Behavioral/Systems/Cognitive

From Distributed Resources to Limited Slots in Multiple-Item Working Memory: A Spiking Network Model with Normalization

Ziqiang Wei,[1,3,4] **Xiao-Jing Wang,**[2] **and Da-Hui Wang**[1,2]

[1]Department of Systems Science and National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China, [2]Department of Neurobiology and Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut 06510, [3]The Solomon H. Snyder Department of Neuroscience, The Johns Hopkins University, Baltimore, Maryland 21205, and [4]Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147

Recent behavioral studies have given rise to two contrasting models for limited working memory capacity: a "discrete-slot" model in which memory items are stored in a limited number of slots, and a "shared-resource" model in which the neural representation of items is distributed across a limited pool of resources. To elucidate the underlying neural processes, we investigated a continuous network model for working memory of an analog feature. Our model network fundamentally operates with a shared resource mechanism, and stimuli in cue arrays are encoded by a distributed neural population. On the other hand, the network dynamics and performance are also consistent with the discrete-slot model, because multiple objects are maintained by distinct localized population persistent activity patterns (bump attractors). We identified two phenomena of recurrent circuit dynamics that give rise to limited working memory capacity. As the working memory load increases, a localized persistent activity bump may either fade out (so the memory of the corresponding item is lost) or merge with another nearby bump (hence the resolution of mnemonic representation for the merged items becomes blurred). We identified specific dependences of these two phenomena on the strength and tuning of recurrent synaptic excitation, as well as network normalization: the overall population activity is invariant to set size and delay duration; therefore, a constant neural resource is shared by and dynamically allocated to the memorized items. We demonstrate that the model reproduces salient observations predicted by both discrete-slot and shared-resource models, and propose testable predictions of the merging phenomenon.

Introduction

Working memory (WM), the ability to internally maintain and manipulate information, is critical for cognition and executive control of behavior (Baddeley, 1992). A hallmark of WM is its limited capacity: we can actively hold a few (~4) unrelated items of information at a time (Miller, 1956; Luck and Vogel, 1997; Cowan, 2005). For visual WM, studies suggest that the limited WM capacity can be accounted for by a fixed number of discrete memory slots ("discrete-slot" model) (Pashler, 1988; Luck and Vogel, 1997; Zhang and Luck, 2008). For instance, in Zhang and Luck's (2008) study, a number of colored squares were flashed on the screen, followed by a brief delay. Then, one of the items was cued and the subject had to report the color of cued square by clicking on a color wheel. The performance data were consistent with a model in which the report has a fixed precision regardless of the set size for a small number of items, and is random for the others, suggesting that the information is stored in discrete slots. Another recent study offered evidence for an alternative explanation for WM capacity limit in terms of a shared, finite resources ("shared-resource" model) with a power-law decay of precision as a function of the set size (Wilken and Ma, 2004; Bays and Husain, 2008). Although the discrete-slot model is intuitively appealing, its neural mechanism is poorly understood. A promising explanation is that each item is actively stored in a subset of neurons which fire synchronously at gamma band and different groups of neurons have different phases; the maximum number of phases limits WM capacity (Lisman and Idiart, 1995; Raffone and Wolters, 2001). Yet little direct neurophysiological evidence has been shown (Fukuda et al., 2010), especially when the items are simultaneously displayed. Moreover, an analog feature such as color is more likely to be encoded by a distributed neural representation (Conway and Tsao, 2009), where the similar colors would interfere with each other (Elmore et al., 2011). For these reasons, it remains unclear about the temporal dynamics of a WM circuit underlying limited capacity.

Received Feb. 15, 2012; revised May 31, 2012; accepted June 27, 2012.

Author contributions: Z.W., X.-J.W., and D.-H.W. designed research; Z.W. and D.-H.W. performed research; Z.W. and D.-H.W. contributed unpublished reagents/analytic tools; Z.W., X.-J.W., and D.-H.W. analyzed data; Z.W., X.-J.W., and D.-H.W. wrote the paper.

This work was supported by NSFC-60974075, 91132702, and the Fundamental Research Funds for the Central Universities (D.-H.W.), NIH-MH062349 and the Kavli Foundation (X.-J.W.). We thank Albert Compte and Wei Ji Ma for discussions, and Moran Furman for help with the model program code. Simulations were carried out at Beijing Normal University.

The authors declare no competing financial interests.

Correspondence should be addressed to either of the following: Xiao-Jing Wang, Department of Neurobiology and Kavli Institute for Neuroscience, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, E-mail: xjwang@yale.edu; or Da-Hui Wang, Department of Systems Science and National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China, E-mail: wangdh@bnu.edu.cn.

DOI:10.1523/JNEUROSCI.0735-12.2012

 0270-6474/12/3211228-13$15.00/0

In this study, we investigated this issue using a spiking neural network of Compte et al. (2000) (with parameter variations), which was designed for WM of an analog quantity like a direction or a position on a color wheel. We found that, whereas the neural representation of cues is distributed in a continuous network, the system behaves in a way consistent with the discrete-slot model, because each item is stored in a distinct bell-shaped activity bump and the network is roughly normalized so that the total activity remains approximately the same for different set sizes, regardless of whether persistent activity bumps are uniformly or randomly distributed in space, and across time in the delay, despite fade-out and merging of bumps. Moreover, we identify two distinct dynamical effects limiting WM capacity, namely excessive (respectively insufficient) recurrent excitation leads to a merging (respectively fade-out) of the activity bumps, which have testable behavioral implications.

Materials and Methods

Model setup. We adopted a ring architecture, suitable for representation of an angular feature by a continuous network with spiking neurons (Compte et al., 2000). The model was originally designed for a spatial WM task, in which the direction, between 0° and 360°, of a spatial cue must be remembered across a delay period (Funahashi et al., 1989). This setting is thus adequate for the Zhang and Luck (2008) experiment, because the position on a color wheel can be described by a directional angle. The model consists of 4096 direction-selective pyramidal cells and 1024 interneurons. Both pyramidal cells and interneurons are modeled as leaky integrate-and-fire neurons (Tuckwell, 1988). The subthreshold membrane potential, $V(t)$, obeys:

$$C_m \frac{dV(t)}{dt} = -g_L(V(t) - V_L) - I_{syn}(t), \quad (1)$$

where $I_{syn}(t)$ is the total synaptic currents to the neuron, C_m is the membrane capacitance, g_L is the leak conductance, and V_L is the resting potential; other parameters are the firing threshold potential, V_{th}, the reset potential, V_{res}, and the refractory period τ. $C_m = 0.5$ nF, $g_L = 0.025$ μS, and $\tau = 2$ ms for pyramidal cells; $C_m = 0.2$ nF, $g_L = 0.020$ nS, and $\tau = 1$ ms for interneurons; $V_L = -70$ mV, $V_{th} = -50$ mV, and $V_{res} = -60$ mV for all neurons (Troyer and Miller, 1997; Wang, 1999).

