times$ 4 (bins) dots in the plot. The brown line represents the best linear fit. **(F)** Same as E, but for 2nd-to-3rd trajectory. **(G)** Same as E, but for 1st-to-3rd trajectory.

Aligned color-location trajectory elicits color-location correlation in recalled trajectories

We further investigated whether the location-color trajectory alignment was truly leveraged in the memory process. Note that participants reproduced the color and location sequences by clicking three positions on the respective rings, i.e., reproducing two spatial trajectories, one for location and one for color (**Figure 1A** [↗](#), reporting period). We therefore could examine the correlation between the reported location and color trajectories in their maps to determine whether the AT condition would result in a correlated pattern based on the reported sequences.

Specifically, we first calculated trajectory error (the circular difference between the reported trajectory and the true trajectory) for location and color features, and then assessed the correlation between the (signed) trajectory error of location and color features, for each subject. As shown in **Figure 2D** [↗](#), AT condition showed significant correlations for both 1st-2nd (one-sample t-test, $t_{(32)} = 5.022$, $p < 0.001$) and 2nd-3rd (one-sample t-test, $t_{(32)} = 3.113$, $p = 0.004$) trajectories, but not for 1st-3rd trajectory (one-sample t-test, $t_{(32)} = 1.579$, $p = 0.124$). In contrast, the MAT condition did not display any significant correlation (one-sample t-test; 1st-2nd: $t_{(32)} = 1.361$, $p = 0.183$; 2nd-3rd: $t_{(32)} = 0.490$, $p = 0.628$; 1st-3rd: $t_{(32)} = -0.582$, $p = 0.565$).

At the group level, motivated by a previous study (H.-H. Li et al., 2021 [↗](#)), we quantified the trajectory correlations by first binning all trials based on the location trajectory error, then extracting the color trajectory error for each bin, for each subject and pooling the data across subjects. As shown in **Figure 2E** [↗](#), a significant correlation was observed for the 1st-2nd ($r = 0.270$, $p = 0.002$) and 2nd-3rd trajectory ($r = 0.279$, $p = 0.002$), but not for the 1st-3rd trajectory ($r = 0.077$, $p = 0.375$). Similarly, the MAT condition did not exhibit any location-color correlation in trajectories (1st-2nd: $r = 0.097$, $p = 0.277$; 2nd-3rd: $r = 0.025$, $p = 0.790$; 1st-3rd: $r = -0.065$, $p = 0.443$; also see Supplementary Figure 1ABC), which excludes the possibility that the reported trajectory correlation was solely due to systematic response bias. We further confirmed the trajectory errors correlations using a generalized linear mixed-effects model (AT: 1st-2nd trajectory, $\beta = 0.071$, $t = 4.215$, $p < 0.001$; 2nd-3rd trajectory, $\beta = 0.077$, $t = 3.570$, $p < 0.001$; 1st-3rd trajectory, $\beta = 0.019$, $t = 1.118$, $p = 0.264$; MAT: 1st-2nd trajectory, $\beta = 0.031$, $t = 1.572$, $p = 0.116$; 2nd-3rd trajectory, $\beta = -0.002$, $t = 0.128$, $p = 0.898$; 1st-3rd trajectory, $\beta = -0.017$, $t = -1.024$, $p = 0.306$).

Together, behavioral findings indicate that memory facilitation arises from an automatic alignment of recalled trajectories across feature domains to compress information. In other words, instead of memorizing two 3-item sequences independently, subjects may just maintain their common trajectories.

Neural decoding of location and color features during encoding

We employed a time-resolved inverted encoding model (IEM) (Brouwer & Heeger, 2009 [↗](#), 2011 [↗](#)) on EEG signals to examine the neural representation of location and color. Specifically, the slope of the reconstructed channel response was estimated to quantify the time-resolved decoding performance for the 1st, 2nd, and 3rd location and color, respectively (see details in Materials and methods) at each time point. We first focused on the encoding period when the 3-disk sequence was physically presented.

As shown in **Figure 3A** [↗](#), the location of each of the 3 disks could be successfully decoded from EEG signals for both AT (1st location: 0.03–0.56 s, 1.14–1.26 s, 1.55–1.84 s; 2nd location: 1.56–2.10 s, 2.65–2.92 s, 3.12–3.35 s; 3rd location: 3.06–3.50 s; corrected cluster $p < 0.001$) and MAT conditions (1st location: 0.03–0.52 s, 1.15–1.39 s, 1.58–1.84 s; 2nd location: 1.52–2.10 s, 3.11–3.36 s; 3rd location: 3.06–3.54 s; corrected cluster $p < 0.001$) during stimulus presentation period. Similarly, color information could also be decoded for both AT (1st color: 0.09–0.50 s, corrected cluster $p < 0.001$; 2nd

color: 1.57–2.01 s, corrected cluster $p < 0.001$; 3rd color: 3.11–3.49 s, corrected cluster $p = 0.002$) and MAT conditions (1st color: 0.04–0.45 s; 2nd color: 1.57–1.91 s, corrected cluster $p < 0.001$; 3rd color: 3.11–3.55 s; corrected cluster $p < 0.001$) (**Figure 3B**). It is noteworthy that location and color features were generated with the constraint that they could not occupy the same position within their respective rings. This thereby ensured the independent decoding of location and color features from the same neural signals. Moreover, the color feature exhibited weaker decoding strength than location, also consistent with behavioral results (**Figure 2A**).

Spontaneous replay of color sequence during location recall

After confirming location and color representations during encoding period, we next examined the neuronal correlates of sequence memory during retrieval. We are particularly interested in the “recall location” period, during which subjects need to remain still without making motor responses but at the same time prepare for subsequent location recall (see **Figure 1A**, and upper panel of **Figure 4**). During this period, subjects need to maintain two sequences: the location sequence which is immediately task-relevant, and the color sequence which is not task-relevant right now but will be recalled later. That is to say, we posit that memories should be converted to sequences as outputs during this period. Behavioral analysis indicates the correlation between recalled location and color trajectories for AT condition (**Figure 2**), which suggests an active combination of common trajectories across features. As a result, we sought neural evidence for the reintegration between the color sequence and the location sequence for AT condition during this period.

As shown in **Figure 4A** (left panel), the currently task-relevant location sequence during AT condition displayed strong decoding performance for the 1st location (0.11–0.46 s, corrected cluster $p = 0.003$), weak but significant decoding performance for the 3rd location (0.27–0.41 s, corrected cluster $p = 0.011$), but not for the 2nd location. Moreover, The MAT condition (**Figure 4A**, right panel) showed the similar location decoding profiles (1st location: 0.11–0.46 s, corrected cluster $p < 0.001$; 3rd location: 0.13–0.35 s, corrected cluster $p = 0.002$). The primacy effect might be due to the fact that the 1st location is the first to be recalled afterwards, and therefore it denotes either the most task-relevant feature or motor preparation. In fact, similar position effect has also been observed for color sequence during the color recalling period (Supplementary Figure 2).

Most importantly, we asked whether the “recall location” period also contains color sequence information, which is not task-relevant at the moment but will be recalled later. As shown in the left panel of **Figure 4B**, we observed significant reactivation of color sequence for AT condition. Specifically, the color sequence undergoes a temporally compressed, forward replay (1st color: 0.10–0.16 s, corrected cluster $p = 0.048$; 2nd color: 0.21–0.27 s, corrected cluster $p = 0.046$; 3rd color: 0.31–0.38 s, corrected cluster $p = 0.089$). In contrast, the MAT condition did not display any neural reactivation of the color sequence (**Figure 4B**, right panel). Direct comparison between AT and MAT conditions showed a significant reactivation difference (2-way repeated ANOVA, $F_{(1,32)} = 14.213$, $p = 0.001$, $\eta_p^2 = 0.308$). Further analysis (**Figure 4C**) reveals that the AT-MAT difference is mainly due to the 1st (paired test, $t_{(32)} = 3.151$, $p = 0.004$, Cohen’s $d = 0.548$) and 2nd items ($t_{(32)} = 1.914$, $p = 0.065$, Cohen’s $d = 0.334$).

We next used a “sequenceness” approach (Kurth-Nelson et al., 2016; Liu et al., 2019) to characterize the sequential replay profile. Specifically, we calculated the cross-correlation coefficients between consecutive items, and then performed a permutation test to examine the statistical significance of the sequential replay profile by shuffling color labels across participants. **Figure 4D** shows a significant temporal lag around 110 ms and 140 ms for the 1st-2nd and 2nd-3rd color pairs, respectively, indicating a forward replay profile with temporal compression within approximately 130 ms (peak of the average of the two cross-correlation time courses).

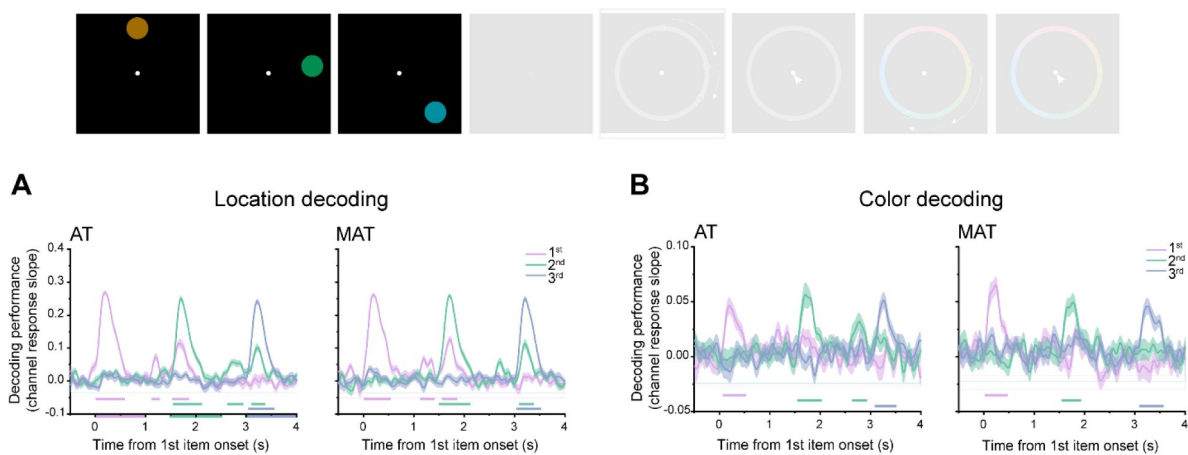


Figure 3.

Neural representation of memory contents during encoding period

(A) Grand average (mean $\pm$ SEM) neural decoding (slope of channel response) of location information for the 1st (purple), 2nd (turquoise) and 3rd (blue) disk as a function of time during the encoding period, for AT (left panel) and MAT conditions (right panel). Horizontal lines with corresponding colors denote significant time ranges (cluster-based permutation test, cluster-defining threshold $p < 0.001$, corrected significance level $p < 0.001$) **(B)** Same as A, but for color feature decoding.

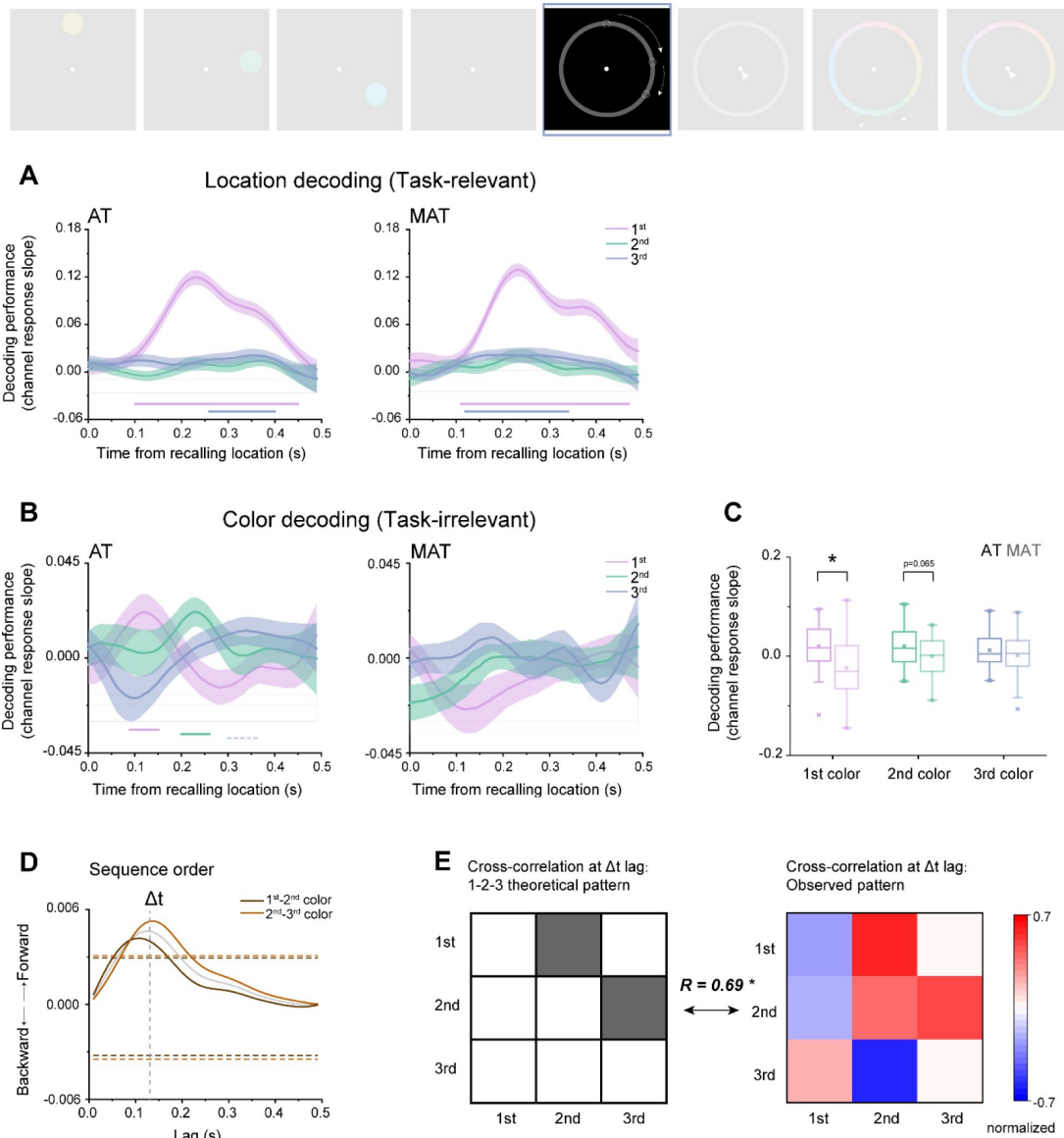


Figure 4.

