Progress in Neurobiology 201 (2021) 102023

Contents lists available at ScienceDirect

Progress in Neurobiology

journal homepage: www.elsevier.com/locate/pneurobio

Temporally coherent perturbation of neural dynamics during retention alters human multi-item working memory

Jiaqi Li [a,b,c,1], Qiaoli Huang [a,b,c,1], Qiming Han [a,b,c,d], Yuanyuan Mi [e,f,**], Huan Luo [a,b,c,*]

[a] School of Psychological and Cognitive Sciences, Peking University, China
[b] PKU-IDG/McGovern Institute for Brain Research, Peking University, China
[c] Beijing Key Laboratory of Behavior and Mental Health, Peking University, China
[d] Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China
[e] Center for Neurointelligence, School of Medicine, Chongqing University, Chongqing 400044, China
[f] AI Research Center, Peng Cheng Laboratory, Shenzhen 518005, China

ARTICLE INFO

Keywords:
Working memory
Recency effect
Short-term plasticity (STP)
Hidden state
Memory manipulation
Dynamic perturbation

ABSTRACT

Temporarily storing a list of items in working memory (WM), a fundamental ability in cognition, has been posited to rely on the temporal dynamics of multi-item neural representations during retention. However, the causal evidence, particularly in human subjects, is still lacking, let alone WM manipulation. Here, we develop a novel "dynamic perturbation" approach to manipulate the relative memory strength of WM items held in human brain, by presenting temporally correlated luminance sequences during retention to interfere with the multi-item neural dynamics. Six experiments on more than 150 subjects confirm the effectiveness of this WM manipulation approach. A computational model combining continuous attractor neural network (CANN) and short-term synaptic plasticity (STP) principles further reproduces all the empirical findings. The model reveals that the "dynamic perturbation" modifies the synaptic efficacies of WM items through STP principles, eventually leading to changes in their relative memory strengths. Our results support the causal role of temporal dynamics of neural network in mediating multi-item WM, and offer a promising, purely bottom-up approach to manipulate WM.

1. Introduction

Temporarily holding a sequence of items in working memory (WM) for subsequence planning or action is an essential function in many cognitive processes, such as language, movement control, episodic memory, and decision making (Dell et al., 1997; Burgess and Hitch, 1999; Cowan, 2001; Doyon et al., 2003; Giraud and Poeppel, 2012). Psychological evidence and computational models suggest that this process relies on attentional refresh (Awh et al., 1998; Baddeley, 2003; Camos et al., 2018; Oberauer, 2019) or neural reactivations during retention (Lisman and Idiart, 1995; Horn and Opher, 1996; Compte et al., 2000; Raffone and Wolters, 2001; Siegel et al., 2009; Heusser et al., 2016; Lundqvist et al., 2016; Michelmann et al., 2016; Huang et al., 2018; Liu et al., 2019; Schuck and Niv, 2019). Interestingly, recent studies and computational modellings advocates a hidden-state WM network view that does not demand persistent firing during maintenance (Stokes, 2015; Fiebig and Lansner, 2017; Trübutschek et al., 2017; Wolff et al., 2017). Instead, items could be temporarily held in synaptic weights via short-term plasticity (STP) principles (Fusi, 2008; Mongillo et al., 2008; Barak and Tsodyks, 2014; Mi et al., 2017). In fact, the two seemingly contradictory proposals – activate state vs. hidden state – could be well combined into a unified framework whereby the reactivation and STP mutually facilitate each other and together subserve WM storage (Miller et al., 2018; Masse et al., 2019, 2020). In this WM network, multiple items, by sharing the same connection to an inhibitory pool, compete with each other in time and take turns to elicit brief bouts of reactivation to strengthen their individual synaptic weights (Fino and Yuste, 2011; Mi et al., 2017). Therefore, the dynamic rivalry between item reactivations and the accompanying synaptic strength changes that accumulate over retention would essentially determine the memory strength of each item in the WM network.

* Corresponding author at: School of Psychological and Cognitive Sciences, Peking University, Room 1703, 52 Haidian Road, Beijing 100087, China.
** Corresponding author at: Center for Neurointelligence, School of Medicine, Chongqing University, Chongqing 400044, China.
E-mail addresses: miyuanyuan0102@cqu.edu.cn (Y. Mi), huan.luo@pku.edu.cn (H. Luo).
[1] These authors contributed equally to the work.

https://doi.org/10.1016/j.pneurobio.2021.102023
Received 9 September 2020; Received in revised form 1 January 2021; Accepted 11 February 2021
Available online 19 February 2021

Previous studies have examined the neural coding of items and the top-down modulation of WM representations (Sauseng et al., 2009; Rose et al., 2016; Mallett and Lewis-Peacock, 2018; Van Ede et al., 2018), yet the causal evidence for the role of retention-stage neural dynamics in mediating multi-item WM is largely absent. In fact, unlike optogenetics tools widely used in animal study, there is yet no time-resolved approach to manipulate the dynamics of WM network in the human brain. Here we developed a novel "dynamic perturbation" approach aiming to modify the relative memory strength of a list of items maintained in WM. Specifically, previous studies show that task-irrelevant color features will be automatically bound to the to-be-memorized feature (e.g., orientation) in WM when they belong to the same object (Luck et al., 1997; Johnson et al., 2008; Hyun et al., 2009; Huang et al., 2018). Therefore, by manipulating the ongoing temporal relationship between the continuously flickering color probes that are bound to each WM item respectively, we could presumably apply temporally correlated perturbations to these item-specific neural assemblies in the WM network. Consequently, this temporal perturbation would impact the dynamical WM network in certain way via modulating the ongoing competition between items, finally leading to the changes in their relative STP-based memory strength.

Six experiments consistently show that the "dynamic perturbation" approach successfully alters the multi-item WM recalling performance on a trial-by-trial basis. We focused on the recency effect (i.e., better memory performance for recently than earlier presented item), a typical WM behavioral index to characterize the relative memory strength of a sequence of items (Burgess and Hitch, 1999; Gorgoraptis et al., 2011; Baddeley, 2012; Jones and Oberauer, 2013). First, when applying synchronized continuous luminance inputs to the color probes ("Synchronization" manipulation) during retention, the recency effect is significantly disrupted. In contrast, the recency effect keeps intact when luminance inputs are temporally uncorrelated ("Baseline" condition). Second, when the luminance sequences of the color probes are temporally shifted with each other in an order that is either the same or reversed as the stimulus sequence ("Order reversal" manipulation), they lead to distinct and even reversed recency effect. Finally, we established a theoretical continuous attractor neural network (CANN) with STP effects (Brody et al., 2003; Romani and Tsodyks, 2015; Trübutschek et al., 2017; Seeholzer et al., 2019) to mimic the "dynamic perturbation" experiments. The simulation results successfully replicated all our experiment findings, further corroborating the neural mechanism for the WM manipulation approach. Taken together, our results constitute new causal evidence for the essential role of STP-based neural dynamics in multi-item WM and provides an efficient, bottom-up approach to manipulate WM in humans.

2. Materials and methods

2.1. Participants

Two hundred and four participants (95 males, age ranging from 18 to 26 years) took part in eight experiments (Expt. 1–5 and Expt. 2-IC in main text, Expt. 6 (probe-absent experiment), Expt. 7 (spatial location experiment) in Supplementary Materials). One subject in Expt. 1, two in Expt. 2, two in Expt. 2-IC, three in Expt. 3, four in Expt. 4, three in Expt. 5, one in Expt. 6 and five subjects in Expt. 7 were excluded due to their overall low memory performance, or self-reported strategies, or not finishing the whole experiment, resulting in 27 subjects for each experiment included in the main text (i.e., Expt. 1–5, Expt. 2-IC). The sample size (27) was determined based on a pretest on 20 subjects using the same paradigm as Experiment 2 (G*Power 3.1, Erdfelder et al., 2009). Moreover, Expt. 6 and Expt. 7 included in supplementary materials had 16 and 26 subjects, respectively. All the participants were naïve to the purpose of the experiments and had normal or corrected-to-normal vision with no history of neurological disorders, and have provided written informed consent before the experiments. All experiments were carried out in accordance with the Declaration of Helsinki and have been approved by the Research Ethics Committee at Peking University.

2.2. Stimuli and tasks

2.2.1. Experiment paradigm

Participants sat in a dark room, 58 cm in front of a CRT monitor (except Display++ monitor in Experiment 4 using the same parameters) with 100 Hz refresh rate and a resolution of 1024 * 768. Subjects performed a working memory task, with their head stabilized on a chin rest. Each trial consisted of three periods: encoding, maintaining, and recalling (Fig. 1A). In the "encoding period", subjects were serially presented with two bars (Experiment 1, 2, 2-IC, 3, 4, 5, 6) displayed at the center. Each bar stimulus lasted for 1 s with a 0.5 s inter-stimulus interval and had different orientation and color features. Subjects were instructed to memorize the orientations of the bar stimuli. In the "maintaining period", participants performed a central fixation task by monitoring an abrupt luminance change (within 25 % of total trials) and reported if they detected changes after the maintenance period. All trials were included into further analysis. Finally, in the "recalling period", an instruction cue was presented for 1.5 s to indicate which orientation should be recalled (1st or 2nd orientation), and participants needed to rotate a horizontal white bar by pressing corresponding keys to the memorized orientation as precise as possible, without time limit.

In each trial, after a 0.5 s fixation period, participants performed a two-item sequence memory task. The 1st orientation was chosen randomly between 0° and 180°, and the 2nd orientation was generated based on the 1st orientation with the tiled angles of ±17°, ±24°, ±38°, ±52°, ±66°, ±80° (Liu and Becker, 2013). The colors of the 1st and 2nd bar were different (randomly selected from red, green, or blue in each trial) and the luminance of all colors has been corrected using gamma correction. During the maintenance period, after a 0.5−1 s blank interval, two task-irrelevant discs (3° in radius) that had the color feature of either the 1st or 2nd memorized bar, referred as the 1st memory-related probe (WM1) and 2nd memory-related probe (WM2) respectively, were displayed at the left or right side of the fixation (7° visual angle) for 5 s. The color (red, green, or blue) and the spatial locations (left or right) of the two probes (WM1 and WM2) were carefully balanced across trials to eliminate possible color-specific effect. Crucially, throughout the 5 s maintenance period, the corrected luminance of the two color discs (WM1 and WM2) was continuously modulated according to two 5 s temporal sequences, respectively (i.e., Seq1 for WM1, Seq2 for WM2), ranging from dark (0 cd/m^2) to bright (15 cd/m^2). Note that Seq1 and Seq2 were generated anew in each trial.

2.2.2. Experiment 1–2 ("temporal synchronization" manipulation)

In Experiment 1, the relationship between Seq1 and Seq2 during the maintenance period varied in three ways: Baseline, alpha-band synchronization (A-Sync), or full-spectrum synchronization (F-Sync). For Baseline condition, Seq1 and Seq2 were two independently generated random time series (Seq1≠Seq2). For F-Sync condition, Seq1 and Seq2 were the same random temporal series (Seq1 = Seq2). For A-Sync condition, Seq1 and Seq2 were the same alpha-band (8−11 Hz) time series (Seq1 = Seq2). Experiment 1 had 216 trials in total (72 trials for each of the three conditions), divided into three blocks. The three conditions were randomly mixed within each experimental block.

Experiment 2 only examined the Baseline and A-Sync conditions. Experiment 2 had 144 trials in total (72 trials for each of the two conditions), divided into three blocks. The two conditions were randomly mixed within each experimental block.

Experiment 2-IC employed the same experimental design and color selection (adding yellow) as Experiment 2, except that the color probes during the delay period had memory-irrelevant colors, i.e., different from the colors of the to-be-memorized bars. Specifically, when the to-be-remembered two bars were blue and red in color, the color probes

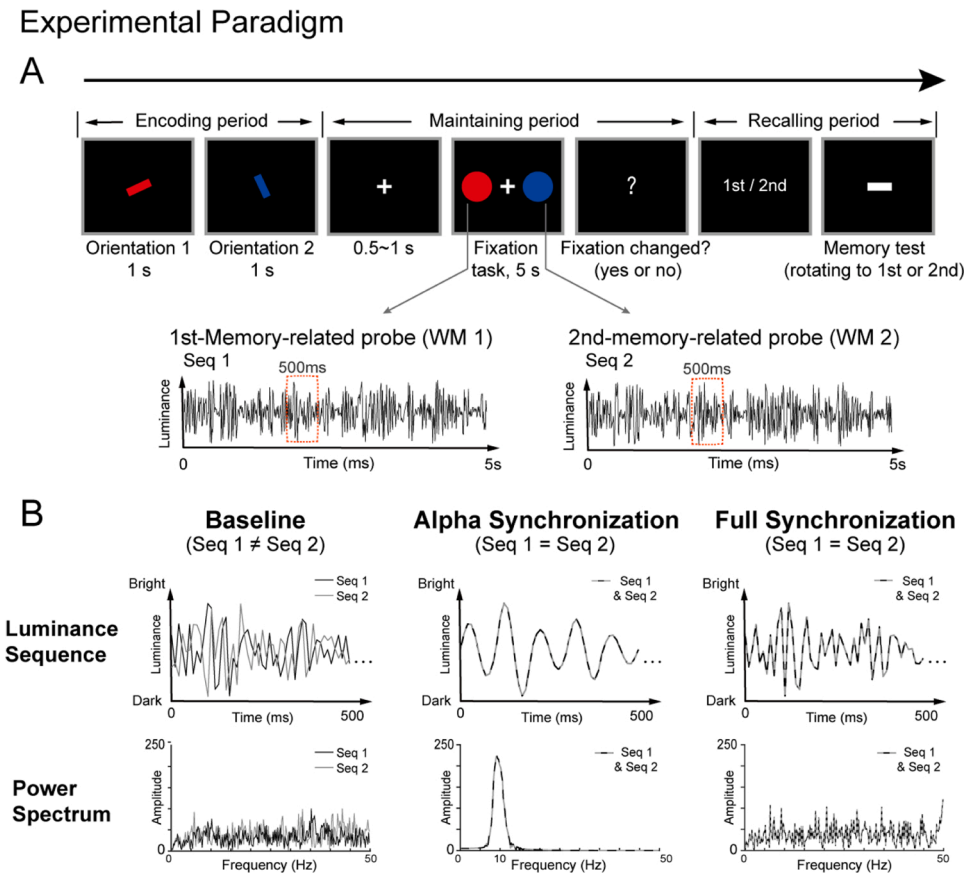


Fig. 1. "Dynamic perturbation" approach: forming temporal associations between WM items via luminance sequences during memory retention.
(A) Experimental paradigm. In each trial, two bar stimuli were serially presented at the center. Subjects needed to memorize the orientations of the two bars (1st and 2nd orientation) and rotated a horizontal bar to the memorized orientation based on the instruction (1st or 2nd). During the in-between 5 s maintaining period, subjects performed a central fixation task, while two task-irrelevant discs that had color feature of either the 1st (WM1) or the 2nd (WM2) memorized bars were displayed at left or right side of the fixation. The color and the spatial locations of WM1 and WM2 were balanced across trials. The luminance of WM1 and WM 2 was continuously modulated according to two 5 s temporal series, respectively (Seq1 for WM1, Seq2 for WM2). (B) Temporal relationship of WM1 and WM2 during maintenance. Upper: Illustration of 500 ms segment of the 5 s Seq1 (dark line) and Seq2 (grey line). Lower: Spectrum of the whole 5 s Seq1 (dark) and Seq2 (grey). Left (Baseline condition): Seq1 and Seq2 were two independently generated random time series. Middle (A-Sync condition): Seq1 and Seq2 shared the same alpha-band (8−11 Hz) time series. Right (F-Sync condition): Seq1 and Seq2 shared the same random time series. Note that Seq1 and Seq2 were generated anew in each trial.

presented during the delay period would be set to yellow and green, and vice versa.