The recurrent currents are mediated by the receptors of AMPA (AMPAR), NMDA (NMDAR), and GABA (GABAR). The current from the spontaneous neural activities outside the local network is modeled as task-irrelevant background noise, I_{noise}. The external current, I_{ext}, encodes the stimuli in a cue array to pyramidal cells. Each neuron thus receives a total synaptic current as:

$$I_{syn}(t) = (I_{AMPA} + I_{NMDA} + I_{GABA}) + I_{noise} + I_{ext}, \quad (2)$$

Currents mediated by AMPAR, NMDAR, and GABAR to neuron i are modeled as:

$$I_{i,AMPA} = (V_i - V_E)\sum_j g_{ji,AMPA} S_{j,AMPA}, \quad (3)$$

$$I_{i,NMDA} = (V_i - V_E)\sum_j \frac{g_{ji,NMDA} S_{j,NMDA}}{1 + [Mg^{2+}]\exp(-0.062V_i/3.57)}, \quad (4)$$

$$I_{i,GABA} = (V_i - V_I)\sum_j g_{ji,GABA} S_{j,GABA}, \quad (5)$$

where $[Mg^{2+}] = 1$ mM (Jahr and Stevens, 1990), $V_E = 0$ mV, and $V_I = -70$ mV. Given a spike train, $\{t_k\}$, in the presynaptic neuron, a gating variable, s, for AMPAR or GABAR follows a fast dynamics, $\frac{ds(t)}{dt} = -\frac{s(t)}{\tau_s} + \sum_k \delta(t - t_k)$; that for NMDAR obeys a slow dynamics (Wang, 1999), $\frac{dx(t)}{dt} = -\frac{x(t)}{\tau_x} + \sum_k \delta(t - t_k)$, $\frac{ds(t)}{dt} = -\frac{s(t)}{\tau_s} + \alpha_s x(t)(1 - s(t))$, with $\alpha_s = 0.5$ kHz and $\tau_x = 2$ ms. τ_s is 2 ms for AMPAR, 10 ms for GABAR, and 100 ms for NMDAR. The gating variable for background noise is independently determined for each neuron by uncorrelated Poisson spiking train at a rate of 1 kHz, and exclusively mediated by AMPAR with the conductances of 2.48 nS for pyramidal cells and 1.9 nS for interneurons, except in Figure 9*C–E* (see below), where the conductance is 2.18 nS for interneurons.

The connectivity between pyramidal cells is structured, consistent with a columnar organization (Goldman-Rakic, 1995; Rao et al., 1999; Constantinidis et al., 2001; Conway and Tsao, 2009). Specifically, the synaptic coupling between neurons i and j, g_{ij}, is the synaptic conductance G_{EE} multiplied by $W(\theta_i - \theta_j)$, where θ_i is the preferred direction of neuron i. This connectivity $W(\theta_i - \theta_j) = J^- + (J^+ - J^-)\exp\left[-\frac{(\theta_i - \theta_j)^2}{\sigma^2}\right]$ (Compte et al., 2000) is normalized as $\frac{1}{360}\int_0^{360} W(\theta_i - \theta_j)d\theta_j = 1$. The parameters J^-, J^+ describe the strength of the cross- and iso-directional connections, respectively, and $J^- = \frac{360 - \sqrt{2\pi}\sigma J^+}{360 - \sqrt{2\pi}\sigma}$. The connectivity width, σ, reflects the effective cross-interaction range of pyramidal cells. The connections onto and from interneurons are uniform: $g_{ij,EI} = G_{EI}$, $g_{ij,IE} = G_{IE}$, $g_{ij,II} = G_{II}$. We used the neuronal and synaptic parameters from Compte et al. (2000), except those of the connectivity and background noise. Specifically, we gradually varied J^+ from 0.02 to 4.62 and σ from 0.25° to 15.5° (Fig. 3*A*). In most of this paper, we showed the results based on two sets of E-E wiring parameters: $J^+ = 4.02$, $\sigma = 5°$ (narrow connectivity) and $J^+ = 3.62$, $\sigma = 11.25°$ (wide connectivity).

Decoding method. Neurons are divided into subpopulations according to the stimuli. By calculating the population vector for the subpopulation of the αth stimulus in a cue array, we decoded its "memory trace" as $\theta_{out,\alpha}(t) = arg[\sum_{j\in\alpha} r_j(t)\exp(i\theta_j)]$, where the summation is over all the pyramidal cells in this subpopulation; $r_j(t) = \frac{1}{T}\int_{t-T}^{t} r_j(\tau)d\tau$ is the average firing rate of pyramidal cell with label θ_j, $T = 0.1$ s. If an activity bump persists throughout the delay period without merging, the result of such decoding method is consistent with the decoding method described previously (Georgopoulos et al., 1989; Zemel et al., 1998; Deneve et al., 1999; Pouget et al., 2000; Amari and Nakahara, 2005).

Simulation protocol. We simulated two types of tasks to examine the WM performance: delayed-recall tasks (DRTs) and change-detection tasks (CDTs). In a DRT (Figs. 1–7, 9*A*,*B*), the network actively maintains the directions in a cue array as bumps in a delay ≤9 s (Fig. 1*B*). Each cue array contains one or more different directions, and $\theta_{in,\alpha}$ denotes the direction of the α_{th} stimulus. A pyramidal cell, with preferred direction θ, thus receives the external input from all n items in a cue array as $I_{ext}(\theta) = \sum_{\alpha=1}^{n} \frac{I_0}{\sqrt{2\pi}\sigma_s}\exp\left[-\frac{(\theta - \theta_{in,\alpha})^2}{\sigma_s^2}\right]$, $I_0 = 0.4$ nA, and $\sigma_s = 2°$. In a uniform cue array, the items are uniformly distributed from 0° to 360° (Fig. 1*C*), while in a random cue array, they are randomly distributed (Fig. 1*D*) with a minimum distance ≥24° (Zhang and Luck, 2009). The cue array is presented to the network from 0.25 s to 0.5 s and then withdrawn. We based the recall of cue items on neural activity in the last 0.25 s of the delay using a population decoding algorithm. One hundred trials were performed for each condition.