Spontaneous color sequence replay during "location recall"

(A) Grand average (mean $\pm$ SEM) decoding performance for 1st (purple), 2nd (turquoise) and 3rd (blue) locations as a function of time during "recall location" period, for AT (left panel) and MAT conditions (right panel). **(B)** Grand average (mean $\pm$ SEM) decoding performance for 1st (purple), 2nd (turquoise) and 3rd (blue) colors as a function of time during "recall location" period, for AT (left panel) and MAT conditions (right panel). (Horizontal solid line: cluster-based permutation test, cluster-defining threshold $p < 0.05$, corrected significance level $p < 0.05$; Horizontal dashed line: marginal significance, cluster-defining threshold $p < 0.1$, $0.05 < \text{cluster } p < 0.1$) **(C)** Grand average decoding performance within the respective significant time range, for 1st (purple), 2nd (turquoise) and 3rd (blue) colors, under AT (dark color) and MAT (light color) conditions. **(D)** Cross-correlation coefficient, calculated to quantify the extent of the neural representations of adjacent two items followed a forward (positive y) or backward (negative y) transition as a function of time lag, between 1st and 2nd colors (brown color) and between 2nd and 3rd (brickred color) colors, and their average (grey color). Dashed vertical line denotes the peak of the averaged cross-correlation time courses. Dashed horizontal lines denote the nonparametric statistical significance threshold ($p < 0.05$, permutation test). **(E)** Left panel: theoretical transition pattern for 3-item forward replay, i.e., 1st-2nd-3rd, characterized by cross-correlation at certain time lag. Right panel: empirical transitional pattern (actual cross-correlation matrix) at 130 ms time lag. A significant correlation was found between the two matrices ($r = 0.690$, $p = 0.040$), further confirming the forward replay of color sequence.

Moreover, motivated by previous studies (Liu, Dolan, et al., 2021 [↗](#); Liu, Mattar, et al., 2021 [↗](#)), we first constructed the theoretical transitional pattern for the 3-item sequence by assuming a Δt temporal lag between consecutive items (i.e., cross-correlation matrix). As shown in **Figure 4E** [↗](#) (left panel), the 1st item at time T could predict the reactivation of 2nd item at T+ Δt , and the 2nd item at time T could predict the appearance of 3rd item at T+ Δt . We then calculated the actual cross-correlation matrix (empirical transitional pattern) at 130 ms time lag which denotes the time lag of consecutive items in our findings (see **Figure 4D** [↗](#), grey line), resulting in a 3 × 3 matrix (**Figure 4E** [↗](#), right panel). A significant correlation was found between the empirical cross-correlation matrix and the theoretical transitional pattern ($r = 0.690$, $p = 0.040$), further confirming the forward replay of color sequence.

Together, when subjects prepare to reproduce location sequence during “recall location”, the currently task-irrelevant color sequence demonstrates a spontaneous sequential replay profile, which highlights the close bound between color and location sequences when they share common trajectory. Together with the color-location trajectory correlation in behavior, the findings suggest that replay-based neural mechanisms in WM mediate sequence combinations based on common structures.

Neural representation of common trajectory and its behavioral correlates

Finally, we accessed the neural representations of the common trajectory structure during encoding and retention periods. Specifically, a linear support vector machine (SVM) was employed to decode the 1st-to-2nd and 2nd-to-3rd trajectory distance, with the 1st-to-3rd trajectory as a control. Since there are only 8 possible circular distances between every two locations on a ring, the chance level of decoding performance is 0.125.

During the encoding period (**Figure 5A** [↗](#)), the 1st-to-2nd trajectory appeared right after the presentation of the 2nd item for AT condition (1.55–2.11 s, corrected cluster $p < 0.001$), which is expected since the relationship between the 1st and 2nd items along the trajectory can only be established when the 2nd item occurs. Similarly, the 2nd-to-3rd trajectory appeared after onset of the 3rd item (3.10–3.48 s, corrected cluster $p < 0.001$) (**Figure 5B** [↗](#)). In contrast, the 1st-3rd trajectory showed nonsignificant neural decoding (**Figure 5C** [↗](#)), consistent with the nonsignificant color-location trajectory correlation in behavior. Most interestingly, the 1st-to-2nd trajectory was reactivated right after the 3rd item (3.19–3.48 s, corrected cluster $p = 0.006$; **Figure 5A** [↗](#)), implicating the formed link between the two trajectories (1st-2nd and 2nd-3rd) for aligned location-color sequences. Therefore, when the color and location sequences share the same trajectory (i.e., AT condition), brain activities tend to co-represent the previously formed 1st-to-2nd trajectory (reactivation) and the newly formed 2nd-to-3rd trajectory during encoding to establish the full trajectory (see Supplementary Figure 1DEF for MAT condition). The trajectory reactivation was also related to WM behavior. Specifically, we divided all participants into two groups based on their 1st-to-2nd color-location trajectory memory correlations in behavior (**Figure 2D** [↗](#)) and calculated their corresponding neural representation of the 1st-to-2nd trajectory, respectively. Both the higher- and lower-correlation groups displayed significant neural decoding of the 1st-to-2nd trajectory right after the 2nd item (**Figure 5D** [↗](#); higher group: 1.56–1.99 s, corrected cluster $p < 0.001$; lower group: 1.64–1.97 s, corrected cluster $p < 0.001$). Meanwhile, only the higher-correlation group exhibited a significant reactivation of the 1st-to-2nd trajectory after the onset of the 3rd item (3.25–3.46 s, corrected cluster $p = 0.018$, 3.55–3.68 s, corrected cluster $p = 0.088$; Higher group vs. Lower group: bootstrap test, $p = 0.037$, two-side). Moreover, we were curious about the correlation between the common trajectory representation and latter forward replay pattern. Therefore, based on the average of trajectory neural representation (1st-2nd and 2nd-3rd trajectories) during presentation of the 3rd item when common trajectory reactivation was observed, we divided all participants into two groups, and observed higher trajectory

representation group was accompanied with clearer forward replay (see Supplementary Figure 3). This finding indicates that memory reorganization formed during encoding can predict serial replay during recalling.

With regards to the retention period, based on previous literatures (de Vries et al., 2020 [DOI](#); J. J. Foster et al., 2016 [DOI](#); Fukuda et al., 2016 [DOI](#); Sutterer et al., 2019 [DOI](#)), we chose to rely on the alpha-band activities to decode shared trajectories (see decoding results based on raw signals in Supplementary Figure 4AB). The aligned condition showed significant and long-lasting decoding of compression trajectories (Supplementary Figure 4CD), while the misaligned condition only showed decoding at the beginning, which might be due to the non-offset response of the 3rd item (Supplementary Figure 4GH). The results, although not as clear as those during encoding and recalling periods, also supports compression trajectory is spontaneously leveraged to reorganize multiple information.

Discussion

Identifying the underlying structure to facilitate efficient information storage in WM is crucial to human intelligence. Here, we investigated whether common structures shared across different feature domains would be spontaneously employed to facilitate memory of multiple sequences and its neural correlates. Since both location and color features could be characterized by positions along a continuous ring, we systematically manipulated the trajectory consistency between the location and color sequences. We show that color-location trajectory alignment is associated with better memory performance than the misaligned condition, and the memory benefit is attributed to structure-induced constraints on individual items that decrease representational uncertainty. EEG recordings provide neural evidence for the employed “compressive” strategy, i.e., reactivation of shared trajectories during encoding and retention, and spontaneous replay of color sequences during location recall. Finally, structure reactivation is related to WM behavior and neural replay. Together, shared common structure is leveraged for storage of multiple sequences through compressed encoding and neural replay, together facilitating efficient information organization in WM.

Events in daily experiences are not isolated but are always linked to each other. Therefore, instead of treating individual events as independent information, a more efficient way is to seek the link between seemingly unrelated events, i.e., “connecting the dots” in the WM system. Here the trajectories for both location and color sequences are defined in a ring coordinate system, providing an abstract-level cognitive map for memory formation. Our results demonstrate that subjects spontaneously realign sequence trajectories across features to facilitate memory of two sequences. In other words, instead of memorizing two 3-item sequences, subjects could just maintain two starting points and a common trajectory, an apparently more efficient way. Not that the trajectory alignment manipulation differs from the Gestalt principles of perceptual organization such as proximity and similarity principles (Goldstone & Medin, 1994 [DOI](#)), and it instead reflects a higher-order relationship between maps. The alignment manipulation also could not be accounted for by associative memory (Aho et al., 2022 [DOI](#); Roads & Love, 2020 [DOI](#)), since the rotational orientation for alignments between the two maps differed on a trial-by-trial basis. Finally, our study is also different from recent works on structure learning and generalization (Dekker et al., 2022 [DOI](#); Garvert et al., 2017 [DOI](#); Liu, Mattar, et al., 2021 [DOI](#); Ren et al., 2022 [DOI](#); Schapiro et al., 2013 [DOI](#)), as our task does not involve pre-exposure training or task-related rewards.

Notably, our findings could also be understood in terms of schematic abstraction, which plays pivotal roles in memory formation (Gilboa & Marlatte, 2017 [DOI](#); Tse et al., 2007 [DOI](#)), such that a congruent schema would facilitate memory retrieval compared to an incongruent condition (Audrain & McAndrews, 2022 [DOI](#); Van Kesteren et al., 2010 [DOI](#)). In fact, cognitive map refers to the internal representation of spatial relations in a specific environment (Tolman, 1948 [DOI](#)), while

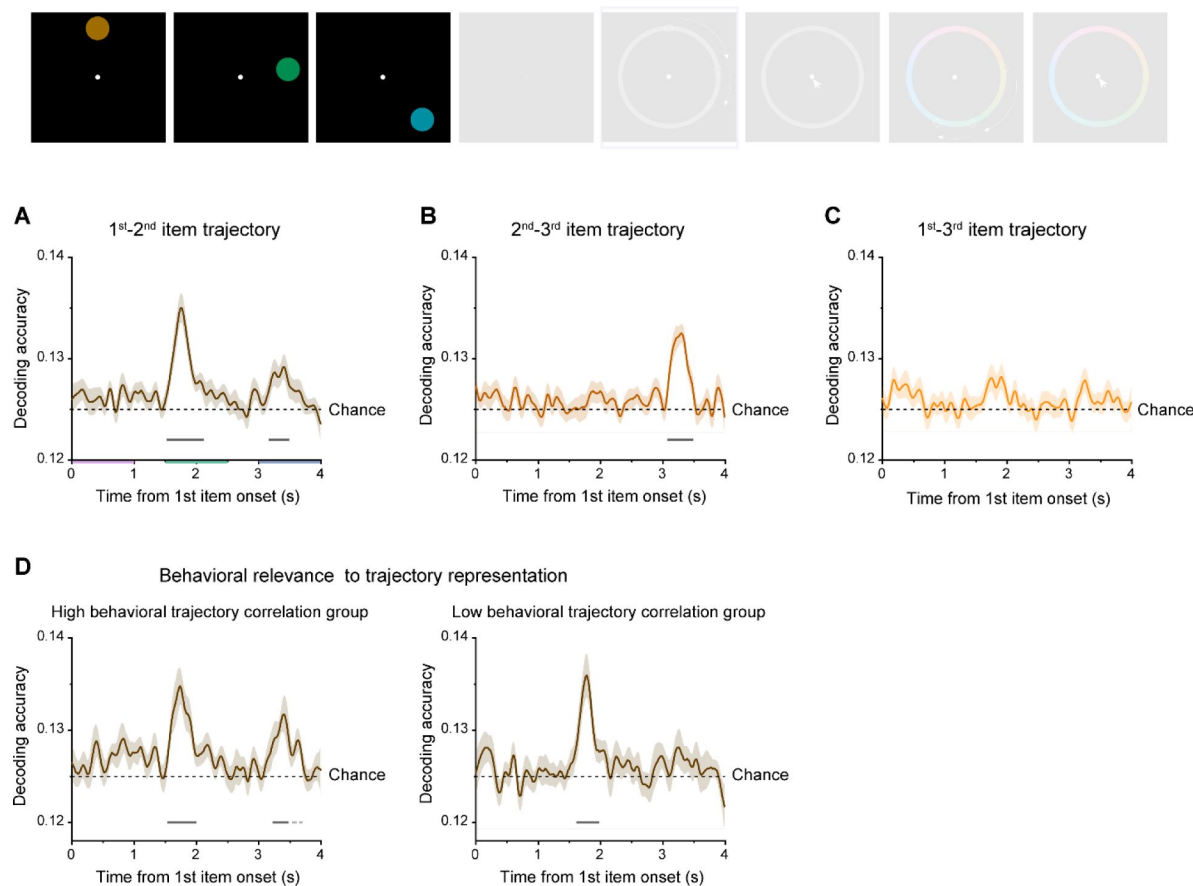


Figure 5.