2.2.3. Experiment 3–5 ("order reversal" manipulation)

Experiment 3–5 employed the same paradigm, except that the temporal relationship between Seq1 and Seq2 were manipulated in a different manner, i.e., Seq1 and Seq2 were a temporal shifter version of each other. Specifically, in each trial, Seq1 was a randomly generated temporal series, and Seq2 was then set as a temporal shifted version of Seq, either leading or lagging Seq2 by 200 ms, corresponding to the "Same Order" and "Reversed Order" conditions, respectively (see illustration in Fig. 3A).

Notably, Seq1 and Seq2 occurred simultaneously rather than asynchronously, since the final 200 ms segment of the leading sequence would be attached to the beginning of the lagged sequence (red line, Fig. 3A). Furthermore, in order to replicate the findings of Experiment 1–2, Baseline and F-Sync conditions were also tested in Experiment 3 and 4, respectively, i.e., Experiment 3 (Same-order, Reversed-order, Baseline) and Experiment 4 (Same-order, Reversed-order, F-Sync). Each condition had 72 trials, resulting in 216 trials for Experiment 3 and 4.

Experiment 5 contained three conditions (Same-order, Reversed-order, Baseline), except that the Seq1-Seq 2 temporal lag was now set to 500 ms. Each condition had 72 trials, resulting in 216 trials for Experiment 5.

To examine the overall "Order Reversal" manipulation effect (Same-order vs. Reversed-order) across Experiment 3 and 4 given the exactly same two conditions in the two experiments, we assessed the behavioral results using a two-sample two-way repeated ANOVA analysis (Experiment * Position * Order).

2.3. Data analysis

A mixture-modelling analysis (Bays et al., 2009) was used to quantify the multi-item WM behavioral performance for each of the item, by quantifying the Von Mises distribution of the reported error. The three-parameter mixture model calculates the respective proportions of target, non-target, and uniform responses. In the present study, we used the target probability to estimate the memory accuracy, which has been widely used to quantify memory performance (Bays et al., 2009; Gorgoraptis et al., 2011; Van Ede et al., 2018). Moreover, the target probability was further normalized by an empirical logit transformation: $\text{logit}(p) = \ln((p+\frac{1}{2n})/(1-p+\frac{1}{2n}))$, where p is target probability and n is the number of observations (Cox and Snell, 1989; de Smith, 2010; Stevens et al., 2016). The normalized target probabilities were used for further statistical tests in all the experiments.

The Meta-analysis across experiments were performed by using Comprehensive Meta-Analysis Software (CMA) across 7 experiments (Expt. 1–5 in main text, Expt. 6–7 in Supplementary Materials). We calculated the recency effect for Baseline condition across Experiment 1, 2, 3, 5, 6, 7, the recency effect for Synchronization condition across Experiment 1, 2, 4, 6, 7, and the interaction effect between Baseline and Synchronization conditions across Experiment 1, 2, 7. Meta-analysis has also been performed on target probability without normalization (Supplementary Fig. 6) and precision measurements (Supplementary Fig. 5).

2.4. Continuous attractor neural networks (CANNs) with short-term synaptic plasticity (STP)

2.4.1. Continuous attractor neural networks

We adopt a continuous attractor neural network model (CANN) with short-term synaptic plasticity (STP) to implement the encoding of orientated bars in the neural system (Brody et al., 2003; Trübutschek

et al., 2017; Seeholzer et al., 2019). As shown in Fig. 5A, neurons are aligned on a ring according to their preferred orientations (θ), which are in the range of $(-\pi/2, \pi/2]$ with the period boundary condition (Ben-Yishai et al., 1995; Blumenfeld et al., 2006; Wu et al., 2016). Neurons are connected recurrently via STP-based synapses (Fung et al., 2012; Mi et al., 2016). All neurons on the ring are further connected to a common inhibitory neural pool (Kim et al., 2017; Mi et al., 2017).

Denote $h_E(\theta, t)$ the synaptic inputs at time t to neurons with orientation preference θ, whose dynamics is written as

$$\tau\frac{\partial h_E(\theta,t)}{\partial t} = -h_E(\theta,t) + \int_{-\pi/2}^{\pi/2}\widetilde{J}(\theta,\theta')u(\theta',t)x(\theta',t)r_E(\theta',t)d\theta' \\ -J_{EI}r_I + I_{ext}(\theta,t) + I_0 + \sigma_1\zeta_1(\theta,t), \tag{1}$$

where τ is the time constant of synaptic currents. $r_E(\theta, t)$ refers to the firing rate of neurons that encodes orientation θ and is given by $r_E(\theta,t) = \alpha\ln[1+\exp(h_E(\theta,t)/\alpha)]$, which is a smoothed threshold-linear function ($r_E(h_E<0)\approx 0$ and $r_E(h_E>0)\approx h_E$) with α determining the firing rate of a neuron around $h \approx 0$ (Mi et al., 2017). r_I refers to the firing rate of the inhibitory neural pool, and J_{EI} is the connection strength from the inhibitory neural pool to neurons on the ring. $I_{ext}(\theta, t)$ refers to the external inputs to the neuronal group encoding orientation θ, i.e., loaded memorized items during the encoding period, flickering probes during the maintaining period, and the recalling signals during the retrieval period. $I_0 + \sigma_1\varsigma_1(\theta, t)$ denotes the background input, with $\varsigma_1(\theta, t)$ as the Gaussian white noise of zero mean and unit variance, and σ_1 as the noise strength. In the simulation, since the number of neurons is limited, the integration in Eq.1 is written as:

$$\int_{-\pi/2}^{\pi/2}\widetilde{J}(\theta,\theta')u(\theta',t)x(\theta',t)r_E(\theta',t)d\theta' = \frac{\pi}{N}\sum_{k=1}^{N}\widetilde{J}(\theta,\theta^k)u(\theta^k,t)x(\theta^k,t)r_E(\theta^k,t), \tag{2}$$

where N is the number of neurons.

$\widetilde{J}(\theta,\theta')$ is the absolute synaptic strength between neurons encoding θ and θ' (Romani and Tsodyks, 2015; Trübutschek et al., 2017), which is set to be:

$$\widetilde{J}(\theta,\theta') = \begin{cases} J\cos[B\times(\theta-\theta')], & if B\times(\theta-\theta')\in[-\arccos(-J_0/J),\arccos(-J_0/J)], \\ -J_0, & else, \end{cases} \tag{3}$$

where J, J_0 and B are constants determining the strength and the range of the neuronal interactions, and J is the maximum value of the synaptic efficacy of neurons. Notably, the neuronal connection profile has a bell-shape (i.e., strong connection to neighborhood and weak connection to distal neurons) and is translation-invariant in the feature space, i.e. $\widetilde{J}(\theta, \theta')$ is a function of $(\theta - \theta')$.

The dynamics of inhibitory neural pool is given by

$$\tau\frac{dh_I}{dt} = -h_I + J_{IE}\int_{-\pi/2}^{+\pi/2}r_E(\theta,t)d\theta, \tag{4}$$

where r_I and h_I denote the firing rate and synaptic current of the inhibitory neural pool, respectively, with $r_I(h_I) = \alpha\ln(1+\exp(h_I/\alpha))$, and J_{IE} as the connection strength from neurons on the ring to the inhibitory neural pool.

Overall, the effect of the recurrent connections $\widetilde{J}(\theta,\theta')$ enables the network to generate localized neural responses at specific location on the ring, and the role of global inhibition is to introduce competition among neuronal groups, which would be further modulated by STP.

2.4.2. Short-term synaptic plasticity (STP)

In Eq.1, the two dynamical variables u and x implement the STP effect (Markram et al., 1998; Mongillo et al., 2008; Barak and Tsodyks, 2014; Miller et al., 2018). Specifically, $u(\theta, t)$ represents the release probability of neural transmitters, and $x(\theta, t)$ denotes the fraction of available neural transmitters. As illustrated in Fig. 5A, once neurons at θ generate spikes, $u(\theta, t)$ increases, resulting in the short-term facilitation (STF) effect; meanwhile, $x(\theta, t)$ decreases, resulting in the short-term depression (STD) effect. After spiking, $u(\theta, t)$ gradually returns to its baseline value $u(\theta, t) = 0$ with a time constant τ_f and $x(\theta, t)$ to its baseline value $x(\theta, t) = 1$ with a time constant τ_d. The dynamics of u and x are given by:

$$\frac{\partial u(\theta,t)}{\partial t} = \frac{-u(\theta,t)}{\tau_f} + U(1-u(\theta,t))r_E(\theta,t), \\ \frac{\partial x(\theta,t)}{\partial t} = \frac{1-x(\theta,t)}{\tau_d} - u(\theta,t)x(\theta,t)r_E(\theta,t), \tag{5}$$

where U is a constant controlling the increment of $u(\theta, t)$ due to neuronal firing ($r_E(\theta, t)$). Here we set the network with parameters that are compatible with experimental measurements in the prefrontal cortex (Wang et al., 2006). Moreover, following previous models (Mongillo et al., 2008), we set the synapses as STF-dominating, i.e., $\tau_f >> \tau_d$. As a result, after neuronal firing, the synaptic strength would be maintained at high values for a long time, which serves as the neural correlate to hold the information of the stimulus (Mongillo et al., 2008; Barak and Tsodyks, 2014; Trübutschek et al., 2017). Furthermore, due to STP, the instantaneous synaptic efficacy of neurons encoding θ at time t is given by $Ju(\theta, t)x(\theta, t)$, which holds the trace of the memorized item, i.e., the larger the value of $Ju(\theta, t)x(\theta, t)$, the higher the probability of the memorized item to be retrieved.

2.4.3. External inputs to CANNs

The network dynamics could be decomposed into three periods: the encoding period when WM items are sequentially loaded into the network, the retention period when the dynamic perturbation is applied to, and the recalling period when WM information is retrieved. The external inputs in the three periods are different:

During the encoding period, we set I_{ext} to be:

$$I_{ext}(\theta,t) = \begin{cases} a_{ext}(t)\cos\left(B_{ext}^{\text{loading}}\times(\theta-\theta_i)\right), & if B_{ext}^{\text{loading}}\times(\theta-\theta_i)\in[-\arccos(0),\arccos(0)], \\ 0, & else, \end{cases} \tag{6}$$

where θ_i ($i = 1$ & 2) represents the orientation of the i^{th} WM item to be loaded into the network. $a_{ext}(t) = A_{ext}^{\text{loading}}$, for $t \in [0, T_{1\text{oading}}]$, is the signal strength at time t, with T_{loading} as the duration of the to-be-memorized item and A_{ext}^{loading} as a constant. B_{ext}^{loading} is a parameter controlling the accuracy of the signal, i.e., the larger the value of B_{ext}^{loading} is, the higher the accuracy of the signal. Importantly, given the object-based characteristics in WM, i.e., automatic binding of color to orientation (Luck et al., 1997; Johnson et al., 2008; Hyun et al., 2009; Oberauer and Lin, 2017; Huang et al., 2018; Manohar et al., 2019), the flickering color probes during retention would carry partial information about the corresponding orientation features. Therefore, we model the flickering color probes during retention as the following external inputs:

$$I_{ext}(\theta,t)=\begin{cases} a^1_{\text{ext}}(t)\cos\left(B^{\text{probe}}_{ext}\times(\theta-\theta_1)\right) & if\, B^{\text{probe}}_{ext}\times(\theta-\theta_i)\in[-\arccos(0),\arccos(0)],i=1,2,\\ +a^2_{\text{ext}}(t)\cos\left(B^{\text{probe}}_{ext}\times(\theta-\theta_2)\right), & \\ 0, & else, \end{cases} \tag{7}$$

where θ_i (for $i=1$ & 2) denotes the orientation of i^{th} loaded memory item. $a^i_{\text{ext}}(t)$, for $t\in[0,T_{\text{probe}}]$, is the signal strength at time t. T_{probe} is the duration of the flickering probes. $B^{\text{probe}}_{\text{ext}}$ is a constant controlling the binding relevance of the probe signal. To reflect that the probe signal is stochastic, we set $a^i_{ext}(t)$ to be a random number in the range of $\left[0,A^{\text{probe}}_{ext}\right]$.