In the CDT (Figs. 8, 9*C–E*), we used the same protocol as that in a DRT for WM retention process, where the color is encoded as a value of hue from 0° to 360° (color set green: 90° to 150°; blue: 210° to 270°; red: 0° to 30° and 330° to 360°). We used the wide connectivity network in Figure 8 (see below) and narrow connectivity network in Figure 9 (see below). In each trial, a cue array (with 2, 3, 4, 6 or 8 colors) and a test array (with the same set size as the cue), separated by a 1 s delay, are shown, and a decision must be made on whether they are the same. In half of the trials, the test arrays, θ_{test}, are identical to the cue arrays, θ_{in}, namely "same" trials, where the amplitude of change is $\Delta = 0°$; while in the other half of the trials, one color in the cue array is changed to a color with an amplitude, Δ, from 10° to 90° away from its value,

namely "diff" trials. To make such a decision based on the memory, we used a downstream "match-nonmatch" neural circuit, which was previously developed by Engel and Wang (2011). Furthermore, we simplified this neural circuit as a sigmoid function (Fig. 8*B*), which decreases with the difference between the memory read-outs, θ_{out}, and test items, θ_{test}:

$$P(same) = \min_{\alpha}\left\{a + \frac{b}{1+\exp[-(|\theta_{out,\alpha}-\theta_{test,\alpha}|-\theta_0)/d\theta]}\right\}, \quad (6)$$

where $a = 0.8214$, $b = -0.8243$, $\theta_0 = 23.57°$, and $d\theta = 6.32°$ for Figure 8 (according to the human behavioral data of Wilken and Ma, 2004; see below); $a = 0.9214$, $b = -0.9243$, $\theta_0 = 28.57°$, and $d\theta = 6.32°$ for Figure 9*C*,*E* (according to the human behavioral data of Lin and Luck, 2009; see below).

Specifically, in a Lin and Luck CDT (Fig. 9*C–E*), we adopted 3 types of cue arrays, namely *far* (low similarity), *close* (high similarity), and *far+close*, and performed 3 types of tests, namely *same*, *diff1* and *diff2* (750 trials for each condition). On *far* trials, each color in a cue array is randomly chosen from a different color set (green, blue or red); on *close* trials, all colors are randomly chosen from the same set; on *far+close* trials, two colors are chosen from the same set, while the other one is chosen from a different set. In *same* tests, the test array is identical to the cue array; in *diff1* tests, one color in the cue array is changed to a color 30° away from the original color (for colors with high similarity, this changed color is on its divergent side); in *diff2* tests (only for colors with high similarity in *close* and *far+close* trials), one color with high similarity is changed to an intermediate level on its convergent side. A mixture of the *same* trials (50%) and *diff1* trials (50%) is equivalent to the behavioral experiments of Lin and Luck (2009). In both *close* and *far+close* trials, the minimum distance between sampled colors are ≥20°.

Quantification of WM performance and capacity. In DRTs, the WM performance can be measured using parameters P_m and s.d. from the discrete-slot model by a von Mises fit $f(x|P_m,\kappa) = \frac{P_m e^{\kappa\cos x}}{360 I_0(\kappa)} + \frac{1-P_m}{360}$, where the response offset x is the difference between the reported and cued directions, κ is a concentration parameter (Fisher, 1993). P_m quantifies the proportion of memorized items and $s.d. = \sqrt{-2\log\left(\frac{I_1(\kappa)}{I_0(\kappa)}\right)}$ describes the memory resolution, where $I_1(\kappa)$ and $I_0(\kappa)$ are the modified Bessel functions with order 1 and 0, respectively. The WM performance can also be measured using parameter $P_s = \frac{1}{s.d.}$ from the shared-resource model, where s.d. is the circular standard deviation (Fisher's fit) $s.d.^2 = -2\log\left(\frac{1}{N}\left|\sum_{n=1}^{N}\exp(ix_n)\right|\right)$ (Bays et al., 2009). When the distribution of the response offset x is nearly uniform, e.g., the set size is large, Fisher's fit would overestimate P_s (or underestimate s.d.). The improved estimation of P_s should be $P_s = \frac{1}{s.d.} - P_0$, where $P_0 = \int \frac{N}{\sqrt{x}\exp[x+N\exp(-x)]}dx$ is the expected precision under uniform distribution of the data (Bays et al., 2009). In Figure 3*D* (see below), we normalized P_s by the maximum value across different set sizes. To generate the response offset distribution, we randomly chose a value from (0°, 360°) as the report of any fade-out bump, and assessed P_s, P_m, and s.d. using the unbinned data and the Matlab code from Bays et al. (2009). In both models, we checked and confirmed that the circular means of the simulated data are around zeros (data not shown).

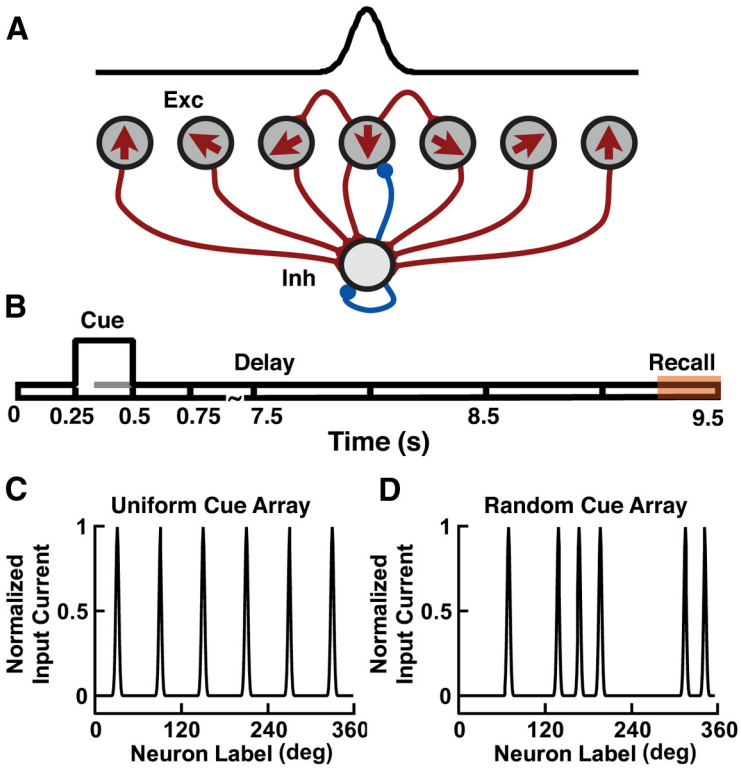

Figure 1. Network model structure and simulation protocol. ***A***, Model scheme. The network is composed of spiking excitatory pyramidal cells (Exc) and inhibitory interneurons (Inh). Pyramidal cells are uniformly placed on a ring, labeled by their preferred directions (shown by arrows). The connections between pyramidal cells are structured as a Gaussian function of the difference in the preferred directions (top), and the connections onto and from the interneurons are uniform. ***B***, Simulation protocol. A cue array is presented to the network from 0.25 s to 0.5 s, followed by a delay period up to 9 s. ***C***, ***D***, Sample cue arrays of 6 uniformly and randomly distributed directions, respectively.