Common trajectory representation and its behavioral relevance

(A) Grand average (mean $\pm$ SEM) neural decoding of 1st-to-2nd as a function of time during the encoding period, for AT condition. **(B)** Same as A, but for 2nd-to-3rd trajectory. **(C)** Same as A, but for 1st-to-3rd trajectory. **(D)** Participants were divided into two groups (higher-correlation group and lower-correlation group), based on their 1st-to-2nd color-location trajectory memory behavioral correlation. Grand average (mean $\pm$ SEM) neural decoding of 1st-to-2nd trajectory as a function of time, for higher-correlation group (N = 16; left panel) and lower-correlation group (N = 16; right panel). (Horizontal solid line: cluster-based permutation test, cluster-defining threshold $p < 0.05$, corrected significance level $p < 0.1$; Horizontal dashed line: cluster-based permutation test, cluster-defining threshold $p < 0.05$, corrected significance level $p < 0.1$).

schematic abstraction denotes a more broad range of circumstances, whereby the gist or structure of multiple environments or episodes can be integrated (Bartlett, 1932 [↗](#); Farzanfar et al., 2023 [↗](#)). In other words, schema refers to a highly abstract framework of prior knowledge that captures common patterns across related experiences, which does not necessarily occur in a spatial framework as cognitive maps do. In the current design, as we specifically manipulate the consistency of spatial trajectory distance between color and location sequences defined in a spatial map, cognitive map might be a more conservative term to frame our findings, although in essence it reflects general schema-based WM reorganization.

The fact that without task requirement, human participants still spontaneously extracted underlying common structure and leveraged it to organize multiple item storage reflects the intelligence of our brain to achieve efficient information coding (Attneave, 1954 [↗](#)). Indeed, there is a mountain of research suggesting that participants exploit statistical regularities to form efficient WM representation without explicit instructions to do so (e.g., Brady et al., 2009 [↗](#); Brady & Tenenbaum, 2013 [↗](#)). However, it remains largely unknown about the underlying neural mechanism. Here, the common structure we manipulated is inspired by the theory of cognitive map (Behrens et al., 2018 [↗](#); Bellmund et al., 2018 [↗](#); O'keefe & Nadel, 1978 [↗](#)), which has argued that reasoning in abstract domains follows similar computational principles as in spatial domains. This theory has been supported by accumulated neuroscientific evidence suggesting common neural substrates for knowledge representation across domains (Constantinescu et al., 2016 [↗](#); Garvert et al., 2017 [↗](#); Park et al., 2021 [↗](#); Schuck et al., 2016 [↗](#); Solomon et al., 2019 [↗](#); Theves et al., 2019 [↗](#)). In line with these evidence, recent behavioral studies further prove the integration of information representation across domains based on common computing principles. For example, learning process is accelerated when two different feature maps are aligned (Aho et al., 2022 [↗](#)), and distance-dependent generalization is observed across two different domains in order to search for correlated rewards (Wu et al., 2020 [↗](#)). Here, we extend the functional role of cognitive map in efficiently organizing information across domains in human WM and reveal compressive encoding and neural replay in facilitating multiple sequence storage.

Neural replay refers to the sequential reactivation in the same or reversed order as previous experience. It was first observed in the rodent hippocampus and mainly for spatial navigation (Wilson & McNaughton, 1994 [↗](#)), but has recently been found in many higher-level non-spatial tasks in human brains (Kurth-Nelson et al., 2016 [↗](#); Liu et al., 2019 [↗](#); Liu, Mattar, et al., 2021 [↗](#); Schapiro et al., 2018 [↗](#); Schuck & Niv, 2019 [↗](#); Zhang et al., 2018 [↗](#)). In fact, neural replay has been posited to represent abstract structure (Huang et al., 2021 [↗](#); Liu et al., 2019 [↗](#)), structure-based inference (Liu, Mattar, et al., 2021 [↗](#)), and generalization (Barry & Love, 2022 [↗](#)). Here, when subjects prepare to recall location sequences (“location recall”), neural replay occurs for color sequences but not for location sequences, supporting the spontaneous nature of neural replay, since color features are not task-relevant right now. The results also exclude other interpretations, such as motor preparation, eye movements, attentional sampling, sequential rehearsal, etc., since if that is the case, we would expect a similar neural replay profile for location sequences that are to be serially recalled soon. Furthermore, color neural replay only appears for the color-location aligned sequence (AT condition) but not for the misaligned sequences (MAT condition), implicating that neural replay serves to consolidate sequences that share a common structure. Therefore, our findings demonstrate new roles of neural replay in structure representation, that is, mediating structure alignment between sequences in the WM system.

It is posited that structure and content are represented in a factorized manner (Behrens et al., 2018 [↗](#); Bengio et al., 2013 [↗](#)), and sequence structure representation that is independent of attached contents guides the replay of new experiences (Liu et al., 2019 [↗](#)). Factorization representation is thought to help fast generalization of a previously learned structure to new contents (Sheahan et al., 2021 [↗](#); Zhou et al., 2020 [↗](#)). Indeed, the ability to spontaneously perceive relational structures is posited to signify the major distinction between human and nonhuman primates (Dehaene et al., 2015 [↗](#); Zhang et al., 2022 [↗](#)). Meanwhile, previous modelling works also

suggest that higher-order structures incorporated in WM would serve as constraints on individual-item representations to reduce representational uncertainty (Brady & Tenenbaum, 2013 [DOI](#); Ding et al., 2017 [DOI](#)). In this work, we provide behavioral and neural evidence that structure is not only dissociated from content representation, i.e., factorization coding, but also can be aligned in a spontaneous manner, i.e., linking structures, which together contribute to efficient representation of memory information.

Materials and methods

Participants

Thirty-six participants (18 males, age ranging from 17 to 25 years) were recruited to accomplish our multi-sequence working memory task. Three participants were removed, since they could not finish the whole experiment. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to previous studies (J. Li et al., 2021 [DOI](#); Wolff et al., 2017 [DOI](#)). All participants had normal or corrected-to-normal vision with no history of neurological disorders. They were naïve to the purpose of the experiments, and provided written informed consent prior to the start of the experiment. This study was approved by the Research Ethics Committee at Peking University, and carried out in accordance with the Declaration of Helsinki.

Stimuli and tasks

Participants sat in a dark room, 60 cm in front of a Display++ monitor with 100 Hz refresh rate and a resolution of 1920×1080 , and their head stabilized on a chin rest. At the beginning of trial, three disks ($1.5^\circ \times 1.5^\circ$ visual angle) were sequentially presented at different locations of the screen, with different colors. The spatial location of each disk was independently drawn from a fixed set of 9 locations, which were evenly distributed on an imaginary circle with radius of 7° visual angle from central fixation and spaced 40° from the nearest locations, with a small random jitter ($\pm 1^\circ - \pm 3^\circ$) added to each. The color of the each disk was also independently selected from a fixed set of 9 colors, which were evenly distributed along a circle in Commission Internationale de l'Eclairage (CIE) $L^*a^*b^*$ space, and equidistant from the gray point at $L^* = 50$, $a^* = 0$, and $b^* = 0$ (Brouwer & Heeger, 2009 [DOI](#)), and spaced by 40° , with a small random jitter ($\pm 1^\circ - \pm 3^\circ$).

Each disk was presented for 1 s, with 0.5 s interval between two adjacent disks. After 2 s delay during which only the fixation point remained on screen, a grey ring appeared for 0.5 s with the same radius (7° visual angle) from central fixation to instruct participants to recall three spatial locations without any movement. Then, a cursor appeared at fixation, and participants should report the remembered locations sequentially in their presented order by using a mouse to click on the grey location ring. After delivering three spatial location responses, a color ring was presented for 0.5 s (7° visual angle in radius) to instruct participants to recall three colors without any movement. Similarly, a cursor then appeared, and participants were asked to report the remembered colors sequentially in their presented order by clicking on the color ring. To reduce the complexity of the task, the color wheel was oriented the same way for individual participant.

Note that even though color and location were different features, their values were both chosen from nine positions/values based on their respective ring ($0^\circ - 320^\circ$ in 40° increments), with the constraint that the color value and location value for the same item can't be the same. In order to investigate whether common structure would organize multiple information storage in different domains, we modulated trajectory consistency. Specifically, in the aligned trajectory condition (AT), both the 1st-to-2nd and 2nd-to-3rd disk trajectory distances in location domain were the same to that in color domain. In other words, by rotating certain degree, the whole trajectory (from the 1st to 3rd point) was matched in the location and color maps. At the same time, we varied the rotated degree to align the two maps on a trial-by-trial basis, such that we could not predict color sequence solely based on location sequence. This manipulation was critical to independently

decode color sequence and location sequence. In misaligned trajectory condition (MAT), the whole trajectories (1st-2nd-3rd) in the two maps were different, while partial trajectory (either 1st-2nd or 2nd-3rd) can be the same for some trials. Therefore, in MAT condition, we couldn't rotate one map to exactly match the other map. Note that random selection for individual item both for AT and MAT conditions resulted in varied trajectories, which can move in clockwise or anticlockwise direction and the direction can even be reversed on the 3rd item. Trials from AT and MAT conditions were interleaved, aiming to investigate the spontaneous information organization process in a more natural way and avoid prediction about trial type. In each trial three locations were chosen independently from 9 values (0°–320° in 40° increments, each occurred 36 times with random order), but with a constraint that they should at least differ by 40°. The same rule was applied to three colors. Moreover, the color and location value from the same object were also constrained to be different. Participants should complete 648 trials in total, which was divided into two sessions on two separate days, separated by at most one week. It took approximately 3 hours to accomplish one session (including breaks).

EEG acquisition and preprocessing

The EEG data was recorded using a 64-channel EasyCap and two BrainAmp amplifiers (BrainProducts). Horizontal electrooculography (EOG) was recorded by an additional electrode around the participants' right eye. The impedances of all electrodes were kept below 10 k. The EEG data was preprocessed offline using FieldTrip software (Oostenveld et al., 2011 [DOI](#)). Specifically, the data was first referenced to the average value of all channels, band-pass filtered between 2 and 50 Hz, and down-sampled to 100 Hz. The data was then baseline-corrected, by selecting the time range from 300 ms to 100 ms before the presentation of the 1st disk in each trial as baseline to be subtracted. Then, independent component analysis (ICA) was performed independently for each participant to remove eye-movement and artifact components, and the remaining components were back-projected onto the EEG electrode space. To further identify artifacts, we calculated the variance (collapsed over channels and time) for each trial. Trials with excessive variances were removed. Note that following decoding approach was based on the whole electrodes, except that location decoding in the encoding period was based on the posterior electrodes (P7, P5, P3, P1, Pz, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz and O2), considering eye movement was not strictly controlled in the present study.

Data analysis

Behavioral performance analysis

For each spatial location and color, the response error was first quantified by the circular difference between the reported location (color) and the true target location (color) in each trial. The memory precision was then estimated by calculating the reciprocal of circular standard deviation of response error. To explore the similarity of the perceived trajectory in spatial location and color domains, we calculated the circular correlation of the perceived trajectory (trajectory memory error) between the two domains for each participant. Trajectory response error was quantified by the circular difference between the reported trajectory and the true trajectory, e.g., the 1st-to-2nd trajectory error in location was calculated by the difference between the 1st location error and 2nd location error. In group level, we quantified the circular correlations of trajectory error by first sorting trials into four bins based on their location trajectory error for each participant, then binning the trials, computing the color trajectory error for each bin, and pooling the data across participants.