During the recalling period, subjects were given an instruction indicating which WM item to be retrieved. Here we model this instruction as an external input which triggers the retrieval of the corresponding WM item in the network. The external input is written as:

$$I_{ext}(\theta,t)=\begin{cases} a_{ext}(t)\cos\left(B^{\text{recalling}}_{ext}\times(\theta-\theta_i)\right)+\sigma_2\varsigma_2(\theta,t), & \\ 0, & \\ if\, B^{\text{recalling}}_{ext}\times(\theta-\theta_i)\in[-\arccos(0),\arccos(0)], & \\ else, & \end{cases} \tag{8}$$

where θ_i (for $i=1$ & 2) is the orientation of the i^{th} memory item, $a^i_{\text{ext}}(t)=A^{\text{recalling}}_{ext}$, for $t\in[0,T_{\text{recalling}}]$ is the strength of the recalling signal, with $T_{\text{recalling}}$ as the duration of the recalling period and $A^{\text{recalling}}_{ext}$ a constant. $B^{\text{recalling}}_{ext}$ is a constant controlling the accuracy of the recalling signal. $\varsigma_2(\theta,t)$ is a Gaussian white noise of zero mean and unit variance, and σ_2 is the noise strength.

Overall, during the encoding period, we set A^{loading}_{ext} and B^{loading}_{ext} to have large values, so that the network can generate responses at the corresponding locations, reflecting that the items are successfully loaded in the network. During the retention period, since the flickering probes carry only partial information about the memorized items, we set A^{probe}_{ext} and B^{probe}_{ext} to be smaller than A^{loading}_{ext} and B^{loading}_{ext} (not strong enough to reload the memory items), but sufficiently large to modulate synapse strengths via STP. During the recalling period, since the recalling instruction contained no stimulus feature information, we set $A^{\text{recalling}}_{ext}$ and $B^{\text{recalling}}_{ext}$ to be further smaller than A^{probe}_{ext} and B^{probe}_{ext}. Overall, the strengths and accuracies of external inputs in different periods satisfy the conditions, i.e., $B^{\text{loading}}_{ext}>B^{\text{probe}}_{ext}>B^{\text{recalling}}_{ext}$ and $A^{\text{loading}}_{ext}>>A^{\text{probe}}_{ext}>A^{\text{recalling}}_{ext}$.

2.5. Model simulation

2.5.1. The simulation protocol

The CANN + STP model was then used to simulate all the experimental conditions. Specifically, the orientations of the two colored bar signals were set to be the same as that in the behavioral experiments. They were loaded into the network sequentially, each lasting 1 s and the time interval between them was 0.5 s. This was stimulated by applying the external inputs $I_{ext}(\theta_1,t)$ and $I_{ext}(\theta_2,t)$ to the network sequentially. After 0.6~1.4 s, two different flickering color probes were presented to the network for 5 s. Finally, a recalling signal was presented to the network for 1 s to retrieve the instructed memory item. We mimicked different "dynamic perturbation" conditions by applying external inputs with corresponding characteristics similar to that in behavioral experiments (details as below).

2.5.2. External inputs for different "dynamic perturbation" conditions

The external inputs representing the flickering probe signals during retention were constructed as follows (i.e., $a_{ext}(t)$). Similar to the behavioral experiments, we first generated a noise sequence X1 of length T_{probe}/dt ($T_{\text{probe}}=5s$, denoting the duration of the maintaining period; dt = 0.01s, denoting the size of time bin) that are uniformly distributed in the range of $\left[0,A^{\text{probe}}_{ext}\right]$, and its spectrum was calculated. We then equalized the spectrum power of X1 at all frequencies to get a new sequence X2. Finally by scaling the mean of X2 to be $A^{\text{probe}}_{ext}/2$, we got a sequence X3. The three sequences and their variations were then used to construct the flickering probe signals under different conditions.

2.5.3. Baseline condition

By randomly shuffling the sequence X3, we got another sequence X4. We set X3 to be the sequence of the color probe strengths that are bound with the orientation θ_1, and X4 as the sequence of color probe strength bound with the orientation θ_2, i.e., X3 = Seq1 and X4 = Seq2, as used in behavioral experiments (Baseline condition, Fig. 6A).

2.5.4. Full-spectrum synchronization condition (F-sync)

We set X3 to be the sequence of the signal strengths for both color probes bound with θ_1 and θ_2, respectively, i.e., X3 = Seq1 = Seq2 as used in behavioral experiments (F-Sync condition; Fig. 6B).

2.5.5. Alpha-band synchronization condition (A-sync)

By filtering the sequence X2 within the alpha band (8−11 Hz), we obtained a sequence X2′. By scaling the mean of X2′ to be $A^{\text{probe}}_{ext}/2$, we got a new sequence X3′. We set X3′ to be the sequence of the strengths for both color probes bound with θ_1 and θ_2, respectively, i,e, X3′ = Seq1 = Seq2 as used in the behavioral experiments (A-Sync condition; Fig. 6C).

2.5.6. Same-order condition

As shown in Fig. 6D, we shifted the last 200 ms sequence of X3 into its front to get X4. We set X3 to be the sequence of the color probe strengths bound with θ_1, and X4 the sequence of the color probe strengths bound with θ_2, i.e., X3 = Seq1 and X4 = Seq2, as used in the behavioral experiments (Same-order condition; Fig. 6D).

2.5.7. Reversed-order condition

The same sequences were constructed as in the same-order condition, but we set X4 to be the sequence of the color probe strengths bound with θ_1, and X3 the sequence of the color probe strengths bound with θ_2, i.e., X4 = Seq1 and X3 = Seq2, as used in the behavioral experiments (Reversed-order condition; Fig. 6E).

2.5.8. Model predictions

The Same-order and Reversed-order conditions were simulated with varying time lags (10, 50, 90, 110, 130, 150, 170, 190, 210, 230, 270, 310, 350 ms).

2.5.9. Statistical analysis of model simulation results

The simulations were performed for each experimental condition, respectively. Six pairs of orientation angles with differences of ±17°, ±24°, ±38°, ±52°, ±66°, and ±80° were used. For each condition

(probe-absent, baseline, synchronization manipulation, same-order, reversed-order, etc.), there were 20 simulation runs and each run contained 200 simulation trials. The results of each run (as if the performance of one subject in the behavior experiments) were then calculated by averaging over the 200 trials within each run.

2.5.10. Decoding WM information during the recalling period

In the recalling period, we used the population vector to read out the orientation value retrieved from the network (Georgopoulos et al., 1982; Wu et al., 2002), expressed as:

$$\theta_{\text{decode}} = \frac{\int \theta \langle r_E(\theta,t)\rangle d\theta}{\int \langle r_E(\theta,t)\rangle d\theta}, \tag{9}$$

where $\langle r_E(\theta,t)\rangle$ is the averaged firing rate of neurons at θ during the recalling period.

Next, the same behavioral analysis and normalization procedure employed in behavioral experiments were used to quantify the memory retrieval performance of the network model.

2.5.11. The definition of two neuronal groups

To illustrate the competing network dynamics that mediates the observed experimental results, i.e., how the process is modulated via STP, we defined two neuronal groups. Group 1 are neurons whose preferred orientations are in the range of $|\theta - \theta_1| < 7^o$, and group 2 are those whose preferred orientations are in the range of $|\theta - \theta_2| < 7^o$. We then calculated the mean firing rate of neurons in each group, denoted as $\langle r_i \rangle$, for $i = 1,2$, and regarded it as the activity level of the corresponding neuronal group. To quantify the level of synaptic strength in each neuronal group, we calculated the mean synaptic efficacy of all neurons within each group, which is given by $\langle Ju_i \rangle = \langle Ju(\theta,t)\rangle$, where the average was performed over all neurons in the group i, for $i = 1, 2$, separately. Note that here synapses are STF-dominating with STD variable $x \approx 1$.

3. Results

3.1. "Dynamic perturbation" approach: temporal association of memory-related, task-irrelevant color probes during retention

Subjects performed a WM task in all the experiments. As shown in Fig. 1A, participants viewed two serially presented bar stimuli and needed to memorize the orientations of the two bars ("encoding period"). After 5 s memory retention, participants rotated a horizontal bar to the memorized 1st or 2nd orientation based on the instruction ("recalling period"). Importantly, the two bar stimuli, in addition to having distinct to-be-memorized orientations, also had different colors. During the "maintaining period", subjects performed a central fixation task, while two discs that had color feature of either the 1st or the 2nd memorized bars (1st memory-related probe, WM1; 2nd memory-related probe, WM2) were displayed at left or right side of the fixation. The color (red, green, or blue) and the spatial locations (left or right) of the two probes (WM1 and WM2) were balanced across trials.

Crucially, throughout the 5 s retention period, the luminance of the two color probes (WM1 and WM2) was continuously modulated by two 5 s temporal series respectively (i.e., Seq1 for WM1, Seq2 for WM2), the relationship of which varied in specific way as detailed below. Under Baseline condition, the two luminance sequences (Seq1 and Seq2) were two independently generated random time series, and thus no temporal association between items were introduced (Fig. 1B, left). Importantly, we developed two types of manipulation that aim to apply temporally correlated perturbation – Synchronization (Fig. 1B) and Order Reversal (Fig. 3A) – to the two flickering color probes. For "synchronization" condition, the probes were modulated by the same luminance sequence,

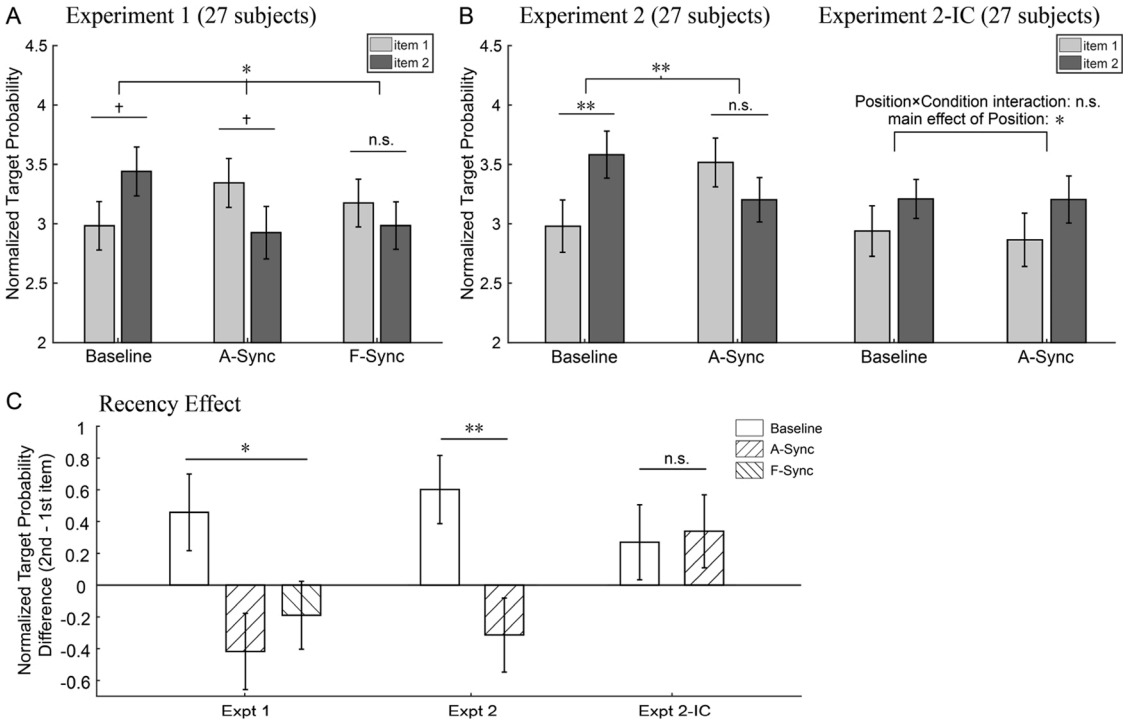


Fig. 2. "Temporal synchronization" manipulation disrupts recency effect.
(A) Experiment 1 results (N = 27). Grand averaged (mean±SEM) normalized target probability for the 1st (light grey) and 2nd (dark grey) orientation for Baseline, A-Sync, and F-Sync conditions. (B) Experiment 2 (N = 27) and Experiment 2-IC (irrelevant color probes) (N = 27) results. Grand averaged (mean±SEM) normalized target probability for the 1st (light grey) and 2nd (dark grey) orientation for Baseline and A-Sync conditions, in Experiment 2 (left) and Experiment 2-IC (right). (C) Grand averaged (mean±SEM) recency effect (2nd minus 1st normalized target probability) for Baseline (white box), A-Sync (rightward checked box), and F-Sync (leftward checked box) conditions in Experiment 1, Experiment 2, and Experiment 2-IC. Noting the synchronization-induced recency disruption in Experiment 1 and 2, but not in Experiment 2-IC. (**: p<=0.01, *: 0.01<p<=0.05, +: 0.05<p<=0.1).

either the same random temporal sequence (F-Sync; Fig. 1B, right), or the same alpha-band filtered random time series (A-Sync; Fig. 1B, middle). Including A-Sync condition is indeed motivated by previous findings revealing prominent alpha-band (8−11 Hz) activations in WM (Klimesch, 1999; Jensen, 2002). We next examined whether and how the dynamic manipulation impacts the subsequent memory performance, particularly the recency effect, a behavioral index reflecting the relative memory strength of WM items (Burgess and Hitch, 1999; Baddeley, 2012). For Baseline condition when no temporal association is introduced, we would expect intact recency effect, as shown previously (Gorgoraptis et al., 2011; Huang et al., 2018). In contrast, conditions that bring temporally correlated perturbation such as Synchronization or Order Reversal would presumably interfere with neural dynamics and impact (i.e. eliminate or even reverse) the recency effect.