In this study, we also developed two parameters for WM performance: correct rate of the reports, $P_c = \frac{No.\ of\{|\theta_{out,\alpha}-\theta_{in,\alpha}|<\theta_{th}\}}{No.\ of\ items\ in\ cue\ array}$, describes the fraction of report which is close to the cue, and standard deviation of the reports, $S.D. = \sqrt{\sum_{\alpha}(\theta_{out,\alpha}-\theta_{in,\alpha})^2}$, describes the memory precision of the reports from merging and persistent bumps. We used correct threshold $\theta_{th} = 5°$ (low threshold) for most of the performance curve, and showed those at $\theta_{th} = 8°$ for comparison (Fig. 5*B*,*D*) as comparison. Notably, the performance curves are comparatively robust for different correct thresholds for uniform cue arrays, therefore we defined WM capacity as the set size maximizing the product of set size and its P_c, using uniform cue arrays. Moreover, we compared the fitting curves using SD $\left(f(x|S.D.) = \left[erf\left(\frac{180}{\sqrt{2S.D.^2}}\right)\right]^{-1}\exp\left(-\frac{x^2}{2S.D.^2}\right)\right.$ (namely "our model") and using 2-parameter von Mises fit, P_m and s.d. $\left(f(x|P_m,\kappa) = \frac{P_m e^{\kappa\cos x}}{360 I_0(\kappa)} + \frac{1-P_m}{360}\right.$ (namely discrete-slot model) in Figure 5*A* (see below).

For the CDT in Figure 8*A* (see below), we measured the WM performance using parameters hit rate, false alarm rate, and correct rate of the reports, P_c. We defined the hit rate as the probability to respond to "different" in the diff trials, $1 - P\{same, \Delta > 0°\}$, and the false-alarm rate

as the probability to respond to different in the same trials, $1 - P\{same, \Delta = 0°\}$. The correct rate across the diff and same trials is: $\frac{1}{2}(1 - P\{same, \Delta > 50°\}) + \frac{1}{2}P\{same, \Delta = 0°\}$, where half of the trials are the same trials ($\Delta = 0°$); the other half of them are the diff ones ($\Delta > 50°$) (Luck and Vogel, 1997; Vogel et al., 2001). In the diff trials, the amplitude of Δ varies from 60° to 90° (step is 10°) with the same probability. Particularly, probabilities to choose diff at these amplitudes are almost saturated (Fig. 8*D*).

Measures of population activity. The average firing rate of pyramidal cells is $\bar{r}(t_0 + T) = \frac{1}{NT}\int_{t_0}^{t_0+T}\sum_{i=1}^{N} r_i(t)dt$, where $T = 0.25$ s, t_0 is the time 0.25 s before the end of the delay, and $r_i(t)$ is the firing rate of the i_{th} pyramidal cell at time t. The total width of activity bumps is $W_{tot} = \sum_{\alpha} W_{\alpha}$, where W_{α} is the width of α_{th} activity bump assessed according to the spatial profile of firing rates (twice the standard deviation of the Gaussian fit) within 0.25 s preceding the end of the delay.

The instantaneous average recurrent excitatory synaptic conductance, $G_{\alpha}(t)$, and the instantaneous average firing rate of pyramidal cells, $R_{\alpha}(t)$ (Fig. 7*D*,*E*) of the αth bump are calculated as follows:

$$G_{\alpha}(t) = \frac{1}{N_{\alpha}}\sum_{i\in\alpha}\sum_{j\in all}\frac{g_{ji,NMDA}S_{j,NMDA}(t)}{1 + [Mg^{2+}]\exp(-0.0062V_i/3.57)}, \quad (7)$$

$$R_{\alpha}(t) = \frac{1}{N_{\alpha}}\sum_{i\in\alpha} r_i(t). \quad (8)$$

We calculated the average firing rates $\bar{R} = \frac{1}{T}\int R(t)dt$ and the average excitatory synaptic conductance of each activity bump $\bar{G} = \frac{1}{T}\int G(t)dt$ (Fig. 7*F*) using the period, $T = 1$ s, preceding the end of the delay period.

Results

Population coding gives rise to the discrete-slot model in a continuous attractor network with normalization

Using a continuous network model of spiking neurons selective for an angle θ representing an analog feature such as the position on a color wheel (Fig. 1*A*), we investigated WM capacity by examining how the system responded to the presentation of an array of directions (Fig. 1*B–D*). Figure 2*A* shows the spatiotemporal spiking neural activity pattern of a network with wide connectivity ($J^+ = 3.62$, $\sigma = 11.25°$), with firing rates plotted as a color-coded map, for a uniform array of 2, 3, 4 or 6 directions. Several characteristics are worth noting. First, pyramidal cells spontaneously discharge at a low rate (<3 Hz) without tuning to any specific directions before the onset of the cue array. Second, when the cue array is presented, the pyramidal cells, whose preferred directions are close to the stimuli in the cue array, increase their firing rates and form distinct bell-shaped activity profiles (bumps) that encode the directions of the corresponding stimuli. Third, these activity bumps continuously develop after the cue array withdrawn. When the set size is small (Fig. 2*A*, top), the WM load is low, and all the activity bumps can persist throughout the WM delay with slight drifts. For instance, Figure 2*A* (upper left) shows that two activity bumps are elicited in the cueing stage and persist during a 1 s delay with almost the identical bump width. Therefore, the representations of directions are actively maintained in WM and can be read out accurately after the delay. On the other hand, when the set size is large, the WM load is high; some activity bumps may fade out or merge in the WM delay. For instance, in a sample trial with 6 directions (Fig. 2*A*, lower right), one activity bump persists throughout the delay, three bumps fade out and two bumps merge in the early phase of the delay. Hence, after a short delay, i.e., 1 s, the information of three fade-out cues is lost; that of the original directions of two merging cues is blurred.

We assessed the network performance based on the readout from neural population activity in the last 0.25 s of the delay using a population decoding algorithm (Materials and Methods). Consistent with the observations from human visual experiments (Luck and Vogel, 1997; Zhang and Luck, 2008), the network model shows high (poor) performance at small (large, respectively) set sizes. With a 1 s delay (black, Fig. 2*B–D*), both the correct rate of reports ($P_c \approx 100\%$) and memory resolution (SD $\approx 2°$) are high when the set size is smaller than a critical number (~4), while P_c sharply decreases to a low level (~20%) and SD drastically increases to a plateau (~18°), once the set size exceeds this critical number (Fig. 2*C*,*D*). We found that this critical set size not only defines the WM capacity, which maximizes the product of the correct rate of reports and the set size (Fig. 2*B*), but also sets an upper bound of the number of the distinct activity bumps. In our model, since recurrent network dynamics continue to unfold over time, merging or fade-out could occur later in the delay period, therefore WM capacity depends on the delay duration, as it is shown by comparison of performance with 1 s versus 9 s delay (black vs gray, Fig. 2*B–D*). We will return to this model prediction later. The plateau of response precision implies that the network represents the memorized directions with an almost constant accuracy, and the low correct response rate indicates that the network forgets some of them if the set size outnumbers WM capacity. Therefore, even though our model is a continuous network and cue directions are encoded by a distributed neural population, it reproduces the defining behavior of a discrete-slot model.