Time-resolved location and color decoding

Similar as previous studies (Brouwer & Heeger, 2009; Huang et al., 2021), in order to assess the time-resolved location and color information from the EEG signals, we implemented the inverted encoding model (IEM) to reconstruct the location and color information from the neural activities at each time point. The IEM assumes that the response in each sensor could be approximated as a linear sum of underlying neural populations encoding different values of the feature-of-interest (i.e., tuning channels). Here, the number of location and color tuning channels were both set to 9. Following previous work (Ester et al., 2015; Yu et al., 2020), the idealized feature tuning curves of nine channels were defined as nine half-wave rectified sinusoids centered at different location (color) values (0°, 40°, 80°, and so on) and raised to the 8th power.

We began by modeling the response of each EEG sensor as a linear sum of nine information channels, characterized by $B_1 = WC_1$, in the training data set, where B_1 (m sensors $\times$ n trials) represents the observed response at each sensor, C_1 (k channels $\times$ n trials) represents the predicted channel responses, w (m sensors $\times$ k channels) represents the weight matrix that characterizes the linear mapping from ‘channel space’ to ‘sensor space’. Therefore, given B_1 , and C_1 , the weight matrix w (m sensors $\times$ k channels) was calculated by using least-squares regression $\hat{W} = B_1 C_1^T (C_1 C_1^T)^{-1}$. Finally, the channel responses (C) for the test data set (B) could be extracted using the estimated $\hat{W}$, by $\hat{C}_2 = (\hat{W}^T \hat{W})^{-1} \hat{W}^T B_2$.

Regarding the division of training and test set, a leave one-out cross-validation was implemented, such that data from all but one block was acted as B_1 to estimate $\hat{W}$ while data from the remaining block was acted as B_2 to estimate $\hat{C}_2$. This procedure ensures the independence between training set and testing set. The entire analysis was repeated until all blocks could be held out as a test set. The observed channel responses $\hat{C}_2$ were then circularly shifted to a common center (0°) in reference to the location/color-of-interest in each trial, and averaged across trials for further analysis.

Consistent with previous studies (J. J. Foster et al., 2017; Huang et al., 2021), decoding performance was characterized by the slope of the estimated channel responses at each time by flipping the reconstructed curves across the center, averaging both sides, and performing linear regression. We further smoothed the slope time courses with a Gaussian kernel (s.d. = 40 ms) (Huang et al., 2021; Wolff et al., 2017).

Note that considering the color wheel is always oriented the same way for each participant, it is possible that participants tend to map colors to locations and remembered in a location code. However, this possibility shouldn’t affect the interpretation of the comparison between AT and MAT conditions. In fact, the underlying logic of the current design is based on the facts that thinking spatially is intuitive and color and location sequences can be spontaneously combined or integrated based on the shared common trajectory structure. In other words, decoding of color sequences could be understood as neural representation of a series of corresponding locations along the ring that are independent of the physical locations of the items.

Forward sequence measure

Following previous studies (Huang et al., 2018; Kurth-Nelson et al., 2016; Liu et al., 2019), cross-correlation was applied to examine whether the color reactivation pattern tended to follow certain order, e.g., a forward (1st-2nd-3rd) or reverse order (3rd-2nd-1st). If it was a forward sequence, the decoded performance of the 1st item at time T should be correlated with the decoding performance of 2nd item at time $T + \Delta t$, and correlated with the decoding performance of 3rd item at time $T + 2\Delta t$, where Δt defines a lag between neural representations of two consecutive items. We first calculated the cross-correlation between the 1st and 2nd items and between 2nd and 3rd items at each time lag. Then we subtracted the reverse direction (2nd-1st, 3rd-

$2^{\text{nd}}-3^{\text{rd}}$) or reverse order ($3^{\text{rd}}-2^{\text{nd}}-1^{\text{st}}$). If it was a forward sequence, the decoded performance of the 1^{st} item at time T should be correlated with the decoding performance of 2^{nd} item at time $T + \Delta t$, and correlated with the decoding performance of 3^{rd} item at time $T + 2*\Delta t$, where Δt defines a lag between neural representations of two consecutive items. We first calculated the cross-correlation between the 1^{st} and 2^{nd} items and between 2^{nd} and 3^{rd} items at each time lag. Then we subtracted the reverse direction ($2^{\text{nd}}-1^{\text{st}}$; $3^{\text{rd}}-2^{\text{nd}}$) from the forward direction ($1^{\text{st}}-2^{\text{nd}}$; $2^{\text{nd}}-3^{\text{rd}}$) respectively at each time-lag, in order to exclude the autocorrelation effect (Kurth-nelson et al., 2016 [\[4\]](#)). The resulted cross-correlation time courses were then averaged to determine the time lag for two consecutive items, here, the time point of the peak of the averaged cross-correlation time course. As mentioned above, for a forward replay pattern, at time lag Δt , we would expect to observe significant transition from the 1^{st} to 2^{nd} items, and from the 2^{nd} to 3^{rd} items, while nonsignificant transition/correlation for the rest pairs, characterized by a theoretical forward transition pattern in **Figure 4E** [\[4\]](#) (left panel). Meanwhile, the actual cross-correlation matrix can be estimated by computing the correlation coefficients for every pair ($1^{\text{st}}-1^{\text{st}}$, $1^{\text{st}}-2^{\text{nd}}$, $1^{\text{st}}-3^{\text{rd}}$, $2^{\text{nd}}-1^{\text{st}}$, $2^{\text{nd}}-2^{\text{nd}}$, $2^{\text{nd}}-3^{\text{rd}}$, $3^{\text{rd}}-1^{\text{st}}$, $3^{\text{rd}}-2^{\text{nd}}$, $3^{\text{rd}}-3^{\text{rd}}$) at the defined time lag (**Figure 4D** [\[4\]](#)). Finally, we quantified the similarity between the observed transition pattern and theoretical forward transition pattern.

Time-resolved trajectory decoding

We implemented a linear support vector machine (SVM) to decode trajectory distance. Considering there were eight possible distances between every two items, i.e., $\pm 160^\circ$, $\pm 120^\circ$, $\pm 80^\circ$, $\pm 40^\circ$ (chance level is 1/8), an eight-way decoder (One-VS-rest multiclass classifier) was used to decode trajectory. A 5-fold cross-validation scheme was used, and the classification accuracy was averaged across the folds. We repeated this process 50 times with each containing a new random partition of data into 5 folds, and then computed their mean accuracy. Note that color and location sequences shared the same trajectory distances in AT condition, while MAT condition involved different trajectories for location and color. The trajectory decoding in MAT condition was based on the trajectory distance in location domain, considering location information showed much stronger representation than color information (**Figure 2** [\[4\]](#)). Same decoding approach was then employed on alpha band power during maintaining period. Here, alpha-band (8 – 12 Hz) power time courses was extracted by conducting a time-frequency analysis using Morlet wavelets with a width of seven cycles (Fieldtrip toolbox).

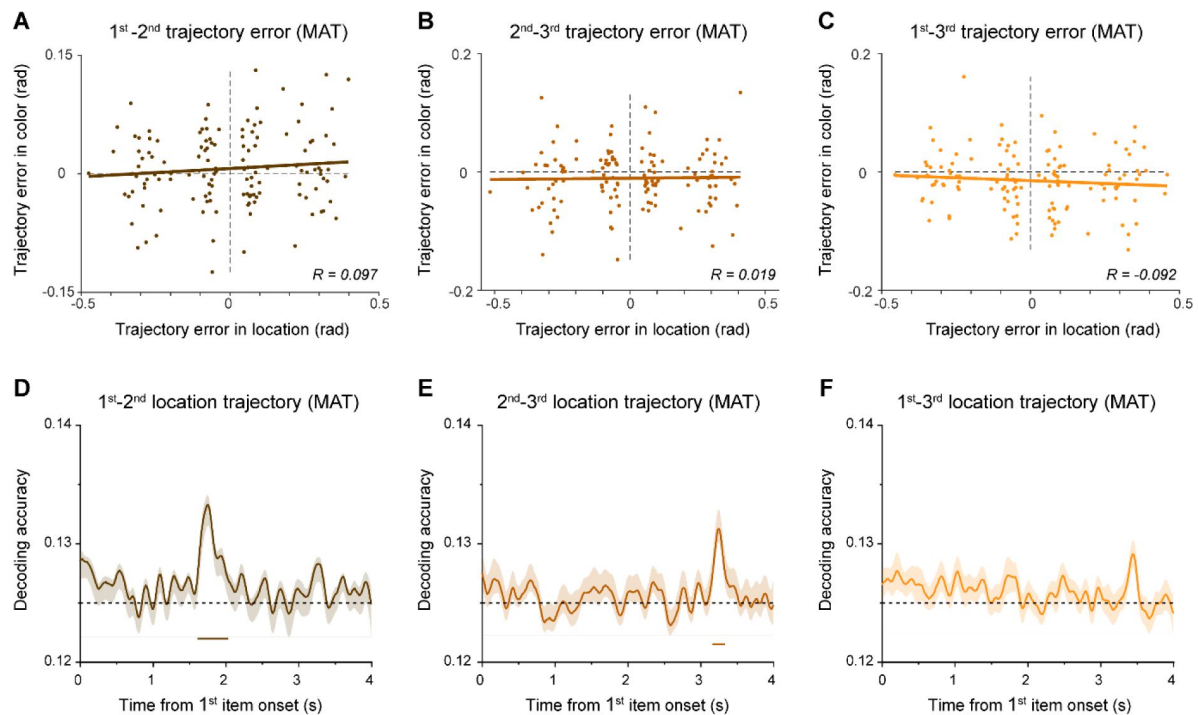
Statistical analysis

To determine statistical significance of decoding performance time courses, we performed cluster-based permutation test (FieldTrip, cluster-based permutation test, 1000 permutations) (Maris & Oostenveld, 2007 [\[4\]](#)). We first identified clusters of contiguous significant time points ($p < 0.05$ or $p < 0.001$ (during encoding period), two-tailed) from the calculated statistics (one-sample t-test, against 0 (slope value of the reconstructed channel response) for location/color decoding, or against 0.125 (classifier chance level) for trajectory distance decoding), and cluster-level statistics was calculated by computing the size of the clusters. Next, a Monte-Carlo randomization procedure was conducted to estimate the significance probabilities for each cluster. Specifically, 0 (for location/color decoding) or 0.125 (for trajectory decoding) with the same sample size was generated and shuffled with the original data 1000 times, and the cluster-level statistics were then calculated from the surrogate data to estimate the significance probabilities for each original cluster.

To determine statistical significance of cross-correlation coefficient (forward direction minus reverse direction), we performed a permutation test by shuffling color labels across participants 1000 times and followed the same procedure to calculate the cross-correlation time courses of the surrogate data, from which the 0.05 threshold level was estimated.

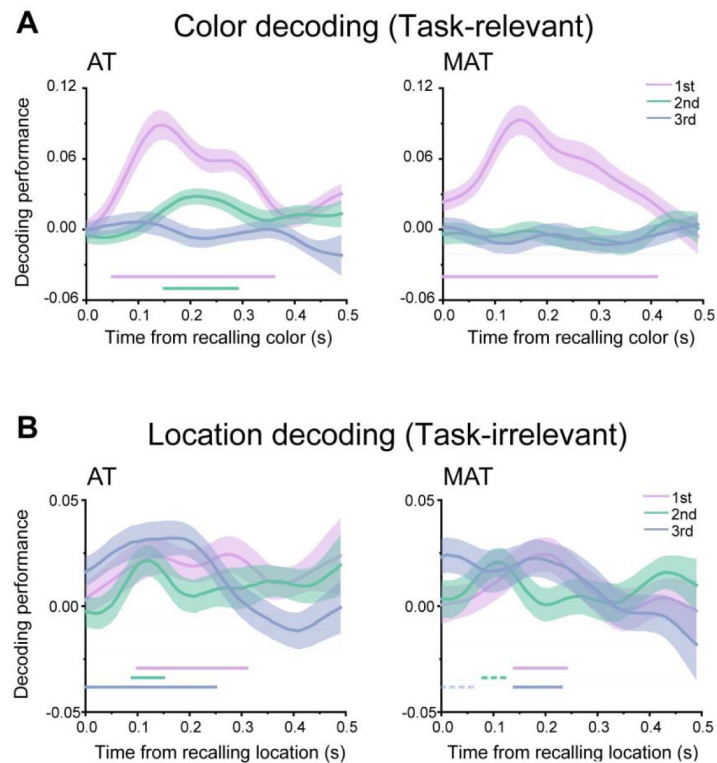
Acknowledgements

This work was supported by the National Science and Technology Innovation STI2030-Major Project (2021ZD0204103 to H.L.), National Natural Science Foundation of China (31930052 to H.L.), and Humboldt Research Fellowship for Postdocs to Q.H. We thank Christian F. Doeller and Muzhi Wang for their helpful comments.