It is noteworthy that all the luminance sequences were generated anew in each trial. To quantify the memory performance for the 1st and the 2nd orientation, we used Paul Bays' model (Bays et al., 2009), a largely used approach that could be implemented through an open-source toolbox, to calculated the target probability. We then converted the target probability results using an empirical logit transformation (Cox and Snell, 1989; de Smith, 2010; Stevens et al., 2016) to make it normally distributed (i.e., normalized target probability) for further statistical analyses.

3.2. *"Synchronization" manipulation disrupts recency effect (Experiment 1–2)*

Twenty-seven subjects participated in Experiment 1 (Fig. 2A). First, consistent with our hypothesis, the Baseline condition showed recency effect (2nd > 1st item; paired t-test, $t_{(26)} = -1.895$, $p = 0.069$, Cohen's $d = -0.431$), as typically observed when subjects load a list of items into WM (Gorgoraptis et al., 2011; Huang et al., 2018), confirming that temporally uncorrelated perturbation would not affect the relative memory strength of WM items. A control experiment that employed the same paradigm but without flickering color probes during retention showed the same recency effect as the Baseline condition (Supplementary Fig. 1). In contrast, when Seq1 and Seq2 were temporally synchronized with each other (A-Sync and F-Sync) by sharing the same luminance sequence, the recency effect was largely disrupted (paired t-test, A-Sync: 2nd < 1st item, $t_{(26)} = 1.745$, $p = 0.093$, Cohen's $d = 0.377$; F-Sync: $t_{(26)} = 0.890$, $p = 0.382$, Cohen's $d = 0.183$), suggesting their relative memory strengths were successfully modified. Two-way repeated ANOVA (Position * Sync) across the three conditions revealed significant interaction effect ($F_{(2,52)} = 3.231$, $p = 0.048$, $\eta_p^2 = 0.111$), and nonsignificant main effect for either position ($F_{(1,26)} = 0.224$, $p = 0.640$, $\eta_p^2 = 0.009$) or condition ($F_{(2,52)} = 0.307$, $p = 0.737$, $\eta_p^2 = 0.012$), indicating that this dynamic perturbation specifically disrupted recency effect without affecting the overall memory performance. Post-hoc comparisons further confirmed the disruption of recency effect for both A-Sync and F-Sync compared to Baseline (2-way repeated ANOVA; A-Sync and Baseline: interaction effect, $F_{(1,26)} = 5.015$, $p = 0.034$, $\eta_p^2 = 0.162$; F-Sync and Baseline: interaction effect, $F_{(1,26)} = 3.454$, $p = 0.074$, $\eta_p^2 = 0.117$). Moreover, the two synchronization conditions did not show significant difference, indicating their similar recency disruption effects (2-way repeated ANOVA; A-Sync and F-Sync: interaction effect, $F_{(1,26)} = 0.477$, $p = 0.496$, $\eta_p^2 = 0.018$).

To further confirm the findings, we ran a separate group of subjects (N = 27) for only the A-Sync and Baseline conditions (Experiment 2). As shown in Fig. 2B, the results were well consistent with Experiment 1, i.e., the Baseline condition showed recency effect (paired t-test, $t_{(26)} = -2.804$, $p = 0.009$, Cohen's $d = -0.553$), whereas the A-Sync condition did not (paired t-test, $t_{(26)} = 1.349$, $p = 0.189$, Cohen's $d = 0.308$). Two-way repeated ANOVA (Positions * Sync) revealed significant interaction effect ($F_{(1,26)} = 8.067$, $p = 0.009$, $\eta_p^2 = 0.237$) but nonsignificant main effect for either position ($F_{(1,26)} = 0.855$, $p = 0.364$, $\eta_p^2 = 0.032$) or sync condition ($F_{(1,26)} = 0.214$, $p = 0.647$, $\eta_p^2 = 0.008$), further supporting the synchronization-induced disruption effect.

Moreover, to examine whether the synchronization-induced disruption effect was dependent on memory-related probes, we ran another experiment by presenting memory-irrelevant color probes (Experiment

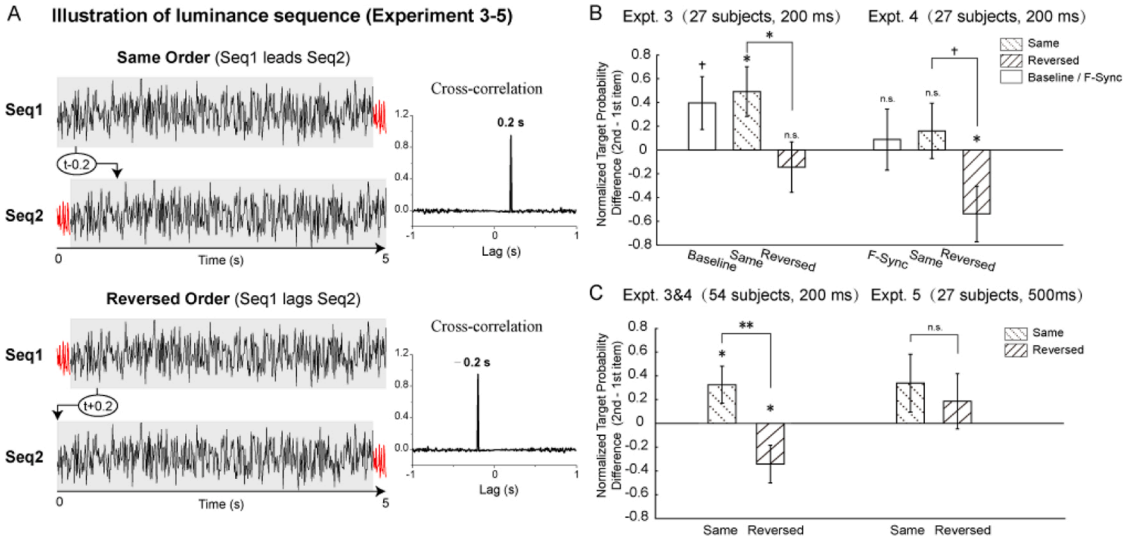


Fig. 3. "Order Reversal" manipulation disrupts and even reverses recency effect.
(A) Experiment 3–5 employed the same experimental paradigm as Experiment 1, except that the luminance sequences (Seq1 and Seq2) were modulated in a different way. Seq1 was a temporally shifted version of Seq2, either leading ("Same-order", upper panel) or lagging Seq2 ("Reversed-order", lower panel) by 200 ms, i.e., the final 200 ms segment of the leading sequence was shifted to the beginning of the lagged sequence (denoted in red). Note that Seq1 and Seq2 occurred simultaneously. Right: the Seq1-Seq2 cross-correlation coefficient as a function of temporal lag (-1 to 1 s). The luminance sequences were generated anew in each trial. (B) Two separate subject groups participated in Experiment 3 (N = 27; Baseline, Same-order, Reversed-order) and Experiment 4 (N = 27; F-Sync, Same-order, Reversed-order) respectively. Grand averaged (mean ± SEM) recency effect (2nd minus 1st normalized target probability) for all conditions in Experiment 3 and Experiment 4. Noting the decrease in recency effect for Reversed-order than Same-order conditions in both experiments. (C) Left: pooled recency effects (N = 54; mean ± SEM) for Same-order and Reversed-order conditions across Experiment 3 and 4. Right: Experiment 5 (500 ms temporal lag) results. Grand average recency effects (N = 27; mean ± SEM) for Same-order and Reversed-order conditions (**: p< = 0.01, *: 0.01 < p< = 0.05, +: 0.05 < p<=0.1).

2-IC, N = 27). As shown in Fig. 2B (right), the Baseline and Synchronization conditions now (Experiment 2-IC) did not show significant difference in recency effect (two-way repeated ANOVA, interaction effect, $F_{(1,26)} = 0.035$, $p = 0.853$, $\eta_p^2 = 0.001$, main effect for position, $F_{(1,26)} = 4.588$, $p = 0.042$, $\eta_p^2 = 0.150$), different from Experiment 2 (Experiment 2 vs. Experiment 2-IC, Two-sample two-way repeated ANOVA, Experiment*condition*position: $F_{(1,52)} = 4.041$, $p = 0.050$ $\eta_p^2 = 0.072$). Thus, memory-irrelevant color features failed to introduce temporal associations between WM items and in turn could not impact WM performance.

As summarized in Fig. 2C, temporal synchronization of the memory-related probes during retention essentially disturbed the recency effect (see similar but less significant trend when spatial locations were synchronized, Experiment 7, Supplementary Fig. 2). It is noteworthy that the color probes were completely task-irrelevant in terms of both memory task (memorizing orientations) and attentional task (central fixation task). Moreover, this disruption effect was not accompanied by any overall memory performance changes (i.e., no main effect of condition), indicating that the manipulation does not interrupt memory retention at a general level but indeed specifically modulates the relative memory strength of WM items (also see precision measurements in Supplementary Fig. 3).

3.3. "Order Reversal" manipulation reverses recency effect (Experiment 3–5)

After establishing the efficacy of the "synchronization" manipulation, we next examined the "Order Reversal" manipulation. Specifically, the same experimental paradigm was used (see Fig. 1A), except that the luminance sequences of the two color discs were modulated in a different way. As shown in Fig. 3A, Seq1 was a temporally shifted version of Seq2 which was also a randomly generated white noise sequence in each trial, either leading (Same-order, upper panel) or lagging Seq2 (Reversed-order, lower panel) by 200 ms. In other words, although Seq1 and Seq 2 were randomly generated sequences in each trials, they kept certain temporal relationship, i.e., Same-order (Seq1 leading Seq2) or Reversed-order (Seq1 lagging Seq2). Moreover, note that Seq1 and Seq2 occurred simultaneously rather than asynchronously, i.e., the final 200 ms segment of the leading sequence was appended to the beginning of the lagged sequence (denoted in red, Fig. 3A). The selection of 200 ms was motivated by the time constant in the STP neural model (Mongillo et al., 2008; Mi et al., 2017; Trübutschek et al., 2017) as well as previous neurophysiological studies revealing a central role of theta-band rhythm in organizing memory replay (Lisman and Idiart, 1995; Buszaki, 2002; Bahramisharif et al., 2018; Huang et al., 2018; Herweg et al., 2020). Two separate groups of subjects participated in Experiment 3 (N = 27; Same-order, Reversed--order, Baseline) and Experiment 4 (N = 27; Same-order, Reversed--order, F-Sync), respectively. Importantly, the two conditions (Same-order vs. Reversed-order) could not be distinguished in visual perception at all.

First, as shown in Fig. 3B, consistent with previous results (Experiment 1–2), recency effect persisted for Baseline condition (paired t-test, $t_{(26)} = -1.778$, $p = 0.087$, Cohen's $d = -0.388$; left panel) and vanished for Synchronization condition (paired t-test, $t_{(26)} = -0.341$, $p = 0.736$, Cohen's $d = -0.095$; right panel). Most interestingly, as shown in Fig. 3B, the recency effect significantly decreased for Reversed-order compared to Same-order condition in both Experiments (paired t-test, Expt. 3: $t_{(26)} = 2.344$, $p = 0.027$, Cohen's $d = 0.581$; Expt. 4: $t_{(26)} = 1.839$, $p = 0.077$, Cohen's $d = 0.577$). Two-sample two-way repeated ANOVA (Experiment*Position*Order) revealed significant Position*Order interaction effect ($F_{(1,52)} = 8.161$, $p = 0.006$, $\eta_p^2 = 0.136$), supporting that the "Order Reversal" manipulation successfully altered the WM behavior. We further combined the Same-order and Reversed-order conditions across Experiment 3 and 4 since the two conditions were exactly the same and derived from two independent subject groups. As shown in Fig. 3C (left panel), the pooled results (N = 54) showed clear decrease and even reversal in recency effect for Reversed-order condition compared to Same-order condition (paired t-test, $t_{(53)} = 2.884$ $p = 0.006$, Cohen's $d = 0.576$). Notably, the only difference between the two conditions was the temporal relationship between the two luminance sequences (i.e., Seq1 leads Seq2, Seq2 leads Seq1) which was even indistinguishable in visual perception, yet they led to distinct recency effect. This strongly advocates that the dynamic perturbation through different temporal directions largely interferes with the WM neural representations.

Finally, we assessed whether introducing a temporal lag outside of the STP time window would also influence WM behavior. Instead of 200 ms, Experiment 5 (N = 27) employed a 500 ms temporal lag between Seq1 and Seq2. As shown in the right panel of Fig. 3C, the Same-order and Reversed-order conditions did not show significant recency difference (paired t-test, $t_{(26)} = 0.393$, $p = 0.697$, Cohen's $d = 0.123$). Thus, the dynamic perturbation should be applied within the STP-related temporal window (e.g., short-term depression time constant) to efficiently influence WM.

3.4. Meta-analysis

Finally, to increase the detecting sensitivity and examine the overall effects, we performed a meta-analysis (Comprehensive Meta-Analysis Software, CMA) across seven experiments, including Expt. 1–5 in the main text as well as Expt. 6 (probe-absent experiment) and Expt. 7 (spatial location experiment) in Supplementary materials. As shown in Fig. 4, the Baseline condition (Experiment 1, 2, 3, 5, 6, 7) showed significant recency effect ($p = 2e^{-6}$), whereas the Synchronization conditions (Experiment 1, 2, 4, 7) did not ($p = 0.395$). Furthermore, Baseline and Synchronization conditions (Experiment 1, 2, 7) revealed significant difference ($p = 4e^{-5}$). Note that the meta-analysis on target probability without normalization (Supplementary Fig. 6) and precision measurements (Supplementary Fig. 5) showed similar pattern.