Interestingly, in our model, the neural population activity is normalized, in the sense that the total width of activity bumps and the average firing rate of pyramidal cells are almost independent of the set size. Although the width of a single activity bump in the upper left panel of Figure 2*A* is obviously wider than that in the upper right panel, the total width of activity bumps are the same. Even the activity bump fades out in Figure 2*A* lower right, the total width of activity bumps remains essentially constant, as the merging bumps expand while the fade-out bumps shrink. To quantify this intuitive observation, we calculated the total width of activity bumps and the average firing rate of all pyramidal cells in 0.25 s preceding the end of the delay period, and found that both are almost constant despite the set size and the delay durations with or without fade-out and merging (Fig. 2*E*,*F*). If we define the WM resources as the activity of pyramidal cells, network normalization indicates that the system recruits a roughly constant amount of memory resource (Bays and Husain, 2008; Buschman et al., 2011), which is dynamically shared by the memorized items (Fig. 2*A*). With a set size smaller than WM capacity, each activity bump has the same width; the neural representations of items in WM thus equally share the resources. With a set size above WM capacity, an activity bump may merge with another activity bump or fade out, and the WM resources are dynamically shifted from a fade-out activity bump to another activity bump or absorbed by merging activity bumps, which agrees with the defining behavior of shared-resource model (Bays and Husain, 2008). Moreover, to maintain the persistent activity, a distinct activity bump must recruit a minimum number of pyramidal cells to make the local excitation strong enough. Therefore, the normalization, which implies a fixed total bump

width, gives rise to a maximum number of distinct activity bumps.

Working memory capacity depends on the strength and width of recurrent excitatory connections

The model has structured excitatory recurrent connections and unstructured inhibitory connections, allowing us to focus on the effects of excitatory-to-excitatory (E-E) connections on WM capacity. Specifically, we gradually varied the E-E connection strength J^+ and spatial footprint σ to examine how WM capacity depends on the local recurrent excitation (Fig. 3*A*). With weak or narrow connections (navy blue), recurrent excitation is insufficient to support any persistent activity bump. Otherwise, WM capacity varies from 2 to 7, which is consistent with the human reports of the single-feature WM capacity (Xu and Chun, 2006). The E-E connectivity thus plays an important role in modulating WM capacity. We found that WM capacity monotonically increases with J^+ given a fixed σ (horizontal white line, Fig. 3*A*). The increase of J^+ enhances iso-directional and weakens cross-directional E-E connections. Consequently, neurons within an activity bump receive stronger mutual excitation among themselves, but neurons in different bumps excite each other less effectively, hence self-maintenance of distinct activity bumps is favored and WM capacity is larger. In contrast, with increasing σ given a fixed J^+, WM capacity increases at first, then decreases (vertical white line, Fig. 3*A*). A narrower connectivity (smaller σ) results in less pyramidal cells recruited to represent each item, as well as decreased mutual excitation; it thus is detrimental to the maintenance of an activity bump and leads to a smaller WM capacity. On the contrary, if σ is large, there would be an excessive number of pyramidal cells for the representation of each item (bumps are wide). These wide bumps merge with a high probability, and WM capacity is thus small. Therefore, a large WM capacity requires strong recurrent excitation with an optimal spatial footprint.