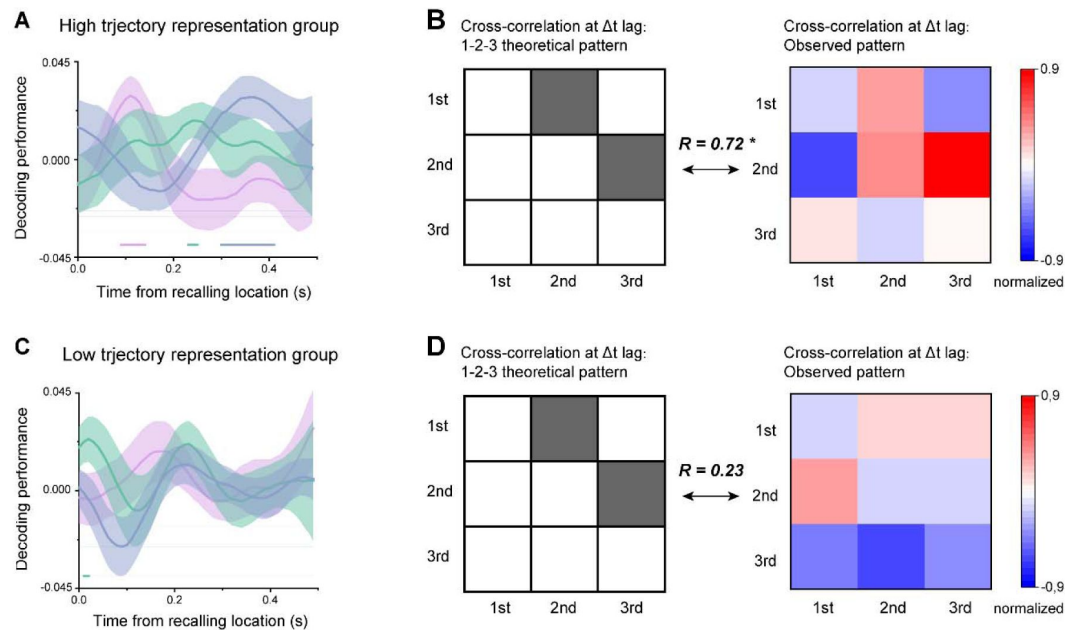
Supplementary Figure 1

(A) Scatterplot of 1st-to-2nd trajectory memory error for location sequence (X-axis) and Color sequence (Y-axis) under MAT condition. Note that the trajectory error of all trials within each subject was divided into 4 bins according to the location trajectory error, resulting in 33 (subject number)*4 (bins) dots in the plot. The brown line represents the best linear fit. **(B)** Same as A, but for 2nd-to-3rd trajectory. **(C)** Same as A, but for 1st-to-3rd trajectory. **(D)** Grand average (mean ± SEM) neural decoding of 1st-to-2nd as a function of time during the encoding period, for MAT condition. **(E)** Same as A, but for 2nd-to-3rd trajectory. **(F)** Same as A, but for 1st-to-3rd trajectory.



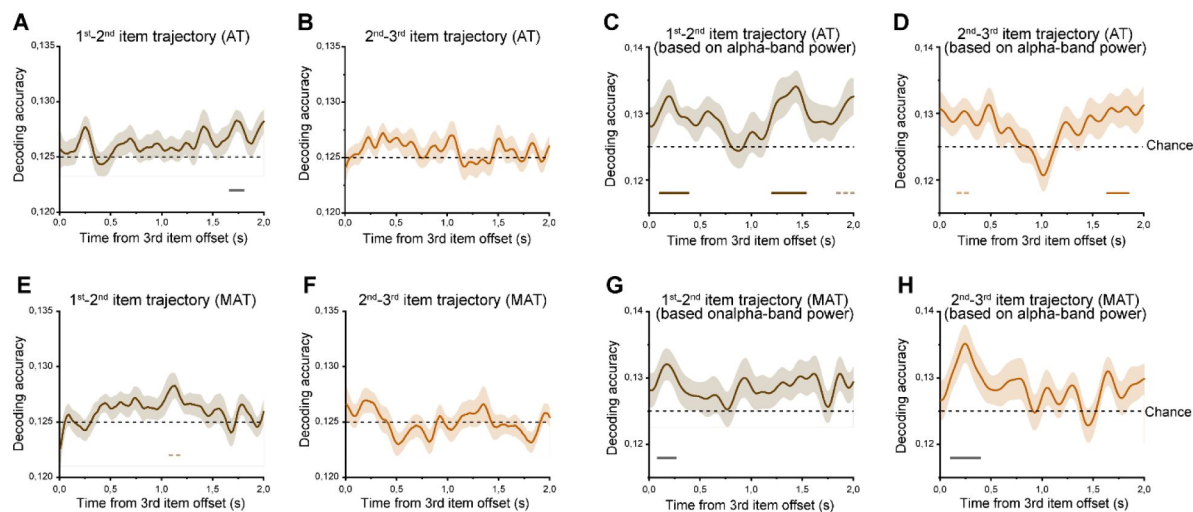
Supplementary Figure 2

(A) Left panel: grand average (mean $\pm$ SEM) time courses of the decoding performance for the 1st (purple), 2nd (turquoise) and 3rd (blue) WM colors during color recalling period. Right panel: grand average (mean $\pm$ SEM) time courses of the decoding performance in MAT condition. (B) Left panel: grand average (mean $\pm$ SEM) time courses of the decoding performance for the 1st, 2nd and 3rd WM locations during color recalling period. Right panel: grand average (mean $\pm$ SEM) time courses of the decoding performance in MAT condition. (Horizontal solid line: cluster-based permutation test, cluster-defining threshold $p < 0.05$, corrected significance level $p < 0.05$; Horizontal dashed line: marginal significance, cluster-defining threshold $p < 0.1$, $0.05 < \text{cluster } p < 0.1$)



Supplementary Figure 3

Participants were divided into two groups based on the average of trajectory decoding performance within the respective significant time range. **(A)** Grand average (mean $\pm$ SEM) time courses of the decoding performance for the 1st, 2nd and 3rd colors during location recalling period for high trajectory representation group. **(B)** Left panel: theoretical transition pattern for 3-item forward replay. Right panel: empirical transitional pattern (actual cross-correlation matrix) at 130 ms time lag defined in **Figure 4D**. A significant correlation was found between the two matrices ($r = 0.715$, $p = 0.030$), further confirming the forward replay of color sequence for high trajectory representation group. **(C)** Same as A, but for low trajectory representation group. **(D)** Same as B, but for low trajectory representation group, nonsignificant correlation between the two matrices was observed ($r = 0.230$, $p = 0.553$). (Horizontal solid line: time window with significant activation (t-test, $p < 0.05$, without correction across time))



Supplementary Figure 4

(A) Grand average (mean $\pm$ SEM) neural decoding of 1st-to-2nd as a function of time during maintaining period, for AT condition. Significant reactivation was observed during a relatively late stage of memory maintenance. **(B)** Same as A, but for 2nd-to-3rd trajectory. No significant reactivation was observed. **(CD)** Considering alpha-band power plays a crucial role in holding information during maintaining period, we also applied the same decoding approach on alpha band power during maintaining period. Specifically, we first conducted a time-frequency analysis using Morlet wavelets with a width of seven cycles (Fieldtrip toolbox) and then extracted alpha-band (8 – 12 Hz) power time courses. A linear SVM was then employed on alpha-band power. Both 1st-to-2nd and 2nd-to-3rd trajectories showed significant reactivation. **(EF)** Same as AB, but for misaligned trajectory condition. **(GH)** Same as CD, but for misaligned trajectory condition. (Horizontal solid line: cluster-based permutation test, cluster-defining threshold $p < 0.05$, corrected significance level $p < 0.05$; Horizontal dashed line: cluster-based permutation test, cluster-defining threshold $p < 0.05$, corrected significance level $p < 0.1$).

References

- Aho K., Roads B. D., Love B. C (2022) **System alignment supports cross-domain learning and zero-shot generalisation** *Cognition* **227** <https://doi.org/10.1016/j.cognition.2022.105200>
- Al Roumi F., Marti S., Wang L., Amalric M., Dehaene S. (2021) **Mental compression of spatial sequences in human working memory using numerical and geometrical primitives** *Neuron* **109**:2627–2639 <https://doi.org/10.1016/j.neuron.2021.06.009>
- Attneave F (1954) **Some informational aspects of visual perception** *Psychological Review* **61**
- Audrain S., McAndrews M. P (2022) **Schemas provide a scaffold for neocortical integration of new memories over time** *Nature Communications* **13**
- Baddeley A (2000) **The episodic buffer: A new component of working memory?** *Trends in Cognitive Sciences* **4**:417–423
- Barry D. N., Love B. C (2022) **A neural network account of memory replay and knowledge consolidation** *Cerebral Cortex* <https://doi.org/10.1093/cercor/bhac054>
- Bartlett F. C (1932) **Remembering: A study in experimental and social psychology**
- Bays P. M., Catalao R. F., Husain M (2009) **The precision of visual working memory is set by allocation of a shared resource** *Journal of Vision* **9**:7–7
- Behrens T. E. J., Muller T. H., Whittington J. C. R., Mark S., Baram A. B., Stachenfeld K. L., Kurth-Nelson Z (2018) **What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior** *Neuron* **100**:490–509 <https://doi.org/10.1016/j.neuron.2018.10.002>
- Bellmund J. L. S., Gärdenfors P., Moser E. I., Doeller C. F (2018) **Navigating cognition: Spatial codes for human thinking** *Science* **362** <https://doi.org/10.1126/science.aat6766>
- Bengio Y., Mesnil G., Dauphin Y., Rifai S (2013) **Better mixing via deep representations** *International Conference on Machine Learning* :552–560
- Brady T. F., Konkle T., Alvarez G. A (2009) **Compression in visual working memory: Using statistical regularities to form more efficient memory representations** *Journal of Experimental Psychology: General* **138**
- Brady T. F., Konkle T., Alvarez G. A (2011) **A review of visual memory capacity: Beyond individual items and toward structured representations** *Journal of Vision* **11**:1–34 <https://doi.org/10.1167/11.5.1>
- Brady T. F., Tenenbaum J. B (2013) **A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates** *Psychological Review* **120**
- Brouwer G. J., Heeger D. J (2009) **Decoding and reconstructing color from responses in human visual cortex** *Journal of Neuroscience* **29**:13992–14003 <https://doi.org/10.1523/JNEUROSCI.3577-09.2009>

- Brouwer G. J., Heeger D. J. (2011) **Cross-orientation suppression in human visual cortex** *Journal of Neurophysiology* **106**:2108–2119 <https://doi.org/10.1152/jn.00540.2011>
- Constantinescu A. O., O'Reilly J. X., Behrens T. E. (2016) **Organizing conceptual knowledge in humans with a gridlike code** *Science* **352**:1464–1468
- Cowan N. (2001) **The magical number 4 in short-term memory: A reconsideration of mental storage capacity** *Behavioral and Brain Sciences* **24**:87–114
- de Vries I. E., Slagter H. A., Olivers C. N. (2020) **Oscillatory control over representational states in working memory** *Trends in Cognitive Sciences* **24**:150–162
- Dehaene S., Meyniel F., Wacongne C., Wang L., Pallier C. (2015) **The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees** *Neuron* **88**:2–19
- Dekker R. B., Otto F., Summerfield C. (2022) **Curriculum learning for human compositional generalization** *Proceedings of the National Academy of Sciences* **119**
- Ding S., Cueva C. J., Tsodyks M., Qian N. (2017) **Visual perception as retrospective Bayesian decoding from high- to low-level features** *Proceedings of the National Academy of Sciences of the United States of America* **114**:E9115–E9124 <https://doi.org/10.1073/pnas.1706906114>
- Doeller C. F., Barry C., Burgess N. (2010) **Evidence for grid cells in a human memory network** *Nature* **463**:657–661
- Ester E. F., Sprague T. C., Serences J. T. (2015) **Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory** *Neuron* **87**:893–905 <https://doi.org/10.1016/j.neuron.2015.07.013>
- Farzanfar D., Spiers H. J., Moscovitch M., Rosenbaum R. S. (2023) **From cognitive maps to spatial schemas** *Nature Reviews Neuroscience* **24**:63–79
- Foster D. J., Wilson M. A. (2006) **Reverse replay of behavioural sequences in hippocampal place cells during the awake state** *Nature* **440**:680–683
- Foster J. J., Bsaies E. M., Jaffe R. J., Awh E. (2017) **Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory** *Current Biology* **27**:3216–3223 <https://doi.org/10.1016/j.cub.2017.09.031>
- Foster J. J., Sutterer D. W., Serences J. T., Vogel E. K., Awh E. (2016) **The topography of alpha-band activity tracks the content of spatial working memory** *Journal of Neurophysiology* **115**:168–177
- Fukuda K., Kang M.-S., Woodman G. F. (2016) **Distinct neural mechanisms for spatially lateralized and spatially global visual working memory representations** *Journal of Neurophysiology* **116**:1715–1727
- Garvert M. M., Dolan R. J., Behrens T. E. (2017) **A map of abstract relational knowledge in the human hippocampal-entorhinal cortex** *Elife* **6**
- Gathercole S. E., Baddeley A. D. (2014) **Working memory and language**