3.5. CANN-STP WM computational model

After confirming the effectiveness of the dynamic perturbation approach on WM manipulation, we further developed a computational model that combines CANN and STP principles to unveil the underlying neural mechanism. CANN has been widely used to model orientation tuning (Ben-Yishai et al., 1995; Blumenfeld et al., 2006) and WM-related networks (Brody et al., 2003; Parthasarathy et al., 2019; Seeholzer et al., 2019) in neural systems. STP, referring to a phenomenon that synaptic efficacy changes over time with neuronal activity, is also commonly used to implement WM functions (Mongillo et al., 2008; Barak and Tsodyks, 2014; Mi et al., 2017; Miller et al., 2018). As shown in Fig. 5A, in the CANN, neurons are aligned on a ring according to their preferred orientations and are connected in a translation-invariant manner. Furthermore, neurons on the ring are reciprocally connected to a common inhibitory neuron pool mediating the competition among neuronal groups (Kim et al., 2017), such that only a single bump-shaped neural response could be generated in the network at any moment. The STP effect is modeled by two dynamical variables, u and x, representing the release probability of neural transmitters (leading to short-term facilitation, STF) and the fraction of resources available after neurotransmitter depletion (leading to short-term depression, STD), respectively (Markram et al., 1998; Mongillo et al., 2008). The instantaneous synaptic efficacy of neuron at θ is written as $Ju(\theta,t)x(\theta,t)$, with J as the maximum value of the synaptic efficacy. Notably, $Ju(\theta,t)x(\theta,t)$ determines the memory strength encoded by neuron at θ and holds the trace of the memorized item (Mongillo et al., 2008), i.e., if $Ju(\theta,t)x(\theta,t)$ is large enough, an external cue can trigger the recall of the memorized item. Thus, the larger the value of $Ju(\theta,t)x(\theta,t)$, the higher the accuracy of the memory to be retrieved. Moreover, the synaptic efficacy varies in the time scales of u and x, which are denoted as τ_f and τ_d, respectively.

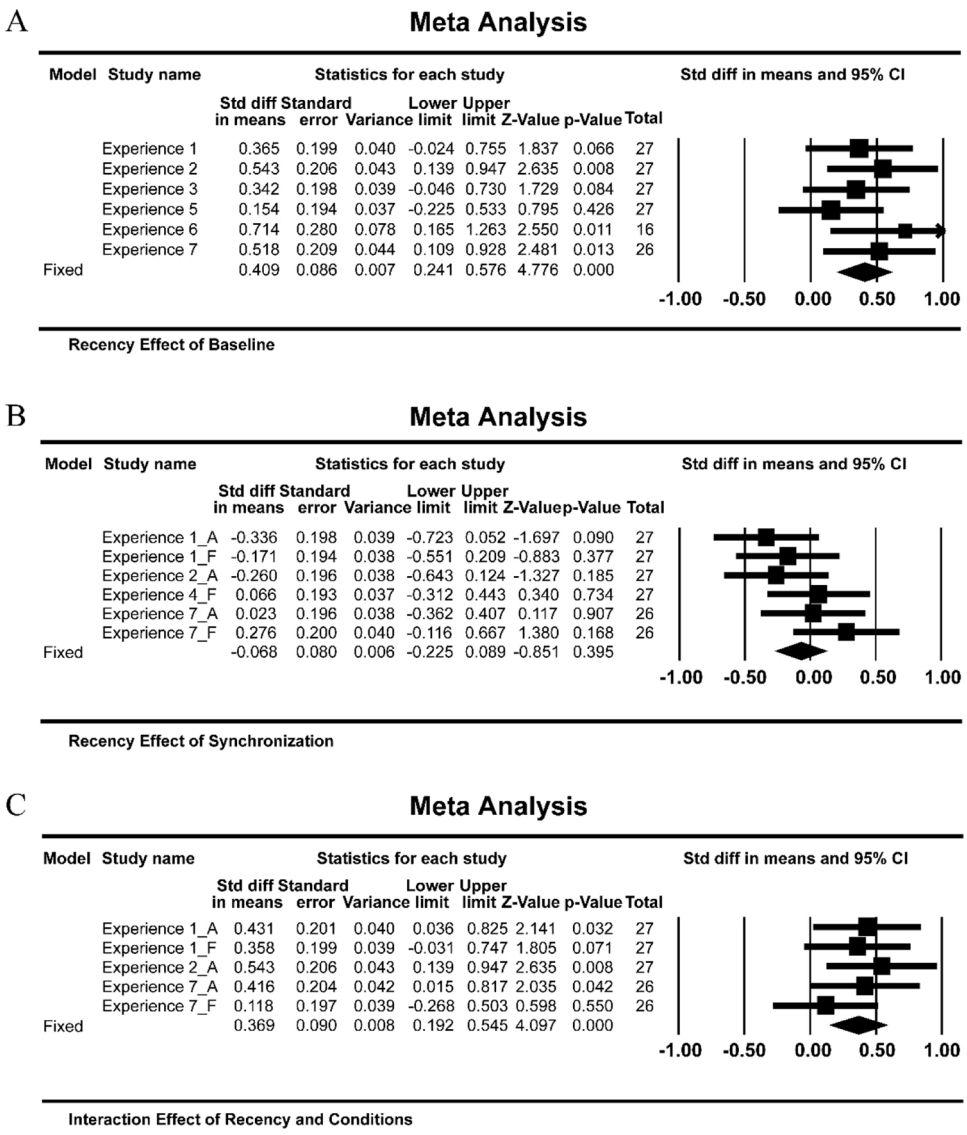


Fig. 4. Meta-analysis of normalized target probability (Experiment 1-7).
Meta-analysis on recency effect across all 7 experiments (Expt. 1–5 in main text, Expt. 6–7 in Supplementary Materials). (A) Significant recency effect for Baseline condition (across Experiment 1, 2, 3, 5, 6, 7). (B) Non-significant recency effect for Synchronization condition (across Experiment 1, 2, 4, 6, 7). (C) Significant difference in recency effect between Baseline and Synchronization conditions (across Experiment 1, 2, 7).

Given the STF dominance in PFC, i.e., $\tau_f \gg \tau_d$, we can effectively approximate the STD variable x to be a constant, and only focus on the approximated synaptic efficacies $Ju(\theta, t)$.

We next used the CANN-STP WM model to elucidate the intrinsic dynamics of neural activities and synaptic efficacies in WM network during the maintaining period as well as the memory recalling performance, for each experimental condition (200 trials in each run, 20 runs for each condition), respectively. Taking one condition as example (Fig. 5B), two orientation features θ_1 and θ_2 are first sequentially loaded into the CANN and evoke two strong transient response called population spikes (PSs) hereafter, at θ_1 and θ_2, respectively. For the convenience of description, we cluster those neurons with their preferred orientations in the range of $|\theta - \theta_i| \leq \Delta$, for $i = 1$ & 2, into two neuronal groups, where $\Delta = 7^o$ is the half of the neuronal connection width. The averaged neural activities and the averaged synaptic strengths of neurons in these two groups reflect the memory strengths they encode, which are denoted as $\langle r_i \rangle$ and $\langle Ju_i \rangle$ (for $i = 1$ & 2), respectively (see Methods). Notably, during retention when stimuli are not present anymore, the synaptic efficacies of the two memory-related neuronal groups would not fall to baseline immediately and would rather remain at high values for a long time and decay slowly as a result of STP. Finally, during recall, an external input representing the recall instruction is presented to the CANN and triggers the retrieval dynamics of the network (see Methods), i.e., correct retrieval is achieved if the corresponding neuronal group is successfully elicited.

The critical manipulation – flickering color probes – are simulated as continuous inputs containing partial information of the memorized items (see Methods), which are endowed with the corresponding temporal characteristics as in each experiment condition, to the corresponding neuronal groups during retention (Fig. 6, left column, I_{ext}). These temporally correlated noisy stimulations perturb the activities of two neuronal groups to generate a series of PSs (analogy to the coherent resonance process (Hu et al., 1993; Pikovsky and Kurths, 1997; Lee et al., 1998)), which trigger a new round of competition between neuronal groups through global inhibition, and in return modulates the relative value of the synaptic efficacies of two neuronal groups (Fig. 6, middle column), ultimately determining the memory strength of WM items and the recalling performance (Fig. 6, right column).

3.6. Model simulations

First, we examined whether our model could reproduce the typical recency effect shown in previous studies (Burgess and Hitch, 1999; Gorgoraptis et al., 2011; Baddeley, 2012; Jones and Oberauer, 2013) as well as our own (Probe-absent experiment, Supplementary Fig. 1), when there is no probe presented during retention. As shown in Fig. 5B, the

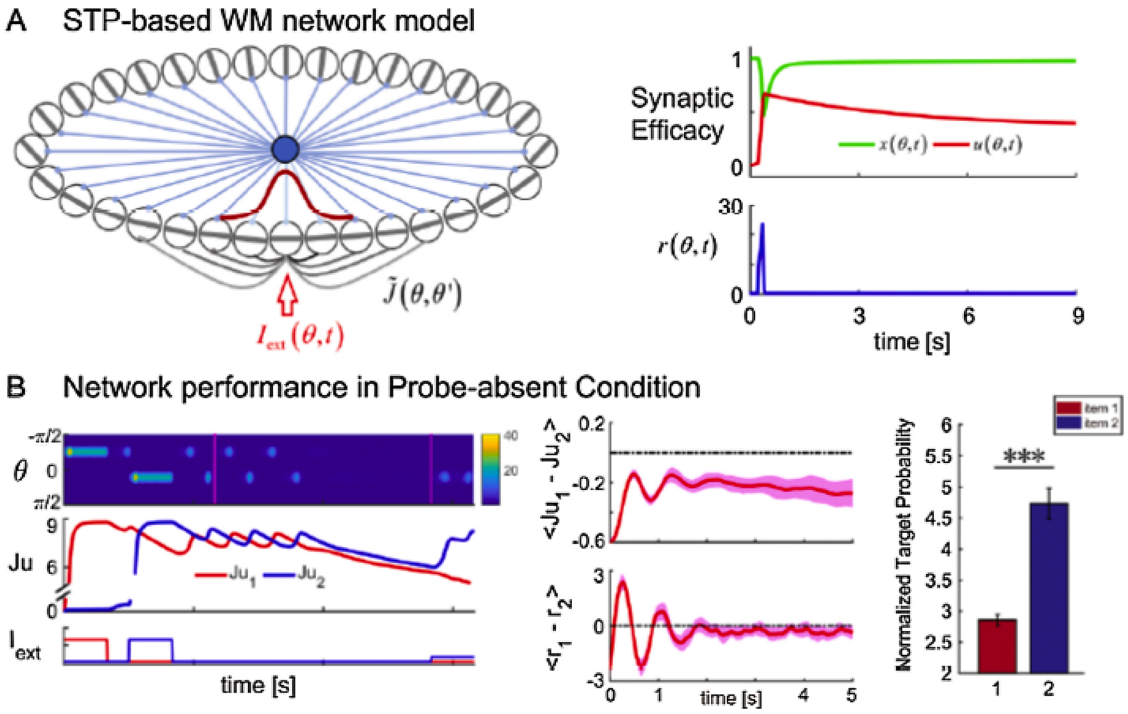


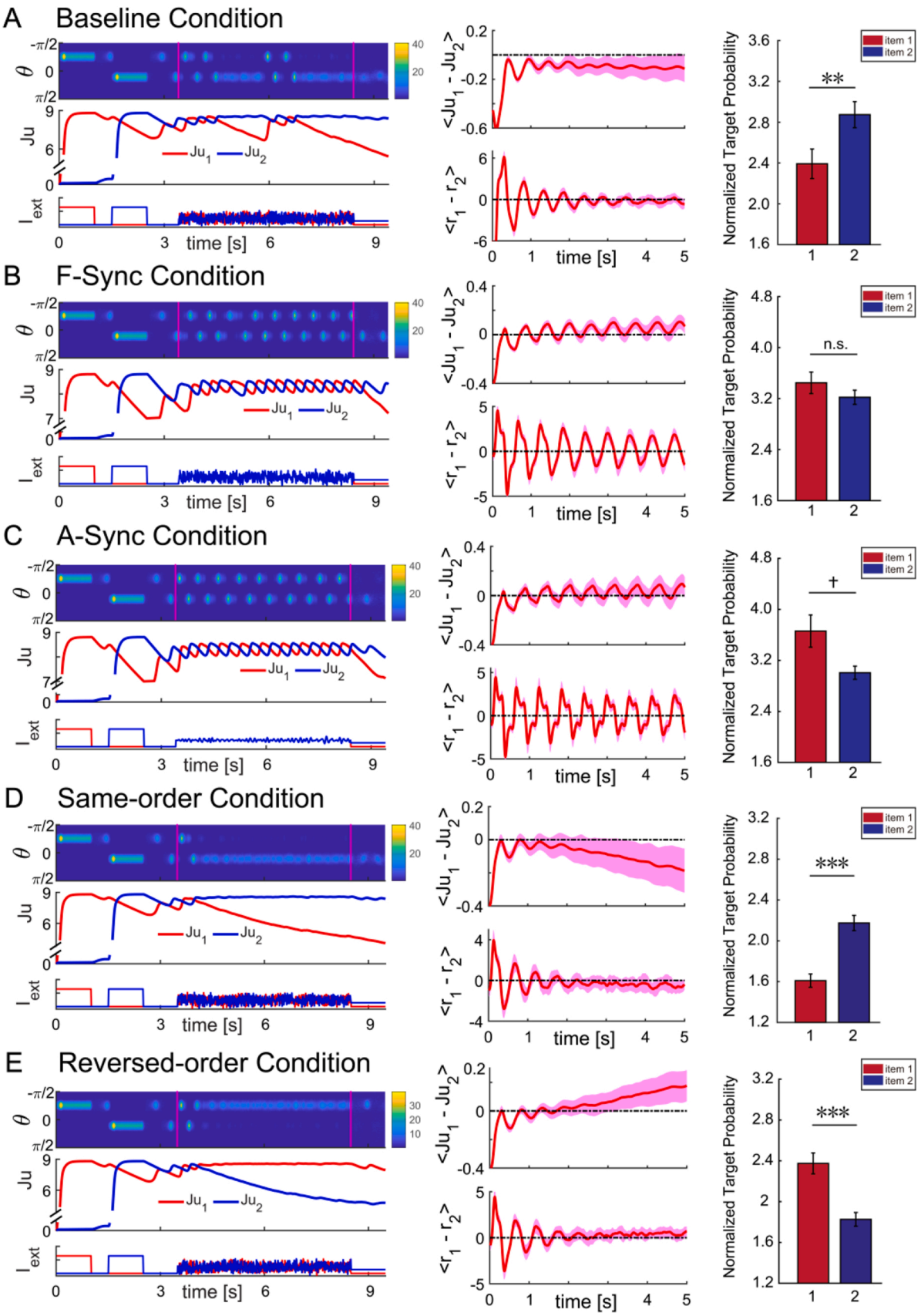
Fig. 5. CANN-STP WM network model and simulation of probe-absent condition.
(A) CANN and STP. Left: Neurons in the CANN are aligned on a ring according to their preferred orientations $\theta \in (-\pi/2, \pi/2]$, and they are connected in a translation-invariant manner in the form of $\tilde{J}(\theta, \theta')$. All neurons are reciprocally connected to a common inhibitory neuronal pool (blue solid circle). In response to an external input $I_{ext}(\theta, t)$, the network generates a bump response (red curve) at the corresponding location according to stimulus strength. Right: Illustration of STP principle. In response to a transient external input, the neuron encoding orientation at θ generates a transient response (firing rate, $r(\theta, t)$), which induces changes in both the neurotransmitter release probability $u(\theta, t)$ (red) and the fraction of available neurotransmitter $x(\theta, t)$ (green), representing STF and STD effect, respectively. The synaptic efficacy of neuron at θ is $Ju(\theta, t)x(\theta, t)$, with J as the maximum value of synaptic efficacy, will remain at high values for a while, due to the larger time constant for STF than STD (i.e., $\tau_d = 0.214s$, $\tau_f = 4.9s$). (B) Simulation results for probe-absent condition. Left: Individual trial simulation example. (Top) Two orientations (1st and 2nd) at $\theta_1 = -\pi/4$ and $\theta_2 = -\pi/4 + 66^o$ (as examples) are loaded into the network sequentially and trigger population spikes in the corresponding neuronal groups (1st and 2nd neuronal group). During retention (the period between two pink lines), the two neuronal groups still show response due to STP, but fall to baseline gradually. In the recalling period, a noisy signal representing the recall instruction, e.g., to recall the 2nd orientation, is presented and elicits the firing of the 2nd neuronal group. (Middle) the temporal course of synaptic efficacies of the 1st (red) and 2nd (blue) neuronal groups throughout the trial. (Bottom) the temporal course of the 1st (red) and 2nd (blue) loaded external inputs (i.e., $a_{ext}(t)$) to the CANN throughout the trial. Middle: Grand averaged (across 20 simulation runs with each having 200 simulation trials) synaptic efficacy difference $\langle Ju_1 - Ju_2 \rangle$ (top, mean ± SEM) and firing rate difference $\langle r_1 - r_2 \rangle$ (bottom, mean ± SEM) between the 1st and 2nd neuronal groups over retention. Right: Grand averaged (mean ± SEM, across 20 simulation runs) recalling performance characterized by the normalize target probability (the same analysis for behavioral data) for the 1st (red) and 2nd (blue) orientation. (***: p< = 0.001.). Parameters and mathematical details are given in Methods and Supplementary Materials.