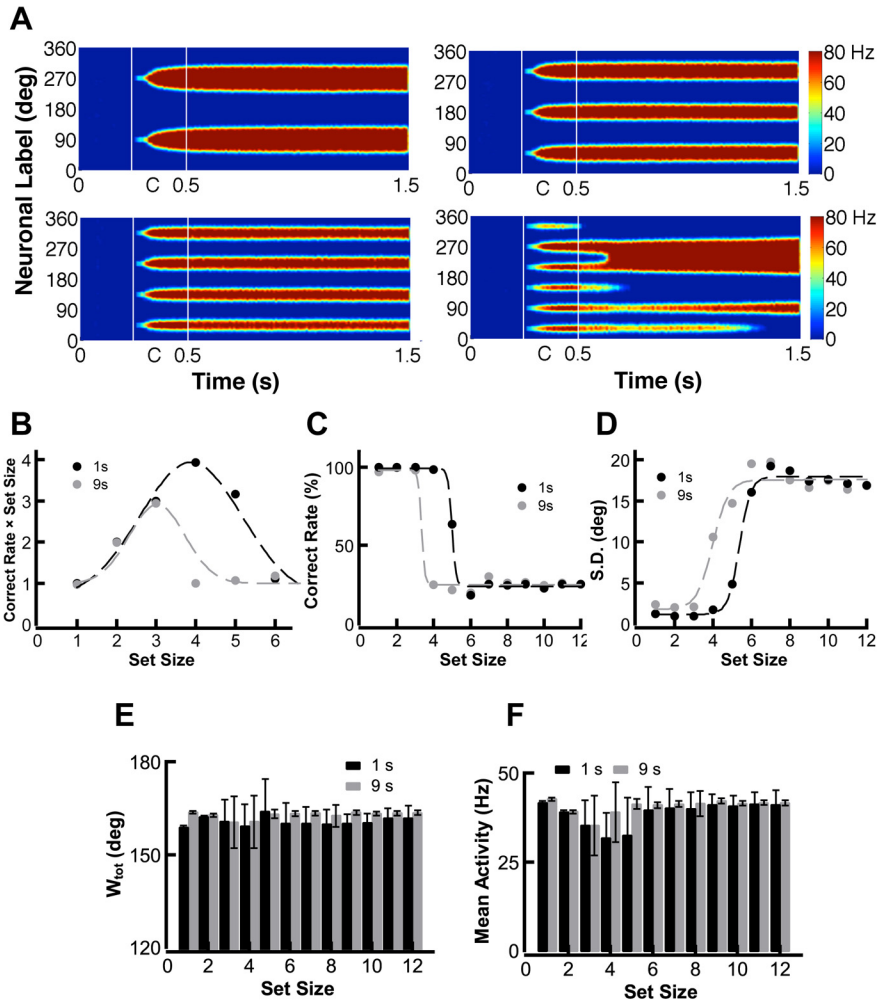

Figure 2. Neural spiking activity, WM performance and normalization. The network has a wide E-E connectivity. ***A***, Spatiotemporal neural activity pattern of pyramidal cells in response to an array of 2, 3, 4, or 6 directions. Pyramidal cells are labeled along the *y*-axis according to the preferred directions. The *x*-axis represents time. Firing rate is color coded. After being briefly presented during the cue period (marked as C on *x*-axis), each stimulus evokes a bell-shaped activity pattern of localized pyramidal cells (bump). In the lower right panel, some activity bumps fade out, some merge with each other, while all the elicited activity bumps in the other three panels persist in a 1 s delay period. Notably, the width of persistent activity bumps decreases with the set size. ***B–D***, The performance of the network. ***B***, The product of the set size and its correct rate of the reports exhibits a maximum. The corresponding set size defines WM capacity. WM capacity is smaller with longer delay duration. ***C***, The correct rate is ~100% for a set size below WM capacity, but declines sharply to ~20% for a set size above it. ***D***, SD is ~2° for a set size below WM capacity, and sharply increases to ~18° for a set size above it. ***E***, ***F***, A constant memory resource by network normalization. The total width of mnemonic activity bumps (***E***) and average firing rate of pyramidal cells (***F***) are almost invariant to set sizes, and roughly the same for 1 s and 9 s delays.

Although WM capacity depends on E-E connections, the typical characteristics of the networks with different E-E connections are similar. Given the uniform cue arrays, the performance curves of the network with narrow connectivity (×**a**, Fig. 3*A*) are similar to those of the network with wide connectivity (×**b**, Fig. 3*A*). Given uniform cue arrays, P_c drastically decreases to a low level and SD sharply increases to a plateau when the set size exceeds WM capacity (~3 in Fig. 2 and ~6 in Fig. 3*B*,*C* for a 9 s delay), which have a step-like shape. However, given the random cue arrays in which the minimum distance between items is ≥24° (Zhang and Luck, 2009), P_c continuously decreases and SD continuously increases with the set size until reaching the same plateaus for the uniform cue arrays (Fig. 3*B*,*C*). Of note, due to the attractor dynamics in our model, the memorized items are stored separately in different discrete slots during retention process (discrete-slot feature), wherefore SD (Fig. 3*C*) for random cue arrays resembles the human experimental data from the discrete-slot model. Using the global inhibition, the roughly constant memory resource is dynamically allocated to the memorized items (shared-resource feature), wherefore the relative precision, which is normalized by the maximum P_s over all trials (Materials and Methods), follows a power-law decay function of set size, which resembles that observed in the shared-resource model (compare Bays and Husain, 2008, their Fig. 3*B*, with our Fig. 3*D*).

Surprisingly, although the persistent activity pattern and performance are considerably different with random versus uniform array of cues, network's normalization is remarkably similar for the two types of cue arrays (Fig. 3*E*,*F*). Therefore, the total WM

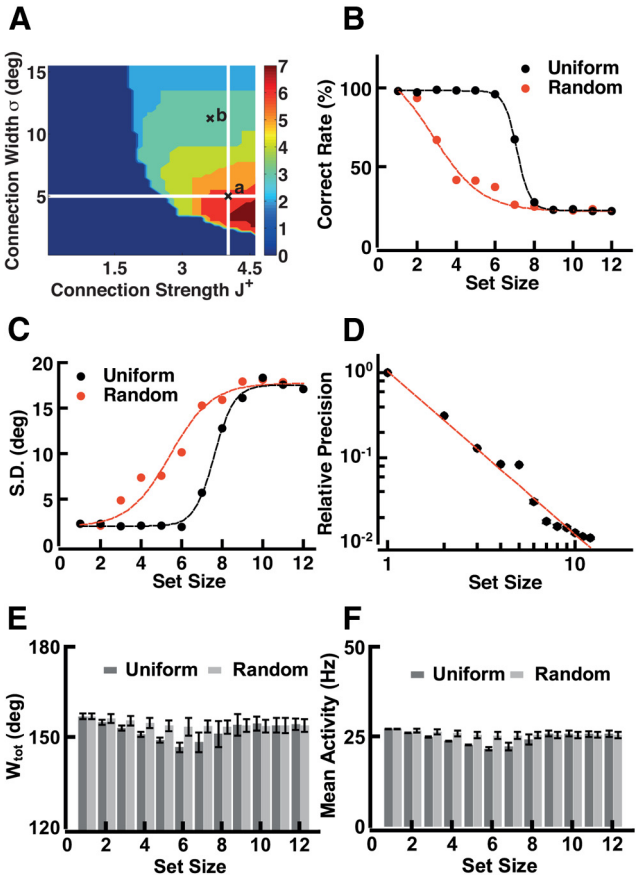

Figure 3. WM capacity depends on E-E connections. ***A***, Dependence of WM capacity on E-E connectivity. The WM capacity is color coded and shown on the plane of two parameters characterizing E-E connections: the strength J^+ and spatial width σ. When J^+ is too weak or σ is too small (navy), the persistent activity is absent. Outside of that region, WM capacity ranges from 2 to 7; it is larger with more narrowly structured local synaptic excitation (smaller σ), which needs to be compensated by larger connection strength J^+ to ensure sufficient recurrent excitation for WM maintenance. For a fixed $J^+ = 4.02$, WM capacity increases at first and then decreases with the increasing connection width, peaking at 7 (vertical white line). For a fixed $\sigma = 5°$, WM capacity increases with the increasing connection strength (horizontal white line). ***B***, ***C***, Performance of a narrow connectivity network (×**a** in ***A***) with uniform and random cue arrays. Correct rate and SD show a step-like transition as set size increases for uniform cue arrays (black), while there is a smooth function for random cue arrays (red). All the fitting curves are sigmoid functions. ***D***, The relative precision for random cue arrays exhibits power-law dependence on set size. The total width of activity bumps (***E***) and average firing rate of pyramidal cells (***F***) are normalized for both uniformly and randomly distributed arrays of directional cues.

resources are independent of the details of external inputs, but are determined by the E-E connectivity profile (comparing Fig. 2*E*,*F* with 3*E*,*F*).

Working memory capacity depends on delay duration

Figure 4*A* shows the same sample spatiotemporal patterns as those in Figure 2*A*, except for a longer delay of 9 s. Notably, activity bumps may fade out or merge at different times in the delay. This explains why the performance is different for a delay of 1 s versus 9 s (Fig. 2*B*), when the set size is in an intermediate range. For a small set size, none of the activity bumps will fade out or merge in a prolonged delay, whereas for a large set size, bump fade-out or merging takes place early. In both cases, WM performance is insensitive to the delay duration. On the other hand, for intermediate "critical" set sizes, bump merging and fade-out exhibit slow stochastic dynamics during the delay. Consequently, WM capacity exhibits a dependence on delay duration (Fig. 4*B*, top). SD also depends on delay duration in a trend mirroring that of correct rate (Fig. 4*B*, bottom), and the set size at which SD starts to saturate is roughly a linear function of WM capacity (Fig. 4*C*).