- Gilboa A., Marlatte H. (2017) **Neurobiology of schemas and schema-mediated memory** *Trends in Cognitive Sciences* **21**:618–631
- Goldstone R. L., Medin D. L (1994) **Time course of comparison** *Journal of Experimental Psychology: Learning, Memory, and Cognition* **20**
- Hafting T., Fyhn M., Molden S., Moser M.-B., Moser E. I (2005) **Microstructure of a spatial map in the entorhinal cortex** *Nature* **436**:801–806
- Huang Q., Jia J., Han Q., Luo H (2018) **Fast-backward replay of sequentially memorized items in humans** *eLife* **7**:1–21 <https://doi.org/10.7554/eLife.35164>
- Huang Q., Zhang H., Luo H (2021) **Sequence structure organizes items in varied latent states of working memory neural network** *eLife* **10**:1–21 <https://doi.org/10.7554/elife.67589>
- Kurth-Nelson Z., Behrens T., Wayne G., Miller K., Luettgau L., Dolan R., Liu Y., Schwartenbeck P (2023) **Replay and compositional computation** *Neuron*
- Kurth-Nelson Z., Economides M., Dolan R. J., Dayan P (2016) **Fast sequences of non-spatial state representations in humans** *Neuron* **91**:194–204
- Li H.-H., Sprague T. C., Yoo A. H., Ma W. J., Curtis C. E (2021) **Joint representation of working memory and uncertainty in human cortex** *Neuron* **109**:3699–3712
- Li J., Huang Q., Han Q., Mi Y., Luo H (2021) **Temporally coherent perturbation of neural dynamics during retention alters human multi-item working memory** *Progress in Neurobiology* **102**:23
- Liu Y., Dolan R. J., Higgins C., Penagos H., Woolrich M. W., Ólafsdóttir H. F., Barry C., Kurth-Nelson Z., Behrens T. E (2021) **Temporally delayed linear modelling (TDLM) measures replay in both animals and humans** *Elife* **10**
- Liu Y., Dolan R. J., Kurth-Nelson Z., Behrens T. E. J (2019) **Human Replay Spontaneously Reorganizes Experience** *Cell* **178**:640–652 <https://doi.org/10.1016/j.cell.2019.06.012>
- Liu Y., Mattar M. G., Behrens T. E. J., Daw N. D., Dolan R. J (2021) **Experience replay is associated with efficient nonlocal learning** *Science* **372**
- Maris E., Oostenveld R (2007) **Nonparametric statistical testing of EEG-and MEG-data** *Journal of Neuroscience Methods* **164**:177–190
- Mathy F., Feldman J (2012) **What’s magic about magic numbers? Chunking and data compression in short-term memory** *Cognition* **122**:346–362
- Nieh E. H., Schottdorf M., Freeman N. W., Low R. J., Lewallen S., Koay S. A., Pinto L., Gauthier J. L., Brody C. D., Tank D. W (2021) **Geometry of abstract learned knowledge in the hippocampus** *Nature* **595**:80–84 <https://doi.org/10.1038/s41586-021-03652-7>
- O’keefe J., Nadel L (1978) **The hippocampus as a cognitive map**
- Ólafsdóttir H. F., Bush D., Barry C. (2018) **The role of hippocampal replay in memory and planning** *Current Biology* **28**:R37–R50

- Oostenveld R., Fries P., Maris E., Schoffelen J.-M (2011) **FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data** *Computational Intelligence and Neuroscience* **2011**
- Park S. A., Miller D. S., Boorman E. D (2021) **Inferences on a multidimensional social hierarchy use a grid-like code** *Nature Neuroscience* **24**:1292–1301
- Ren X., Zhang H., Luo H (2022) **Dynamic emergence of relational structure network in human brains** *Progress in Neurobiology* **219**
- Roads B. D., Love B. C (2020) **Learning as the unsupervised alignment of conceptual systems** *Nature Machine Intelligence* **2**:76–82 <https://doi.org/10.1038/s42256-019-0132-2>
- Schapiro A. C., McDevitt E. A., Rogers T. T., Mednick S. C., Norman K. A (2018) **Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance** *Nature Communications* **9**:1–11 <https://doi.org/10.1038/s41467-018-06213-1>
- Schapiro A. C., Rogers T. T., Cordova N. I., Turk-Browne N. B., Botvinick M. M (2013) **Neural representations of events arise from temporal community structure** *Nature Neuroscience* **16**:486–492
- Schuck N. W., Cai M. B., Wilson R. C., Niv Y (2016) **Human orbitofrontal cortex represents a cognitive map of state space** *Neuron* **91**:1402–1412
- Schuck N. W., Niv Y (2019) **Sequential replay of nonspatial task states in the human hippocampus** *Science* **364** <https://doi.org/10.1126/science.aaw5181>
- Sheahan H., Luyckx F., Nelli S., Teupe C., Summerfield C (2021) **Neural state space alignment for magnitude generalization in humans and recurrent networks** *Neuron* <https://doi.org/10.1016/j.neuron.2021.02.004>
- Skaggs W. E., McNaughton B. L (1996) **Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience** *Science* **271**:1870–1873
- Solomon E. A., Lega B. C., Sperling M. R., Kahana M. J (2019) **Hippocampal theta codes for distances in semantic and temporal spaces** *Proceedings of the National Academy of Sciences* **116**:24343–24352
- Sutterer D. W., Foster J. J., Adam K. C., Vogel E. K., Awh E (2019) **Item-specific delay activity demonstrates concurrent storage of multiple active neural representations in working memory** *PLoS Biology* **17**
- Theves S., Fernandez G., Doeller C. F (2019) **The hippocampus encodes distances in multidimensional feature space** *Current Biology* **29**:1226–1231
- Tolman E. C (1948) **Cognitive maps in rats and men** *Psychological Review* **55**
- Tse D., Langston R. F., Kakeyama M., Bethus I., Spooner P. A., Wood E. R., Witter M. P., Morris R. G (2007) **Schemas and memory consolidation** *Science* **316**:76–82
- Van Kesteren M. T., Fernández G., Norris D. G., Hermans E. J. (2010) **Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans** *Proceedings of the National Academy of Sciences* **107**:7550–7555

- Whittington J. C., Muller T. H., Mark S., Chen G., Barry C., Burgess N., Behrens T. E. (2020) **The Tolman-Eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation** *Cell* **183**:1249–1263
- Wilson M. A., McNaughton B. L (1994) **Reactivation of hippocampal ensemble memories during sleep** *Science* **265**:676–679
- Wolff M. J., Jochim J., Akyürek E. G., Stokes M. G (2017) **Dynamic hidden states underlying working-memory-guided behavior** *Nature Neuroscience* **20**:864–871 <https://doi.org/10.1038/nn.4546>
- Wu C. M., Schulz E., Garvert M. M., Meder B., Schuck N. W (2020) **Similarities and differences in spatial and nonspatial cognitive maps** *PLoS Computational Biology* **16**:1–28 <https://doi.org/10.1371/JOURNAL.PCBI.1008149>
- Yu Q., Teng C., Postle B. R (2020) **Different states of priority recruit different neural representations in visual working memory** *PLoS Biology* **18**:1–21 <https://doi.org/10.1371/journal.pbio.3000769>
- Zhang H., Fell J., Axmacher N (2018) **Electrophysiological mechanisms of human memory consolidation** *Nature Communications* **9** <https://doi.org/10.1038/s41467-018-06553-y>
- Zhang H. *et al.* (2022) **Working Memory for Spatial Sequences: Developmental and Evolutionary Factors in Encoding Ordinal and Relational Structures** *The Journal of Neuroscience* **42**:850–864 <https://doi.org/10.1523/JNEUROSCI.0603-21.2021>
- Zhou J., Jia C., Montesinos-Cartagena M., Gardner M. P. H., Zong W., Schoenbaum G (2020) **Evolving schema representations in orbitofrontal ensembles during learning** *Nature* <https://doi.org/10.1038/s41586-020-03061-2>

Editors

Reviewing Editor

Marius Peelen

Radboud University Nijmegen, Nijmegen, Netherlands

Senior Editor

Laura Colgin

University of Texas at Austin, Austin, United States of America

Reviewer #1 (Public Review):

Summary:

Huang and Luo investigated whether regularities between stimulus features can be exploited to facilitate the encoding of each set of stimuli in visual working memory, improving performance. They recorded both behavioural and neural (EEG) data from human participants during a sequential delayed response task involving three items with two properties: location and colour. In the key condition ('aligned trajectory'), the distance between locations of successively presented stimuli was identical to their 'distance' in colour space, permitting a compression strategy of encoding only the location and colour of the first stimulus and the relative distance of the second and third stimulus (as opposed to remembering 3 locations and 3 colours, this would only require remembering 1 location, 1

colour, and 2 distances). Participants recalled the location and colour of each item after a delay.

Consistent with the compression account, participants' location and colour recall errors were correlated and overall lower compared to a non-compressible condition ('misaligned trajectory'). Multivariate analysis of the neural data permitted decoding of the locations and colours during encoding. Crucially, the relative distance could also be decoded - a necessary ingredient for the compression strategy.

Strengths:

The main strength of this study is a novel experimental design that elegantly demonstrates how we exploit stimulus structure to overcome working memory capacity limits. The behavioural results are robust and support the main hypothesis of compressed encoding across a number of analyses. The simple and well-controlled design is suited to neuroimaging studies and paves the way for investigating the neural basis of how environmental structure is detected and represented in memory. Prior studies on this topic have primarily studied behaviour only (e.g., Brady & Tenenbaum, 2013).

Weaknesses:

The main weakness of the study is that the EEG results could make a clearer case for compression. There is some evidence that distance decoding is present in alpha-band activity in the maintenance delay, but the strongest evidence for this occurs only briefly in the late encoding phase (the re-activation of decoding of the distance between items 1 and 2, Fig. 5A). The link to behaviour (Fig. 5D) seems fairly weak and based on a potentially circular analysis. During location recall, colour decoding re-emerges and is reactivated in sequence, but this finding is consistent both with compression-based and conventional rehearsal mechanisms. Nevertheless, the balance of evidence appears to favour the compression account.

Impact:

This important study elegantly demonstrates that the use of shared structure can improve capacity-limited visual working memory. The paradigm and approach explicitly link this field to recent findings on the role of replay in structure learning and will therefore be of interest to neuroscientists studying both topics.

<https://doi.org/10.7554/eLife.93158.2.sa0>

Author response:

The following is the authors' response to the original reviews.

eLife assessment

This valuable study uses a novel experimental design to elegantly demonstrate how we exploit stimulus structure to overcome working memory capacity limits. While the behavioural evidence is convincing, the neural evidence is incomplete, as it only provides partial support for the proposed information compression mechanism. This study will be of interest to cognitive neuroscientists studying structure learning and memory.

Public Reviews:

Reviewer #1 (Public Review):

Summary:

Huang and Luo investigated whether regularities between stimulus features can be exploited to facilitate the encoding of each set of stimuli in visual working memory, improving performance. They recorded both behavioural and neural (EEG) data from human participants during a sequential delayed response task involving three items with two properties: location and colour. In the key condition ('aligned trajectory'), the distance between locations of successively presented stimuli was identical to their 'distance' in colour space, permitting a compression strategy of encoding only the location and colour of the first stimulus and the relative distance of the second and third stimulus (as opposed to remembering 3 locations and 3 colours, this would only require remembering 1 location, 1 colour, and 2 distances). Participants recalled the location and colour of each item after a delay.

Consistent with the compression account, participants' location and colour recall errors were correlated and were overall lower compared to a non-compressible condition ('misaligned trajectory'). Multivariate analysis of the neural data permitted decoding of the locations and colours during encoding. Crucially, the relative distance could also be decoded - a necessary ingredient for the compression strategy.

Strengths:

The main strength of this study is a novel experimental design that elegantly demonstrates how we exploit stimulus structure to overcome working memory capacity limits. The behavioural results are robust and support the main hypothesis of compressed encoding across a number of analyses. The simple and well-controlled design is suited to neuroimaging studies and paves the way for investigating the neural basis of how environmental structure is detected and represented in memory. Prior studies on this topic have primarily studied behaviour only (e.g., Brady & Tenenbaum, 2013).

Thanks for the positive comments and excellent summary.

Weaknesses:

The main weakness of the study is that the EEG results do not make a clear case for compression or demonstrate its neural basis. If the main aim of this strategy is to improve memory maintenance, it seems that it should be employed during the encoding phase. From then on, the neural representation in memory should be in the compressed format. The only positive evidence for this occurs in the late encoding phase (the re-activation of decoding of the distance between items 1 and 2, Fig. 5A), but the link to behaviour seems fairly weak ($p=0.068$).