loaded items evoke PSs, and the network activity decays to silence rapidly afterwards. Interestingly, during retention when inputs are removed, the synaptic efficacies of the two neuronal groups persist at high values as a result of STP (Fig. 5B, left column). Since the two WM items are loaded at different moments, the averaged synaptic efficacy of the early presented item (1st WM item) would be generally smaller than that of the recently presented item (2nd WM item) (Fig. 5B, middle column). Consequently, given their difference in synaptic efficacy developed during retention, the 2nd WM item would have larger probability to be successfully retrieved than the 1st WM item (Fig. 5B, right column), resulting in the recency effect (paired t-test, $t_{(19)}$ = -6.747, p = $2e^{-6}$, Cohen's d =-2.228).

When probe signals are applied to the network during retention (i.e., dynamic perturbation), their ultimate effects are to perturb the activities of neurons, which subsequently modulate the neuronal synapse strengths via STP and consequently manipulate the memory retrieval performances. Since probe signals are noisy, their effects on the network dynamics fluctuate over time. Overall, dynamic perturbation could induce three different neuronal responses during retention: 1) oscillation, in which the two neuronal groups win the competition (having a larger activity) alternately; 2) group 1 constantly wins, in which neuronal group 1 is at a higher activity level than group 2 and on average $\langle Ju_1 \rangle > \langle Ju_2 \rangle$; 3) group 2 constantly wins, in which the neuronal group 2 is at a higher activity level than group 1 and on average $\langle Ju_2 \rangle > \langle Ju_1 \rangle$. Under different "dynamic perturbation" conditions, the chances for the network falling into the three states are also different. Consistent with the synaptic theory of WM (Mi et al., 2017; Trübutschek et al., 2017), when $\langle Ju_2 \rangle > \langle Ju_1 \rangle$ holds at the end of retention, the network recalling performance would show the recency effect and otherwise the primary effect.

To mimic the Baseline condition, we applied two independent random inputs simultaneously to the two neuronal groups during retention. Since $\langle Ju_2 \rangle > \langle Ju_1 \rangle$ at the beginning of the retention period, the 2nd neuronal group has a higher chance to generate PSs, which in turn increases the synaptic efficacy (Fig. 6A, left column). Meanwhile, due to the competition mediated by global inhibition, the averaged response and synaptic efficacy of the 1st neuronal group are continuously suppressed. As a result, the relative value of the synaptic efficacies between the two neuronal groups is retained (Fig. 6A, middle column), leading to the recency effect in recalling performance (Fig. 6A, right column; paired t-test, $t_{(19)}$ = -3.007, p = 0.007, Cohen's d =-0.788), consistent with the empirical findings (Fig. 2).

For Synchronization conditions, instead of two independent random inputs, the same input series (white noise for F-Sync and alpha-band series for A-Sync) were applied to the two neuronal groups. Since the input signals are synchronized, two neuronal groups continue to



(caption on next page)

Fig. 6. Model simulations for different dynamic manipulation conditions.
Simulation results for (A) Baseline condition, (B) F-Synchronization condition, (C) A-Synchronization condition, (D) Same-order condition and (E) Reversed-order condition. Left: Individual trial examples in different conditions (a typical example in each condition is shown). (Top) Two orientations (1st at $\theta_1 = -\pi/4$ and 2nd at $\theta_2 = -\pi/4 + 66^o$ as examples) are loaded sequentially and the two corresponding neuronal groups generate PSs. The flickering probe signals trigger different neural activity patterns during maintaining period, and the 2nd item is cued in the recalling period. (Middle) the temporal course of synaptic efficacies of the 1st (red) and 2nd (blue) neuronal groups through the trial. (Bottom) the temporal course of the 1st (red) and 2nd (blue) external inputs (i.e., $a_{ext}(t)$) to the CANN through the trial. Middle: Grand averaged (mean ± SEM) (across 20 simulation runs with each having 200 simulation trials) synaptic efficacy difference $\langle Ju_1 - Ju_2 \rangle$ (top) and firing rate difference $\langle r_1 - r_2 \rangle$ (bottom) between the 1st and 2nd neuronal groups over retention. Right: Grand averaged (mean ± SEM, across 20 simulation runs) recalling performance characterized by the normalized target probability for the 1 st (red) and 2nd (blue) orientation. Parameters and mathematical details of probe stimulations are given in Supplementary Materials. (***: p< = 0.001,**: 0.001 < p< = 0.01, *: 0.01 < p< = 0.05, +: 0.05 < p<=0.1).

generate PSs in the same order as loaded, and their synaptic efficacies fluctuate with neural activities according to STP (Fig. 6BC, left and middle columns), which gradually destroys the initial relative value of the synaptic efficacies between the two neuronal groups. Consequently, during recalling (Fig. 6BC, right column), the recency effect is largely disrupted (F-Sync: paired t-test, $t_{(19)} = 1.124$, $p = 0.275$, Cohen's $d = 0.356$; A-Sync: paired t-test, $t_{(19)} = 2.084$, $p = 0.051$, Cohen's $d = 0.757$), also in line with the experimental observations (Fig. 2).

For the "Order Reversal" simulation, the input to the neuronal group that encodes the 1st orientation would either lead (Same-order condition) or lag (Reversed-order condition) the input to the group encoding the 2nd orientation. Intuitively, for the Same-order condition, the noisy input (which contains partial information of the memorized items) would load the WM items in the same order as they are originally loaded into the network. Hence, the Same-order manipulation enhances the relative value of the synaptic efficacies of two neuronal groups, i.e., enlarging the difference $\langle Ju_2 \rangle > \langle Ju_1 \rangle$ and $\langle r_2 \rangle > \langle r_1 \rangle$ (Fig. 6D, middle column), finally leading to the recency effect (Fig. 6D, right column; paired t-test, $t_{(19)} = -6.654$, $p = 2e^{-6}$, Cohen's $d = -1.804$). In contrast, for Reversed-order condition, the noisy input in effect is to load the WM items in the reversed order as they are originally loaded into the network. Therefore, the "Order Reversal" manipulation keeps disturbing the relative value of synaptic efficacies and gradually flips their relationship m, i.e., it becomes $\langle Ju_1 \rangle > \langle Ju_2 \rangle$ and $\langle r_1 \rangle > \langle r_2 \rangle$ (Fig. 6E, left and middle columns). This process thus would cause the reversal of the recency effect to primacy effect (Fig. 6E, right column; paired t-test, $t_{(19)} = 4.361$, $p = 3e^{-4}$, Cohen's $d = 1.419$), again consistent with the experimental results (Fig. 3).

Furthermore, we confirmed that the same results still hold when the separation between two orientations ($|\theta_1 - \theta_2|$) is set to be any value from 3° to 90° (Supplementary Fig. 7). Thus, the orientation difference would not affect the manipulation effect, suggesting that the direct interaction between neuronal groups might be not critical. Taking into this account and to further elucidate the key elements that contribute to the "dynamic perturbation", we built a simplified version of the model. Specifically, we considered only two neuron clusters, and neurons in the same cluster are connected with each other via STP without between-cluster connections. The two neuron clusters are further connected to a common inhibitory neuron pool, which mediates their competition. The simplified model largely replicated the simulation results in the full model (see Supplementary Materials).

Taken together, the CANN-STP WM model successfully reproduces all the empirical results (Figs. 5 and 6), demonstrating that the changes on the ratio of synaptic efficacies of neuronal groups, which are caused by the dynamic perturbation introduced by flickering probes, result in the manipulation of relative memory strength of WM items.

3.7. Model predictions

Our experiments only tested "Order Reversal" manipulation using 200 ms and 500 ms time lags (Fig. 3). We further used the model to predict how systematic changes in the correlation time window between the two probe sequences, in the Same- or Reversed-order conditions, would affect the recency effect (Fig. 7). First, when the time lag is close to zero, both the Same- and Reversed-order conditions show disrupted recency effect and tend to shift to a nonsignificant trend of primary effect, similar to "Synchronization" conditions. Second, when the time lag is too large, both conditions show the recency effect, similar to "Baseline" condition. Most interestingly, when the time lag is within [110 ms, 210 ms], a range close to the temporal constant of STD, i.e., $(0.5 \sim 1) * \tau_d$, the recalling performance for the Same- or Reversed-order conditions begin to display divergence, i.e., recency and

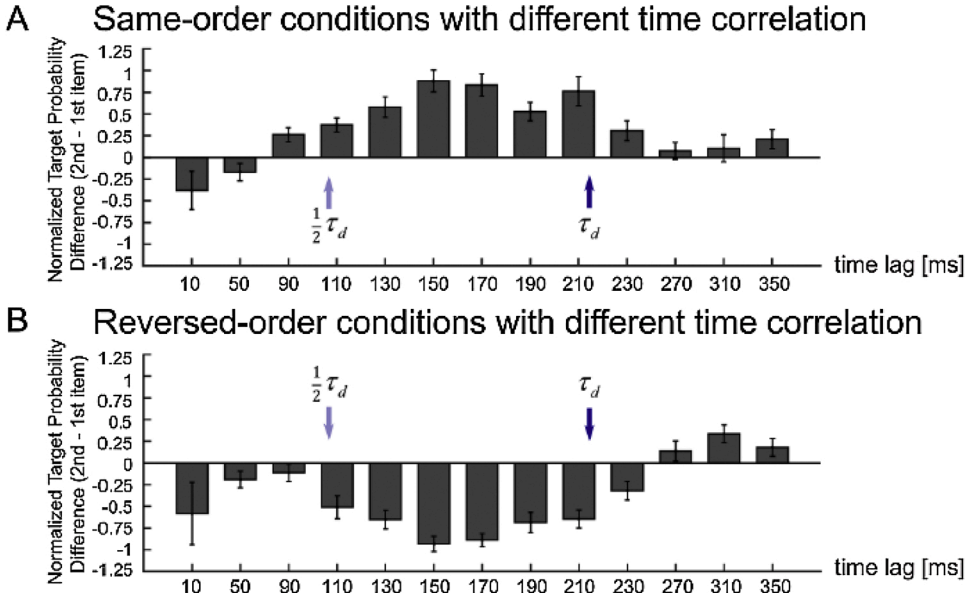


Fig. 7. Model predictions.
Grand averaged (mean±SEM, across 20 simulation runs with each having 200 simulation trials) normalized target probability difference between the 1st and 2nd orientation (2nd minus 1st) at varying time lags (10–350 ms) for (A) Same-order and (B) Reversed-order conditions. τ_d (STD time constant) and $0.5\tau_d$ are marked by dark blue arrows (214 ms) and light blue arrows (107 ms), respectively. Note that when the time lag is close to zero (similar to Synchronization condition), both Same- or Reversed-order conditions show disruption in recency effect, whereas when the time lag is large (similar to Baseline condition), both conditions show recency effect. Only when the correlation time is in the range of [110 ms, 210 ms], the two conditions show divergence, i.e., recency effect for Same-order condition and primacy effect for Reversed-order condition. Parameters are given in Supplementary Materials.

primacy effects for the same-order and reversed-order conditions, respectively. We also varied the value of the STD time constant τ_d and obtained similar result (see Supplementary Materials). Thus, the temporal window for efficient WM manipulation is mainly determined by the STD time constant.