Comparing capacity estimation using different measurements

The behavior of WM performance "near a critical set size" has been examined in a human study as a probe to the forgetting mechanism of WM (Zhang and Luck, 2009). It was assumed that a subject would report a random value in the case he or she forgets the memorized item, or report a value around the original cue when he or she remembers it (Zhang and Luck, 2008). This discrete-slot model was formulated using a 2-parameter von Mises fit. To compare with the psychophysical data from the discrete-slot model, we performed this fit to the distribution of the response offset, $\theta_{out} - \theta_{in}$ (unbinned data), for random cue arrays. With a 1 s delay, we found that the discrete-slot fit of simulated data (Fig. 5*A*) is comparable with that of Zhang and Luck (2008), their Figure 1*c*. We compared quantities, P_m and s.d. of discrete-slot model with P_c and SD of our model using the same data and plotted as functions of set size (Fig. 5*B*,*C*) and delay duration (Fig. 5*D*,*E*). These two quantification methods display the same trend of performance: P_m and P_c decreases (s.d. and SD increases) as a function of set size and delay duration. Of note, (1) P_c with high threshold displays a smooth decrease, which resembles P_m (Fig. 5*B*,*D*); (2) although SD from our model exhibits a continuous and smooth increasing against set size (Bays and Husain, 2008; Bays et al., 2009) and delay duration, its alternative fit, s.d., reaches a plateau when set size is >4 (capacity), which is consistent with the behavioral data in recent studies using brief delays (Zhang and Luck, 2008; Anderson et al., 2011), and the prediction by Fukuda et al. (2010). Notably, s.d. is nearly constant against the delay duration (Fig. 5*E*), indicating that a declined performance with longer delays mostly results from the sudden death of the mnemonic items (fade-out; Zhang and Luck, 2009).

Network mechanism underlying fade-out and merging of activity bumps

We have identified the fade-out and merging as the main dynamic effects limiting WM capacity. Indeed, the probability of a fade-out or a merging bump sharply increases from a low level to a high plateau when the set size increases above the WM capacity, as shown for both wide (Fig. 6*A*) and narrow (Fig. 6*B*) network connectivities. The sum of the fade-out and merging probabilities approaches 100% when set size is much larger than WM capacity.

With the same parameters as in Figure 6*B*, and a set size of 6 (capacity), all the activity bumps persists throughout a delay of 9 s (Fig. 7*A*). However, with a set size of 8 (above capacity), only four activity bumps persist until the end of the delay, whereas two activity bumps fade out (at 67.5° and 337.5°) and two others merge (at 247.5° and 292.5°) into one (Fig. 7*B*). A persistent activity bump has a bell-shaped spatial distribution of neural activity, while the merging activity bump displays a wide plateau in the spatial profile of neural activity, as shown by the population activity for the last 1 s of the delay (Fig. 7*A*,*B*, right).

The single neurons inside these three types of bumps display the distinct firing activities during the delay (Fig. 7*C*). First, neurons around the peak of the persistent bump at 22.5° (*a* in Fig. 7*B*,*C*), spike at a high frequency in the cueing stage and at a moderate frequency ~50 Hz with small fluctuations during the delay, which maintain the memory trace of the cue direction. Second, neurons in the bump at 67.5° (*b* in Fig. 7*B*,*C*) that even-

tually fades out, increase their firing rates upon cue presentation, exhibit persistent activity (~50 Hz) in the early phase of the delay, but abruptly cease firing at ~3 s in the delay. This sudden disappearance of mnemonic activity reveals an "all-or-none" mechanism for losing the memorized information. Third, for neurons within and between two bumps at 247.5° and 292.5° (*c* in Fig. 7*B*,*C*), the firing rates are initially quite different. Over time, however, they all converge to a similar activity level of ~50 Hz late in the delay, when the two bumps eventually merge with each other. Neurons near the centers of the two activity bumps behave similarly as those in a persistent bump, whereas firing rates of neurons located at the edges of the original two distinct bumps slowly ramp up; neurons in the midpoint between the two distinct bumps are essentially silent in the early phase of the delay period, but display a sharp jump of activity to ~50 Hz in the late phase of delay. Therefore a gradual ramping and a sudden transition from spontaneous activity to persistent state in the delay may be manifestations of a merging phenomenon that is observable at the single cell level.

E-E interactions support the persistent activity (Fig. 3*A*), which might be also a key factor determining whether a bump fades out or merges with another. To test this, we calculated two quantities for each activity bump in Figure 7*B*: the instantaneous average recurrent excitatory synaptic conductance, $G(t)$, and the instantaneous average firing rate of pyramidal cells $R(t)$ throughout the delay, and classified them into three groups: persistent bumps (P), fade-out bumps (F), and merging bumps (M) (Fig. 7*D*,*E*). In a fade-out bump, $G(t)$ exhibits a sharp decrease at an unpredictable time in the delay, to a small but nonzero level. $R(t)$ decreases to zero Hz, implying that the excitatory drive they receive is below firing threshold. Note that the sudden drop of $G(t)$ precedes that of $R(t)$, as expected for a fade-out process: the decrease of excitatory currents leads to less spikes in a localized activity bump, which in turn results in further weaker recurrent excitation; the cycle continues until the overall excitatory drive becomes too small and the bump fades out. For the bumps that eventually merge, $G(t)$ and $R(t)$ increase during the merging process and reach a high level afterward. The increase of $G(t)$ preceding that of $R(t)$ displays the process opposite to the observation of fade-out: stronger excitatory currents lead to more spikes in the localized activity bump, which results in even more recurrent excitation; when this positive feedback exceeds a certain level, neurons between the two activity bumps receive enough excitation to switch to a high activity state (Fig. 7*C*, right), and the two activity bumps merge with each other. For persistent bumps, $G(t)$ and $R(t)$ fluctuate but remain roughly constant over time. Their values are smaller than those of merging bumps and larger than those of fade-out bumps. Therefore, insufficient excitation leads to fade-out, while excessive excitation results in merging.