Thanks for raising this important concern. The reviewer is correct that in principle subjects should employ the compression strategy during the encoding phase when sequence stimuli are presented, yet our results show that the 1-2 trajectory could only be decoded during the late encoding phase.

Meanwhile, subjects could not get enough information to form the compressed strategy for the location and color sequences until the appearance of the 3rd item. Specifically, based on the first two items, the 1st and 2nd item, they only learn whether the 1st-2nd trajectories are congruent between location and color features. However, they could not predict whether it would also apply to the incoming 2nd-3rd trajectory. This is exactly what we found in neural decoding results. The 1st-2nd trajectory could be decoded after the 2nd item presentation, and the 2nd-3rd trajectory appears after the 3rd item onset. Most critically, the 1st-2nd trajectory is reactivated after the 3rd item but only for alignment condition, implicating

formation of the full-sequence compression strategy wherein the previously formed 1st-2nd trajectory is reactivated to be connected to the 2nd-3rd trajectory.

Regarding the difference between higher- and lower-correlation groups, previously we used the time window based on the overall 2nd-3rd neural reactivations, which might not be sensitive to reactivation strength. We now re-chose the time window based on the higher-correlation group (bootstrap test, $p = 0.037$, two sides).

Results have been updated (Figure 5; Results, Page 16). Interpretations about the formation of compression strategy during encoding phase have been added to Results (Page 15-16) and Discussion (Page 18).

Stronger evidence would be showing decoding of the compressed code during memory maintenance or recall, but this is not presented. On the contrary, during location recall (after the majority of memory maintenance is already over), colour decoding re-emerges, but in the un-compressed item-by-item code (Fig. 4B). The authors suggest that compression is consolidated at this point, but its utility at this late stage is not obvious.

Thank you for the important question we apologize for omitting previously - neural evidence for the compressive account.

The reason we did not perform neural decoding during maintenance is that previous EEG/MEG studies including our own failed to reveal robust and sustained time-resolved memory decoding during this period. This is posited to arise from “activity-silent” WM states, wherein memories are not necessarily retained in sustained firing but silently stored within connection weights of WM networks (Stokes, *Trends Cogn. Sci.*, 2015; Rose, *Curr Dir Psychol Sci*, 2020). Our previous work showed that by transiently perturbing the ‘activity-silent’ WM using a retrocue or neutral impulse, memories could be reactivated and robustly decoded from neural activities (Huang et al., eLife, 2021). However, due to the lack of transient events during retention in the current design, we do not expect robust decoding results during maintenance. As shown below (AB), this is indeed what we have observed, i.e., no robust neural decoding of trajectories during retention.

We further used alpha-band (8-11 Hz) neural activities, which have been shown to carry WM information (de Vries et al., *Trends Cogn. Sci.*, 2020; Foster et al., *Curr. Biol.*, 2016; Fukuda et al., *J. Neurophysiol.*, 2016; Sutterer et al., *PLOS Biol.*, 2019) to perform decoding analysis of compression trajectories during maintenance. As shown below, the alpha-band decoding results are indeed stronger than raw activities. Importantly, as shown below (CD), the aligned condition indeed showed significant and long-lasting decoding of compression trajectories (1st-2nd, 2nd-3rd) during retention, while the misaligned condition only showed decoding at the beginning (GH), which might be due to the non-specific offset response of the 3rd item. The results, although not as clear as those during encoding and recalling periods, support the reviewer’s hypothesis that the compressive strategy, if exploited, would be demonstrated during both encoding and maintenance periods. New results and related discussion have been added (Page 16, Supplementary Figure 4).

With regards to the observed item-by-item color replay during location recall, the reviewer was concerned that this was not consistent with the compressive account, given the lack of trajectory decoding.

First, item sequences stored in compressive formats need to be converted to sequences during serial recall. In other words, even though color and location sequences are retained in a compressive format (i.e., common 1st-2nd, 2nd-3rd trajectories) throughout the encoding and retention phases, they should be transferred to two sequences as outputs. This is exactly why we performed decoding analysis on individual color and location items rather than trajectories.

Second and most importantly, we observed serial replay of color sequences when recalling locations. In our view, these results constitute strong evidence for common structure, since the spontaneous color replay during location recall for aligned condition highlights the close bound between color and location sequences stored in WM. In fact, item-by-item serial replay has been well acknowledged as a critical neural index of cognitive maps, not only for spatial navigation but also for higher-order tasks (e.g., Liu et al., Cell, 2019; Liu et al., Science, 2021). Therefore, spontaneous color sequence replay during location sequence recall supports their shared underlying cognitive map.

Finally, spontaneous serial replay is also correlated with the reactivation of compressive trajectories during encoding (Supplementary Figure 3). This further indicates that serial replay during recalling is associated with memory reorganization formed during encoding.

Taken together, we posit that memories need to be converted to sequences as outputs, which leads to serial reactivations during recalling. Importantly, the observed spontaneous replay of color sequences for the aligned condition provides strong evidence supporting the associations between color and location sequences in WM.

We have now added relevant interpretations and discussions (Page 11&13).

Reviewer #2 (Public Review):

Summary:

In this study, the authors wanted to test if using a shared relational structure by a sequence of colors in locations can be leveraged to reorganize and compress information.

Strength:

They applied machine learning to EEG data to decode the neural mechanism of reinstatement of visual stimuli at recall. They were able to show that when the location of colors is congruent with the semantically expected location (for example, green is closer to blue-green than purple) the related color information is reinstated at the probed location. This reinstatement was not present when the location and color were not semantically congruent (meaning that x displacement in color ring location did not displace colors in the color space to the same extent) and semantic knowledge of color relationship could not be used for reducing the working memory load or to benefit encoding and retrieval in short term memory.

Weakness:

The experiment and results did not address any reorganization of information or neural mechanism of working memory (that would be during the gap between encoding and retrieval).

We apologize for not presenting clear neural evidence for memory reorganization, particularly neural decoding during WM maintenance and retrieval, in the previous version. As below, we explain why the findings provide converging neural evidence for WM reorganization based on a shared cognitive map.

First, during the encoding phase when location and color sequences are serially presented, our results reveal reactivation of the 1st-2nd trajectories upon the onset of the 3rd item when location and color sequences are aligned with each other. The reactivation of 1st-2nd trajectory right after the emergence of 2nd-3rd trajectory for aligned but not for misaligned sequences strongly supports WM reorganization, since only stimulus sequences that could be

compressed based on shared trajectories (aligned condition) show the co-occurrence of 1st-2nd and 2nd-3rd trajectories. Moreover, the relevance of 1st-2nd reactivation to behavioral measurements of color-location reorganization (i.e., behavioral trajectory correlation, Figure 5D) further indicates its link to WM reorganization.

Second, the reason we originally did not perform neural decoding during maintenance is that previous EEG/MEG studies including our own failed to reveal robust and sustained time-resolved memory decoding during this period. This is posited to arise from “activity-silent” WM states, wherein memories are not necessarily retained in sustained firing but silently stored within connection weights of WM networks (Stokes, *Trends Cogn. Sci.*, 2015; Wolff et al., *Nat. Neurosci.*, 2017; Rose et al., *Curr Dir Psychol Sci.*, 2020). Our previous work showed that by transiently perturbing the ‘activity-silent’ WM using a retrocue or neutral impulse, memories could be reactivated and robustly decoded from neural activities (Huang et al., *eLife*, 2021). However, due to the lack of transient events during retention in the current design, we do not expect robust decoding results during maintenance. As shown in Supplementary Figure 4(AB), this is indeed what we have observed, i.e., no robust neural decoding of trajectories during retention.

We then used alpha-band (8-11 Hz) neural activities, which have been found to carry WM information (de Vries et al., *Trends Cogn. Sci.*, 2020; Foster et al., *Curr. Biol.*, 2016; Fukuda et al., *J. Neurophysiol.*, 2016; Sutterer et al., *PLOS Biol.*, 2019) to perform decoding analysis of compression trajectories during maintenance. As shown below, the alpha-band decoding results are indeed stronger than raw activities. Importantly, as shown in Supplementary Figure 4(CD), the aligned condition indeed showed significant and long-lasting decoding of compression trajectories (1st-2nd, 2nd-3rd) during retention, while the misaligned condition only showed decoding at the beginning (GH), which might be due to the non-specific offset response of the 3rd item. The results, although not as clear as those during encoding and recalling periods, thus also support WM reorganization.

Finally, during the recalling period, we observed automatic serial replay of color sequences when recalling locations. In our view, these results constitute strong evidence for common structure, since the spontaneous color replay during location recall for aligned condition highlights the close bound between color and location sequences stored in WM. In fact, item-by-item serial replay has been well acknowledged as a critical neural index of cognitive maps, not only for spatial navigation but also for higher-order tasks (e.g., Liu et al., *Cell*, 2019; Liu et al., *Science*, 2021). Therefore, spontaneous replay of color sequence during location recall supports their shared underlying cognitive map. Moreover, the spontaneous serial replay is correlated with the reactivation of compressive trajectories during encoding (Supplementary Figure 3). This further indicates that serial replay during recalling is associated with memory reorganization formed during encoding.

Taken together, we have added updated results about the maintenance period (Page 16, Supplementary Figure 4) and included clarifications and interpretations about why the findings during the encoding and retrieval periods support the WM reorganization view (Page 15-16).

There was also a lack of evidence to rule out that the current observation can be addressed by schematic abstraction instead of the utilization of a cognitive map.

The likely impact of the initial submission of the study would be in the utility of the methods that would be helpful for studying a sequence of stimuli at recall. The paper was discussed in a narrow and focused context, referring to limited studies on cognitive maps and replay. The bigger picture and long history of studying encoding and retrieval of schema-congruent and schema-incongruent events is not discussed.

We agree with the reviewer that cognitive map referred here could be understood as schematic abstraction. Cognitive map refers to the internal representation of spatial relations in a specific environment (Tolman 1948). Schematic abstraction denotes a more broad range of circumstances, whereby the gist or structure of multiple environments or episodes can be integrated (Bartlett, 1932; Farzanfar et al., *Nat. Rev. Neurosci*, 2023).

In other words, schema refers to highly abstract framework of prior knowledge that captures common patterns across related experiences, which does not necessarily occur in a spatial framework as cognitive maps do. Meanwhile, in the current design, we specifically manipulate the consistency of spatial trajectory distance between color and location sequences. Therefore, we would argue that cognitive map is a more conservative and appropriate term to frame our findings.

Relevant discussions have been added (Page 3&19).

We apologize for the lack of more generalized discussion and have added schema-related literatures. Thanks for the suggestion.

Recommendations for the authors:

Reviewer #1 (Recommendations For The Authors):

(1) Do time-frequency-domain data (e.g., alpha-band power) in the delay provide evidence for delay-period decoding of trajectory lengths? This might strengthen the case for compression.

Thanks for the suggestion. We now performed decoding analysis of the delay period based on alpha-band power. As shown in supplementary figure 4, both the 1st-2nd and 2nd-3rd trajectories could be decoded for the aligned condition.

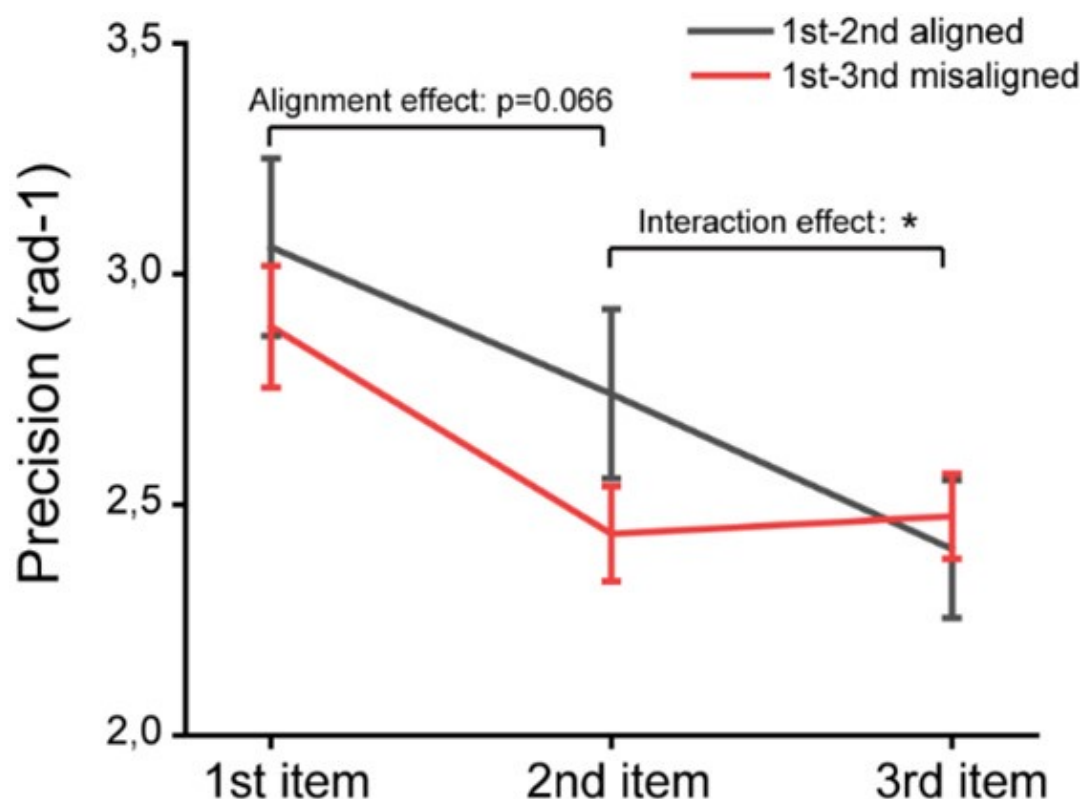
Added in supplementary figure 4 and Page 16.