4. Discussion

We developed a novel "dynamic perturbation" approach to manipulate the relative memory strength of a list of items held in human WM, by using temporally-correlated luminance sequences to interfering with their neural dynamics during WM retention. Six experiments confirm the effectiveness of this method in multi-item WM manipulation. A computational model incorporating CANN and STP principles fully reproduces the empirical findings. Importantly, the model demonstrates that this "dynamic perturbation" induces changes in the relative synaptic efficacies of WM items and eventually causes the modulation of their relative memory strength. Taken together, the temporal dynamics of WM neural representation, determined by both active-state and STP-based hidden-state, plays a fundamental role in mediating the storage of multiple items. Our results also provide a promising, purely bottom-up approach to manipulate human WM behavior.

It has been long posited that temporary maintenance of information relies on a refreshing process (Awh et al., 1998; Baddeley, 2003; Camos et al., 2018; Oberauer, 2019). Animal recordings and human neuroimaging studies show memory-related neural reactivations during retention (Skaggs and McNaughton, 1996; Compte, 2000; Wang, 2001; Foster and Wilson, 2006; Siegel et al., 2009; Michelmann et al., 2016; Huang et al., 2018; Liu et al., 2019; Schuck and Niv, 2019), supporting an active-state WM view. Meanwhile, neural response are also known to induce changes in the synaptic efficacy via STP mechanisms (Mongillo et al., 2008; Barak and Tsodyks, 2014; Mi et al., 2017; Trübutschek et al., 2017). Specifically, the synaptic efficacy, given its slow temporal properties, could remain for a while to preserve memory information without relying on sustained neural firing, namely a hidden-state WM view (Stokes, 2015; Wolff et al., 2017; Miller et al., 2018). In fact, the two theories could be well integrated into a framework where information is maintained in the synaptic weights between reactivations through STP principles, and the dynamic interplay between the active-state and hidden-state over the delay period ultimately leads to memory maintenance. Interestingly, sustained firing and hidden-state have recently been posited to be involved in manipulation and passive storage of WM information, respectively (Trübutschek et al., 2017; Masse et al., 2019, 2020). Hence, manipulation of stored information might require reinstatement of the hidden-state representation into activated state (Stokes, 2015; Trübutschek et al., 2019; Masse et al., 2020). Here, by applying certain temporal perturbation patterns to interfere with the active-state of the WM network, we modulate the hidden-state via STP principles and alter the memory strength.

Holding a list of items in WM elicits an item-by-item sequential reactivation pattern (Siegel et al., 2009; Heusser et al., 2016; Michelmann et al., 2016; Bahramisharif et al., 2018; Huang et al., 2018; Liu et al., 2019; Schuck and Niv, 2019), which arises from an ongoing competition between items through lateral inhibition and shared inhibitory inputs (Fino and Yuste, 2011; Kim et al., 2017; Mi et al., 2017). During this rivalry process, each item takes turns to reactivate and accordingly strengthen its respective synaptic efficacy which slowly builds up and decays over the retention interval. Therefore, the dynamic competition between items and the resulted memory trace held in the synaptic weights would eventually determine their relative memory strength. Here, the two WM items are perturbed in a temporally correlated manner so that their dynamic competition is modulated, initiating changes in their relative memory strength. In contrast, temporally uncorrelated perturbation, i.e., baseline condition, given its lack of stable influence on the competition between WM items, shows no modulation effects. Moreover, it is worth noting that the luminance sequences used in our experiment were generated anew in each trial, and the manipulation thus does not depend on specific sequences but on the temporal associations between them.

We build a computational model that combines CANN and STP to explore the dynamics of the WM neural network– both neural firing rate (active state) and synaptic efficacy (hidden state) – and how the dynamics leads to memory strength changes. In this model, STF and STD, as two forms of STP, play different functions, i.e., STF serves to hold memory information and STD controls the time window for effective manipulation. The dynamic sequences brought by the flickering probes continuously perturb the network state and lead to changes in synaptic efficacies and memory performance. When items are independently disturbed in time (Baseline), their competition are not affected and so the recency effect keeps intact, whereas synchronous perturbance would instead change the competition, which in turn modulates their relative synaptic weights and disrupts recency effect (Synchronization). In the case of Same-order condition, the perturbation strengthens the relative synaptic efficacies for the two consecutively loaded WM items, resulting in the recency effect. In contrast, under the Reversed-order condition, the flipped perturbation pattern indeed reverses the relative synaptic efficacies, i.e., $\langle Ju_2 \rangle > \langle Ju_1 \rangle$ becomes $\langle Ju_2 \rangle < \langle Ju_1 \rangle$, leading to a shift from recency to primacy effect. Notably, in both cases, the temporal correlation of the two probe sequences needs to fall into the STD temporal range for the manipulation to be effective. How to understand this phenomenon? Previous work (Mi et al., 2017), using STP-based mechanism to explain WM capacity, demonstrates that the time separation between consecutive WM is close to the STD time constant. Therefore, when two temporal sequences are applied to perturb the order of memorized items, their correlation time needs to fall into the STD time constant, so that the changes they induce to synaptic efficacy can effectively disturb the retrieval performances of two memory items. Together, our model sheds light on a challenging WM issue and provides promising ways to efficiently manipulate WM from brain's perspective.

Previous empirical and theoretical works have largely specified how multiple features of memorized items are bound together and retained in WM network (e.g., Manohar et al., 2019; Oberauer and Lin, 2017; Huang et al., 2018; Hyun et al., 2009; Johnson et al., 2008; Luck et al., 1997; Manohar et al., 2019; Oberauer and Lin, 2017). Our model is thus based on a simplification that the flickering color probes act directly on neurons encoding corresponding orientation features, i.e., not explicitly incorporating feature binding process. Furthermore, the model particularly aims to examine the neural mechanism for "dynamic perturbation" approach, i.e., how temporally correlated inputs modulate the relative memory strengths of WM items, regardless of the associated luminance sequence, thereby not largely dependent on detailed characteristics of the feature-binding process.

Previous studies, by using behavioral or non-invasive stimulation approaches in the encoding or delay period, have also successfully modulated the recalling performance (Sauseng et al., 2009; Rose et al., 2016; Clouter et al., 2017; Hanslmayr et al., 2019; Martorell et al., 2019). Here, benefiting from using the time-resolved luminance sequences to disturb the WM network dynamics, our study largely differs from previous interference findings in two aspects. First, previous studies are mainly focused on the modulation of general memory performance while our study aims at particularly manipulating the relative memory strength of WM items (i.e., recency effect). Second, by using temporally correlated luminance sequences, our approach could precisely modulate the ongoing temporal relationship between items (e.g., Synchronization, Order reversal within 200 ms or 500 ms). Generally speaking, our "Synchronization manipulation" condition shares similar motivations as associative memory paradigms, during which pairing two stimuli would lead to changes in their association in memory (Shohamy and Wagner, 2008; Duncan et al., 2012; Gershman et al., 2017). Meanwhile, instead of long-term memory, our manipulation modulates working memory on a trial-by-trial basis. In addition, we adopted a purely bottom-up approach by modulating the luminance of

task-irrelevant color probes, thus largely different from previous interference methods. Together, our model sheds light on a challenging WM issue and provides promising ways to efficiently manipulate WM from brain's perspective. Future neuroimaging approaches are also needed to enable the direct assessment of ongoing synaptic strength changes in the WM network that contributes to the "dynamic perturbation".

Finally, recent studies advocate a close connection between WM and attention, and WM has been proposed to be an internal attentional process that shares similar properties and operation principles as attention (Chun et al., 2011; D'Esposito and Postle, 2015; Oberauer, 2019). For instance, the recency effect could be readily explained in terms of attentional gain, such that the lately presented item enters the focus of attention (FOA) and obtains more attention (Oztekin et al., 2010; Oberauer and Lin, 2017). In general, WM and attention are likely to be mediated via largely overlapping neural networks and similar mechanism, making them hard to be fully disentangled. In fact, the memory-related reactivation might also reflect attentional modulation of neural response that would affect subsequent attentional performance (Soto et al., 2005; Olivers et al., 2006; Mallett and Lewis-Peacock, 2018). Similarly, attention-induced response might leave memory traces to affect following memory behavior (Cowan et al., 2005; Rose et al., 2016; Camos et al., 2018; deBettencourt et al., 2019). Moreover, attention has been shown to sample different objects sequentially and rhythmically, also resembling the dynamic reactivation pattern in WM (Jensen et al., 2014; Jia et al., 2017; Fiebelkorn and Kastner, 2019). Therefore, the "dynamic perturbation" might impact the attentional process as well and thus constitutes a promising way to also manipulate attention, planning and decision making.

5. Conclusions

Taken together, we develop a novel "dynamic perturbation" behavioral approach to manipulate the relative memory strength of a list of WM items, by applying temporally correlated luminance sequences to interfere with their neural dynamics during WM retention. A computational model shows that this manipulation modifies the synaptic efficacies of WM items through STP principles, eventually leading to changes in their relative memory strengths. Our results support the causal role of temporal dynamics of neural network in mediating multi-item WM and offer a promising, non-invasive approach to manipulate multi-item WM.

Author contributions

Jiaqi Li, Qiaoli Huang, and Huan Luo designed the experiment. Jiaqi Li and Qiaoli Huang performed the experiments and analyzed the data. Qiming Han contributed to technique issues. Yuanyuan Mi built the computational model and performed simulations. Jiaqi Li, Qiaoli Huang, Yuanyuan Mi, and Huan Luo wrote the paper.

Funding sources

This work was supported by the National Natural Science Foundation of China (31930052 to H.L., 31771146, 11734004 to Y.Y. Mi); Beijing Municipal Science & Technology Commission (Z181100001518002) to H.L.; Beijing Nova Program (Z181100006218118) to Y. Y. Mi; Guangdong Province (2018B030338001) to Y.Y. Mi; Fundamental Research Funds for the Central Universities (2020CDJQY-A073) to Y.Y. Mi.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

We thank Dr. David Poeppel, Dr. Fang Fang, and Dr. Nai Ding for their helpful comments, as well as amounts of anonymous reviews to our previous submission.

Appendix A. The Peer Review Overview and Supplementary data

The Peer Review Overview and Supplementary data associated with this article can be found in the online version, at doi:https://doi.org/10.1016/j.pneurobio.2021.102023.

References

Awh, E., Jonides, J., Reuter-lorenz, P.A., 1998. Rehearsal in spatial working memory. J. Exp. Psychol. Hum. Percept. Perform. 24, 780–790.
Baddeley, A., 2003. Working memory: looking back and looking forward. Nat. Rev. Neurosci. 4, 829–839.
Baddeley, A., 2012. Working memory: theories, models, and controversies. Annu. Rev. Psychol. 63, 1–29.
Bahramisharif, A., Jensen, O., Jacobs, J., Lisman, J., 2018. Serial representation of items during working memory maintenance at letter-selective cortical sites. PLoS Biol. 16, e2003805. Available at: http://dx.plos.org/10.1371/journal.pbio.2003805.
Barak, O., Tsodyks, M., 2014. Working models of working memory. Curr. Opin. Neurobiol. 25, 20–24. https://doi.org/10.1016/j.conb.2013.10.008. Available at:
Bays, P.M., Catalao, R.F.G., Husain, M., 2009. The precision of visual working memory is set by allocation of a shared resource. J. Vis. 9, 7. Available at: http://jov.arvojournals.org/Article.aspx?doi=10.1167/9.10.7.
Ben-Yishai, R., Lev Bar-Or, R., Sompolinsky, H., 1995. Theory of orientation tuning in visual. Proc. Natl. Acad. Sci. U. S. A. 92, 3844–3848.
Blumenfeld, B., Bibitchkov, D., Tsodyks, M., 2006. Neural network model of the primary visual cortex: from functional architecture to lateral connectivity and back. J. Comput. Neurosci. 20, 219–241.
Brody, C.D., Romo, R., Kepecs, A., 2003. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. Curr. Opin. Neurobiol. 13, 204–211.
Burgess, N., Hitch, G.J., 1999. Memory for serial order: a network model of the phonological loop and its timing. Psychol. Rev. 106, 551–581.
Buszaki, G., 2002. Theta oscillations in the hippocampus. Neuron 33, 325–340. Available at: papers3://publication/uuid/90B140FB-AC03-4F7C-9968-EF84D84A3009.
Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., Vergauwe, E., 2018. What is attentional refreshing in working memory? Ann. N. Y. Acad. Sci. 1424, 19–32.
Chun, M.M., Golomb, J.D., Turk-Browne, N.B., 2011. A taxonomy of external and internal attention. Annu. Rev. Psychol. 62, 73–101.
Clouter, A., Shapiro, K.L., Hanslmayr, S., 2017. Theta phase synchronization is the glue that binds human associative memory. Curr. Biol. 27, 3143–3148. https://doi.org/10.1016/j.cub.2017.09.001 e6 Available at:
Compte, A., 2000. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb. Cortex 10, 910–923.
Compte, A., Brunel, N., Goldman-rakic, P.S., Wang, X., 2000. Dynamics Underlying Spatial Working Memory in a Cortical Network Model E, pp. 910–923.
Cowan, N., 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav. Brain Sci. 24, 87–114.
Cowan, N., Elliott, E.M., Saults, S.J., Morey, C.C., Mattox, S., Hismjatullina, A., Conway, A.R.A., 2005. On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. Cogn. Psychol. 51, 42–100.
Cox, D.R., Snell, E.J., 1989. Analysis of Binary Data. CRC Press, 32.
D'Esposito, M., Postle, B.R., 2015. The cognitive neuroscience of working memory. Annu. Rev. Psychol. 66, 115–142.
de Smith, M.J., 2010. Statistical Analysis Handbook. A Comprehensive Handbook of Statistical Concepts, Techniques and Software Tools. StatsrefCom.
deBettencourt, M.T., Keene, P.A., Awh, E., Vogel, E.K., 2019. Real-time triggering reveals concurrent lapses of attention and working memory. Nat. Hum. Behav. 3, 808–816. https://doi.org/10.1038/s41562-019-0606-6. Available at:
Dell, G.S., Burger, L.K., Svec, W.R., 1997. Language production and serial order: a functional analysis and a model. Psychol. Rev. 104, 123–147.
Doyon, J., Penhune, V., Ungerleider, L.G., 2003. Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. Neuropsychologia 41, 252–262.
Duncan, K., Sadanand, A., Davachi, L., 2012. Memory's Penumbra: episodic memory decisions induce lingering mnemonic biases. Science (80-) 337, 485–487.
Erdfelder, E., FAul, F., Buchner, A., Lang, A.G., 2009. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav. Res. Methods 41, 1149–1160.
Fiebelkorn, I.C., Kastner, S., 2019. A rhythmic theory of attention. Trends Cogn. Sci. 23, 87–101. https://doi.org/10.1016/j.tics.2018.11.009. Available at:
Fiebig, X.F., Lansner, X.A., 2017. A spiking working memory model based on Hebbian short-term potentiation. J. Neurosci. 37, 83–96.
Fino, E., Yuste, R., 2011. Dense inhibitory connectivity in neocortex. Neuron 69, 1188–1203. https://doi.org/10.1016/j.neuron.2011.02.025. Available at:
Foster, D.J., Wilson, M.A., 2006. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature 440, 680–683.
Fung, C.C.A., Wong, K.Y.M., Wu, S., 2012. Delay compensation with dynamical synapses. Advances in Neural Information Processing Systems.