(2) Do participants erroneously apply the compression strategy in the misaligned condition? This would not show up in the trajectory error correlation analysis, but might be visible when examining correlations between raw trajectory lengths.

Thanks for raising this interesting suggestion. To test the hypothesis, we chose a typical misaligned condition where 1st-2nd trajectory distances are same between location and color sequences, while the 2nd-3rd trajectory distances are different between the two features.

In this case, participants might exploit the compression strategy for the first two items and erroneously apply the strategy to the 3rd item. If so, we would expect better memory performance for the first two items but worse memory for the 3rd item, compared to the rest of misaligned trials. As shown below, the 1st-2nd aligned trials showed marginally significant higher performance than misaligned trials for the first two items ($t(32) = 1.907$, $p = 0.066$, Cohen's $d = 0.332$). Unfortunately, we did not find significant worse performance for the 3rd item between the two conditions ($t(32) = -0.4847$, $p = 0.631$, Cohen's $d = -0.084$). We observed significant interactions between the last two items and the alignment effect ($t(32) = 2.082$, $p = 0.045$, Cohen's $d = 0.362$), indicating a trend of applying wrong compression strategy to the 3rd item.

Author response image 1.



(3a) Some more detail on some of the methods might help readers. For instance, did trajectories always move in a clockwise direction? Could the direction reverse on the third item? If not, did this induce a response bias? Could such a bias possibly account for the trajectory error correlations

Sorry for the unclear statement. For individual trial, both the color and location features of the three items are randomly selected from nine possible values without any constraint about the directions. That is to say, the trajectories can move in a clockwise or anticlockwise direction, and the direction can also reverse on the third item in some trials. Thus, we think the current design can actually help us to reduce the influence of response bias. Taking a step back, if trajectory error correlations are due to response bias, we should expect consistent significant correlation for all conditions, instead of only observing significant correlation for 1st-2nd and 2nd-3rd trajectories but not for 1st-3rd trajectory and only in aligned trajectory condition but not in misaligned condition. Therefore, we think the trajectory error correlations cannot be simply explained by response bias.

Details have been added (Page 23).

(3b) Is the colour wheel always oriented the same way for a participant? If so, given there are only nine colors, it seems possible that colors are mapped to locations and remembered in a location code instead. This does not seem to be a problem in principle for the behavioural findings, but might change the interpretation of what is being decoded from the EEG. If this is a possibility then this might be acknowledged.

The color wheel is always oriented the same way for each participant. We agree with the reviewer that it is possible that participants tend to map colors to locations and remembered

in a location code. We don't have sufficient evidence to rule out this possibility. One possible way could be running another experiment with varied color wheel during response period. Meanwhile, we would like to point out that the underlying logic of the current design is based on the facts that thinking spatially is intuitive and spatial metaphors like "location" and "distance" is commonly used to describe world, e.g., the well-known mental number line (Dehaene et al., *JEP: General*, 1993). Therefore, we expected participants to associate or integrate location and color maps based on trajectory distance.

The reviewer is correct that the color decoding would reflect spatial location rather than the genuine color feature. This is actually the point of the experimental design, whereby two irrelevant features could be possibly combined within a common cognitive map. Without the realignment of the two feature maps defined in space, subjects could not at all form the strategy to compress the two sequences. In other words, decoding of color sequences could be understood as neural representation of a series of corresponding locations along the ring that are independent of the physical locations of the items.

Interpretations and clarifications have been added (Page 23&26).

(4) Does the discretisation of the stimulus distribution (to only 9 possible locations) make the compression strategy easier to use? If the features had been continuously distributed across the location/colour circle, would participants still pick up on and use the shared trajectory structure?

Thanks for the question. Without further data, it's hard to say whether the discretization of the stimulus distribution would make the compression strategy easier to use or not, compared to continuous distribution. Both outcomes seem possible. On the one hand, discrete stimulus distribution would result in discrete trajectory distribution, which helps participants to realize the common trajectory strategy. On the other hand, discrete stimulus distribution would result in category or label representation, which may weaken the effectiveness of structure compression strategy. We postulate that our findings could be generalized to continuous trajectories in a cognitive map within certain resolution.

(5a) Minor point: I disagree that avoiding the same points for location and colour for a given item allows them to be independently decoded. I would argue the contrary - this kind of constraint should create a small anti-correlation that in principle could lead to spurious decoding of one variable (although this seems unlikely here).

We appreciate the concern. As mentioned above, with discrete stimulus distribution (9 possible values for both color and location domains), it is quite possible that a fraction of trials would share same values in location and color. Therefore, the neural decoding for one domain might be confounded by another domain. To dissociate their neural representations, we imposed constraints that color and location could not occupy the same value for a given item.

We agree that this kind of constraint might create a small anti-correlation, even though it is not observed here. Future studies using continuous stimulus distribution would reduce the correlation or anti-correlation between stimuli.

(5b) Very minor point: 1,000 permutations for significance testing seems on the low side. Since some of the p-values are close to 0.05 it may be worth running more permutations.

Thanks for this suggestion. We got similar results using 1000 or 10000 permutations.

(6) Missing reference: H. H. Li et al., 2021 (line 213) seems not to be on the list of references.

Sorry for the mistake. Added.

Reviewer #2 (Recommendations For The Authors):

The study aimed to discuss the working memory mechanism, instead, it seems to be focused on the encoding and recall strategies after a short while, I recommend updating the manuscript to refer to the relevant cognitive mechanism.

There was a strong voice on the effect of using the cognitive map in working memory, without any tests on if indeed a cognitive map was used (for example the novel link between stimuli and how a cognitive map can be used to infer shortcuts). Was the participant required to have any mental map beyond the schema of the shown color ring?

*In the current experiment, to discuss if the effect is driven by utilizing a cognitive map or schematic abstraction of color-relatedness, further analysis is required to possibly assess the effects of schema on neural activity and behavior. Namely,
(1) Was there any reinstatement of schematically congruent (expected) colors that were probed by location 1, at locations 2 and 3 in the MAT condition?*

Thanks for pointing out this possibility. However, we don't think there will be stable color expectations given location information under the MAT condition. First, as the trajectory distance varied on a trial-by-trial basis, no prior common trajectory knowledge could be used to make inference about the current stimuli in individual trial. Second, the starting points for color and location (1st item) were randomly and independently selected, such that color sequence could not be predicted based on the location sequence for both aligned and misaligned conditions.

(2) Given that response time can be a behavioral marker of schematic conflict, was the response time faster for congruent than incongruent conditions?

Thanks for this question. Unfortunately, due to the experimental design, the response time could not be used as a behavioral marker to infer mental conflicts, since participants were not required to respond as fast as possible. Instead, they took their own pace to reproduce sequences without time limit. They could even take a short break before submitting their response to initiate the next trial.

(3) In case you cannot rule out that utilizing schema is the cognitive mechanism that supports working memory performance (the behavior), please add the classical literature (on the memory of schematically congruent and incongruent events) to the discussion.

Thanks for this suggestion and we have added relevant literatures now (Page 3&19).

(4) On page 6, 'common structure in the cognitive map' is the schema, isn't it?

Correct. Based on our understanding, 'common structure in the cognitive map' is a spatial schema.

(5) In Figure 2 EFG, would you please use a mixed effect model or show evidence that all participants demonstrated a correlation between the location trajectory error and color trajectory error?

Thanks for the suggestion. We have added the mixed effect model results, which are consistent with Figure 2EFG (AT: 1st-2nd trajectory, $\beta = 0.071$, $t = 4.215$, $p < 0.001$; 2nd-3rd

trajectory, $\beta = 0.077$, $t = 3.570$, $p < 0.001$; 1st-3rd trajectory, $\beta = 0.019$, $t = 1.118$, $p = 0.264$; MAT: 1st-2nd trajectory, $\beta = 0.031$, $t = 1.572$, $p = 0.116$; 2nd-3rd trajectory, $\beta = 0.002$, $t = 0.128$, $p = 0.898$; 1st-3rd trajectory, $\beta = -0.017$, $t = -1.024$, $p = 0.306$).

In general, doesn't such correlation just show that good participants/trials were good (some did well in the study and some did poorly throughout?)

We don't think the trajectory error correlation results just reveal that some participants did well and some participants did poorly. If that is the case, we shouldn't observe significant correlation in Figure 2D, where we first run correlation for each participant and then test correlation significance at group level. Indeed, trajectory error correlation between color and location domains characterizes the consistent changes between the two domains.

It is worth to note that the correlation was estimated with signed trajectory errors in color and location domains, which meant that we indeed cared about whether the errors in the two domains were consistently varied in the same direction, i.e., whether longer trajectory memory compared to the actual trajectory in location domain would predict longer trajectory memory in color domain.

Moreover, as shown in Figure 2EFG, by dividing trials into 4 bins according to the location trajectory error for each participant and pooling the data across participants, we observed 4 clusters along x-axis (location trajectory error). This suggests that participants' memory performance is rather consistent instead of being extremely good or bad. Besides, if trajectory error correlation is due to different overall memory performance between participants, we should observe significant trajectory error correlations both in AT and MAT conditions, instead of only under AT condition and for 1st-2nd and 2nd-3rd trajectories but not for 1st-3rd trajectory.

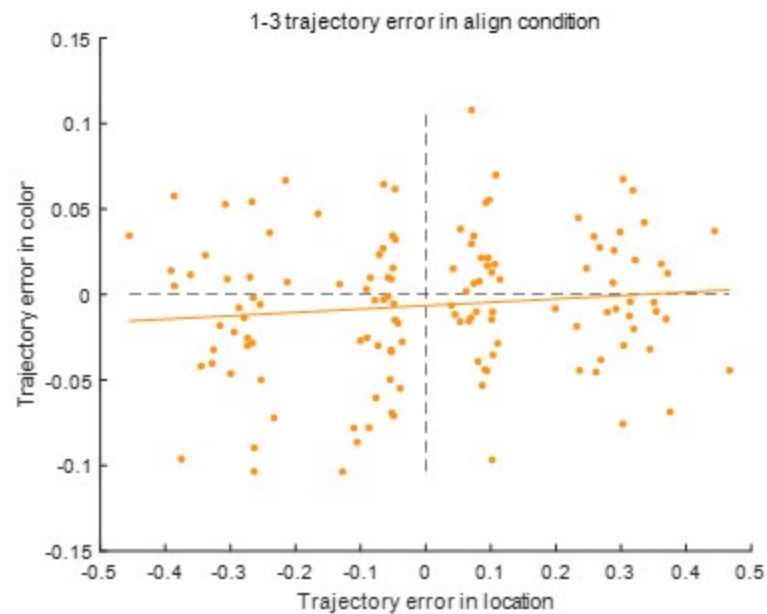
In Figure 2 G, is the marginal error just too big to be sensitive? I am not sure what we are learning here, please clarify.

Sorry for the confusion. To examine this possibility, we excluded errors which are beyond $2.5 * \sigma$, and still observed non-significant 1st-3rd trajectory error correlation between color and location domains ($r = 0.119$, $p = 0.167$).

The 1st-3rd trajectory showed nonsignificant behavioral correlation and neural representation, which suggests that the current sequential memory task would encourage participants to organize all information by relying more on the adjacent items and their distance. Thus, we think the 1st-3rd trajectory would serve as a control trajectory, which helps us not only exclude other possible explanation (e.g., systematic response bias), but also validate current findings both in behavioral and neural level.

Results and statements (Page 10-11) added now.

Author response image 2.



(6) Regarding the first lines on page 11, did you do qualitative research to know if less information was encoded in congruent conditions?

The current experimental design is inspired by the mental compression of spatial sequence studies from Dehaene's lab (Amalric et al., 2017; Roumi et al., 2021), in which they propose that human brain compresses spatial sequence using an abstract language and formalize minimal description length of a sequence as the "language-of-thought complexity." Based on this evidence, we think less information is required to describe congruent condition compared to incongruent condition. This idea is supported by better memory performance for congruent condition. Unfortunately, we couldn't manage to quantify how less information was encoded in congruent condition.

<https://doi.org/10.7554/eLife.93158.2.sa2>