Fusi, S., 2008. A quiescent working memory. Science (80-) 319, 1495–1496.
Georgopoulos, A.P., Kalaska, J.F., Caminiti, R., Massey, J.T., 1982. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J. Neurosci. 2 (11), 1527–1537.
Gershman, S.J., Monfils, M.-H., Norman, K.A., Niv, Y., 2017. The computational nature of memory modification. Elife 6, 1–41.
Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–517. https://doi.org/10.1038/nn.3063. Available at:
Gorgoraptis, N., Catalao, R.F.G., Bays, P.M., Husain, M., 2011. Dynamic updating of working memory resources for visual objects. J. Neurosci. 31, 8502–8511. https://doi.org/10.1523/JNEUROSCI.0208-11.2011. Available at:
Hanslmayr, S., Axmacher, N., Inman, C.S., 2019. Modulating human memory via entrainment of brain oscillations. Trends Neurosci. 42, 485–499. https://doi.org/10.1016/j.tins.2019.04.004. Available at:
Herweg, N.A., Solomon, E.A., Kahana, M.J., 2020. Theta oscillations in human memory. Trends Cogn. Sci. 24, 208–227. https://doi.org/10.1016/j.tics.2019.12.006. Available at:
Heusser, A.C., Poeppel, D., Ezzyat, Y., Davachi, L., 2016. Episodic sequence memory is supported by a theta-gamma phase code. Nat. Neurosci. 19, 1374–1380.
Horn, D., Opher, I., 1996. Temporal segmentation in a neural dynamic system. Neural Comput. 8, 373–389.
Hu, G., Ditzinger, T., Ning, C.Z., Haken, H., 1993. Stochastic resonance without external periodic force. Phys. Rev. Lett. 71, 807–810.
Huang, Q., Jia, J., Han, Q., Luo, H., 2018. Fast-backward replay of sequentially memorized items in humans. Elife 7, 1–21.
Hyun, Jseok, Woodman, G.F., Vogel, E.K., Hollingworth, A., Luck, S.J., 2009. The comparison of visual working memory representations with perceptual inputs. J. Exp. Psychol. Hum. Percept. Perform. 35, 1140–1160.
Jensen, O., 2002. Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. Cereb. Cortex 12, 877–882.
Jensen, O., Gips, B., Bergmann, T.O., Bonnefond, M., 2014. Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. Trends Neurosci. 37, 357–369. https://doi.org/10.1016/j.tins.2014.04.001. Available at:
Jia, J., Liu, L., Fang, F., Luo, H., 2017. Sequential sampling of visual objects during sustained attention. PLoS Biol. 15, 1–20. https://doi.org/10.1371/journal.pbio.2001903. Available at:
Johnson, J.S., Hollingworth, A., Luck, S.J., 2008. The role of attention in the maintenance of feature bindings in visual short-term memory. J. Exp. Psychol. Hum. Percept. Perform. 34, 41–55.
Jones, T., Oberauer, K., 2013. Serial-position effects for items and relations in short-term memory. Memory 21, 347–365.
Kim, S.S., Rouault, H., Druckmann, S., Jayaraman, V., 2017. Ring attractor dynamics in the Drosophila central brain. Science (80-) 356, 849–853.
Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev. 29, 169–195. Available at: http://www.sciencedirect.com/science/article/pii/S0165017398000563.
Lee, S.G., Neiman, A., Kim, S., 1998. Coherence resonance in a Hodgkin-Huxley neuron. Phys. Rev. E – Stat. Phys. Plasmas Fluids Relat. Interdisc. Top. 57, 3292–3297.
Lisman, J.E., Idiart, M.A.P., 1995. Storage of 7 ± 2 short-term memories in oscillatory subcycles. Science (80-) 267, 1512–1515.
Liu, T., Becker, M.W., 2013. Serial consolidation of orientation information into visual short-term memory. Psychol. Sci. 24, 1044–1050.
Liu, Y., Dolan, R.J., Kurth-Nelson, Z., Behrens, T.E.J., 2019. Human replay spontaneously reorganizes experience. Cell 178, 640–652.
Luck, S., Vogel, J., Edward, K., 1997. The capacity of visual working memory for features and conjuctions. Nature 390, 279–281.
Lundqvist, M., Rose, J., Herman, P., Brincat, S.L.L., Buschman, T.J.J., Miller, E.K.K., 2016. Gamma and beta bursts underlie working memory. Neuron 90, 152–164.
Mallett, R., Lewis-Peacock, J.A., 2018. Behavioral decoding of working memory items inside and outside the focus of attention. Ann. N. Y. Acad. Sci. 1424, 256–267.
Manohar, S.G., Zokaei, N., Fallon, S.J., Vogels, T.P., Husain, M., 2019. Neural mechanisms of attending to items in working memory. Neurosci. Biobehav. Rev. 101, 1–12. https://doi.org/10.1016/j.neubiorev.2019.03.017. Available at:
Markram, H., Wang, Y., Tsodyks, M., 1998. Differential signaling via the same axon of neocortical pyramidal neurons. Proc. Natl. Acad. Sci. U. S. A. 95, 5323–5328.
Martorell, A.J., Paulson, A.L., Suk, H.J., Abdurrob, F., Drummond, G.T., Guan, W., Young, J.Z., Kim, D.N.W., Kritskiy, O., Barker, S.J., Mangena, V., Prince, S.M., Brown, E.N., Chung, K., Boyden, E.S., Singer, A.C., Tsai, L.H., 2019. Multi-sensory gamma stimulation ameliorates Alzheimer's-associated pathology and improves cognition. Cell 177, 256–271. https://doi.org/10.1016/j.cell.2019.02.014 e22 Available at:
Masse, N.Y., Yang, G.R., Song, H.F., Wang, X.J., Freedman, D.J., 2019. Circuit mechanisms for the maintenance and manipulation of information in working memory. Nat. Neurosci. 22, 1159–1167. https://doi.org/10.1038/s41593-019-0414-3. Available at:
Masse, N.Y., Rosen, M.C., Freedman, D.J., 2020. Reevaluating the role of persistent neural activity in short-term memory. Trends Cogn. Sci. 24, 242–258. https://doi.org/10.1016/j.tics.2019.12.014. Available at:
Mi, Y., Lin, X., Wu, S., 2016. Neural computations in a dynamical system with multiple time scales. Front. Comput. Neurosci. 10, 96.
Mi, Y., Katkov, M., Tsodyks, M., 2017. Synaptic correlates of working memory capacity. Neuron 93, 323–330. https://doi.org/10.1016/j.neuron.2016.12.004. Available at:
Michelmann, S., Bowman, H., Hanslmayr, S., 2016. The temporal signature of memories: identification of a general mechanism for dynamic memory replay in humans. PLoS Biol. 14, 1–27.
Miller, E.K., Lundqvist, M., Bastos, A.M., 2018. Working memory 2.0. Neuron 100 (2), 463–475.
Mongillo, G., Barak, O., Tsodyks, M., 2008. Synaptic theory of working memory. Science (80-) 319, 1543–1546.
Oberauer, K., 2019. Working memory and attention – a conceptual analysis and review. J. Cogn. 2, 1.
Oberauer, K., Lin, H.Y., 2017. An interference model of visual working memory. Psychol. Rev. 124, 21–59.
Olivers, C.N.L., Meijer, F., Theeuwes, J., 2006. Feature-based memory-driven attentional capture: visual working memory content affects visual attention. J. Exp. Psychol. Hum. Percept. Perform. 32, 1243–1265.
Oztekin, I., Davachi, L., Mcelree, B., 2010. Are representations in working memory distinct from representations in long-term memory? Neural evidence in support of a single store. Psychol. Sci. 21, 1123–1133.
Parthasarathy, A., Tang, C., Herikstad, R., Cheong, L.F., Yen, S.C., Libedinsky, C., 2019. Time-invariant working memory representations in the presence of code-morphing in the lateral prefrontal cortex. Nat. Commun. 10, 1–11. https://doi.org/10.1038/s41467-019-12841-y. Available at:
Pikovsky, A.S., Kurths, J., 1997. Coherence resonance in a noise-driven excitable system. Phys. Rev. Lett. 78, 775–778.
Raffone, A., Wolters, G., 2001. A cortical mechanism for binding in visual working memory. J. Cogn. Neurosci. 13, 766–785.
Romani, S., Tsodyks, M., 2015. Short-term plasticity based network model of place cells dynamics. Hippocampus 25, 94–105.
Rose, N.S., LaRocque, J.J., Riggall, A.C., Gosseries, O., Starrett, M.J., Meyering, E.E., Postle, B.R., 2016. Reactivation of latent working memories with transcranial magnetic stimulation. Science (80-) 354, 1136–1139.
Sauseng, P., Klimesch, W., Heise, K.F., Gruber, W.R., Holz, E., Karim, A.A., Glennon, M., Gerloff, C., Birbaumer, N., Hummel, F.C., 2009. Brain oscillatory substrates of visual short-term memory capacity. Curr. Biol. 19, 1846–1852. https://doi.org/10.1016/j.cub.2009.08.062. Available at:
Schuck, N.W., Niv, Y., 2019. Sequential replay of nonspatial task states in the human hippocampus. Science 80-, 364.
Seeholzer, A., Deger, M., Gerstner, W., 2019. Stability of Working Memory in Continuous Attractor Networks under the Control of Shortterm Plasticity. Available at:. https://doi.org/10.1371/journal.pcbi.1006928.
Shohamy, D., Wagner, A.D., 2008. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron 60, 378–389. https://doi.org/10.1016/j.neuron.2008.09.023. Available at:
Siegel, M., Warden, M.R., Miller, E.K., 2009. Phase-dependent neuronal coding of objects in short-term memory. Proc. Natl. Acad. Sci. U. S. A. 106, 21341–21346. https://doi.org/10.1073/pnas.0908193106. Available at:
Skaggs, W.E., McNaughton, B.L., 1996. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. Science (80-) 271, 1870–1873.
Soto, D., Heinke, D., Humphreys, G.W., Blanco, M.J., 2005. Early, involuntary top-down guidance of attention from working memory. J. Exp. Psychol. Hum. Percept. Perform. 31, 248–261.
Stevens, S., Valderas, J.M., Doran, T., Perera, R., Kontopantelis, E., 2016. Analysing indicators of performance, satisfaction, or safety using empirical logit transformation. BMJ 352, i1114.
Stokes, M.G., 2015. "Activity-silent" working memory in prefrontal cortex: a dynamic coding framework. Trends Cogn. Sci. 19, 394–405. https://doi.org/10.1016/j.tics.2015.05.004. Available at:
Trübutschek, D., Marti, S., Ojeda, A., King, J.R., Mi, Y., Tsodyks, M., Dehaene, S., 2017. A theory of working memory without consciousness or sustained activity. Elife 6, 1–29.
Trübutschek, D., Marti, S., Ueberschär, H., Dehaene, S., 2019. Probing the limits of activity-silent non-conscious working memory. Proc. Natl. Acad. Sci. U. S. A. 116, 14358–14367.
Van Ede, F., Chekroud, S.R., Stokes, M.G., Nobre, A.C., 2018. Decoding the influence of anticipatory states on visual perception in the presence of temporal distractors. Nat. Commun. 9 https://doi.org/10.1038/s41467-018-03960-z. Available at:
Wang, X.-J., 2001. Synaptic reverberations underlying mnemonic persistent activity. Trends Neurosci. 24, 455–463.
Wang, Y., Markram, H., Goodman, P.H., Berger, T.K., Ma, J., Goldman-Rakic, P.S., 2006. Heterogeneity in the pyramidal network of the medial prefrontal cortex. Nat. Neurosci. 9, 534–542.
Wolff, M.J., Jochim, J., Akyürek, E.G., Stokes, M.G., 2017. Dynamic hidden states underlying working-memory-guided behavior. Nat. Neurosci. 20, 864–871.
Wu, S., Amari, S.I., Nakahara, H., 2002. Population coding and decoding in a neural field: a computational study. Neural Comput. 14 (5), 999–1026.
Wu, S., Wong, K.Y.M., Fung, C.C.A., Mi, Y., Zhang, W., 2016. Continuous attractor neural networks: candidate of a canonical model for neural information representation. F1000Research 5, 1–10.