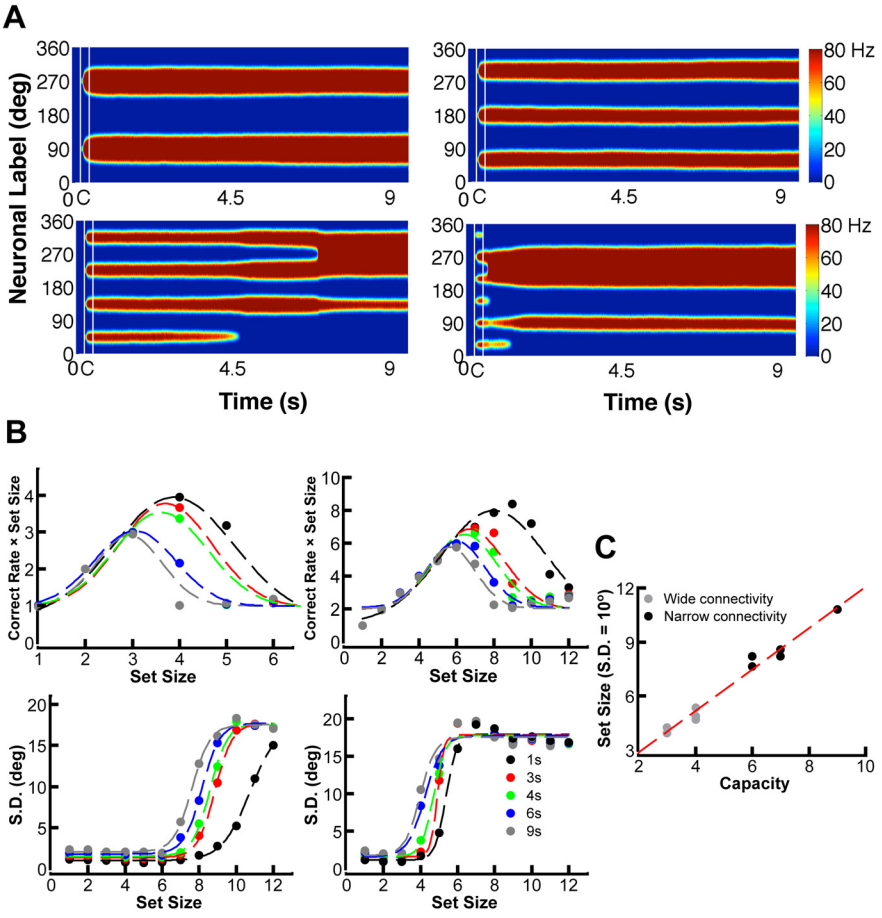

Figure 4. The effects of delay duration on performance. ***A***, Same sample trials as in Figure 2*A*, except shown for a 9 s delay. For a set size, 4, near WM capacity (lower left), bump fade-out or merging may occur late in the delay. ***B***, Top, WM capacity depends on the delay duration and the E-E connections (left for wide; right for narrow connectivity). Bottom, SD increases as a function of set size for different delay durations, with wide (left) or narrow (right) connectivity. The network performance is essentially independent of the delay duration for small or large set size. However, for an intermediate set size, the performance of the network deteriorates with a prolonged delay period, as found in the human experiment (Zhang and Luck, 2009). ***C***, The set size at which SD reaches a threshold level (10°) is linear with WM capacity for all the conditions considered, different delays, and narrow and wide connectivities.

To better examine the correlation between recurrent excitation and neural activity, we calculated the average firing rates $\bar{R}$ and the average excitatory synaptic conductance of each activity bump $\bar{G}$. Figure 7*F* shows $\bar{R}$ plotted against $\bar{G}$ for different activity bumps using a uniform cue array of set size 8. Three groups can be clearly discerned: the values of $\bar{R}$ and $\bar{G}$ for merging bumps are larger than those for persistent bumps, which are larger than those for fade-out activity bumps. Specifically, insufficient local excitation (in nS), $\bar{G} < 32$, leads to fade-out; strong recurrent excitation, $32 < \bar{G} < 35$, ensures a persistent bump; and excessive recurrent excitation, $\bar{G} > 35$, results in merging of activity bumps.

Working memory capacity estimation using change-detection tasks

In addition to the DRTs, the CDT is an alternative experimental scheme widely used to assess WM capacity. However, it is not trivial to base the change-detection performance on that of DRTs, because a CDT includes three stages: (1) sample stage, for encoding the visual inputs; (2) retention stage, for working memory; (3) retrieval stage, for a decision upon the memory, and any of these stages can independently influence the *post hoc* perfor-

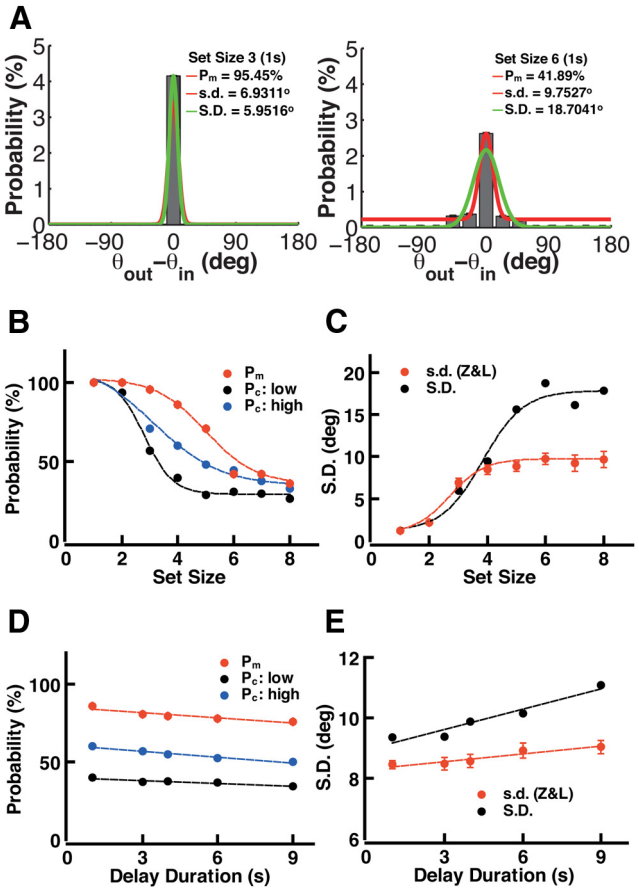

Figure 5. Performance fit using different models. The simulation uses the wide-connectivity network and random cue arrays. ***A***, Typical response offset histograms for set size 3 and 6 (left and right, respectively) with 1 s delay, fitted by discrete-slot model (P_m and s.d., red line) and our model (S.D., green line). ***B***, P_m and P_c decrease as a sigmoid function of set size (delay duration is 1 s). P_c at high threshold (blue line) decreases more smoothly than that at low threshold (black line), which resembles P_m. ***C***, s.d. and S.D. increase as a sigmoid function of set size (delay duration is 1 s). s.d. reaches a plateau as set size is >4 (capacity). ***D***, P_m and P_c decrease as a function of delay duration (set size is 4 at the capacity); P_c at low threshold is nearly constant against the time. ***E***, s.d. and S.D. increase as a function of delay duration (set size is 4 at the capacity); s.d. is nearly constant against the time.

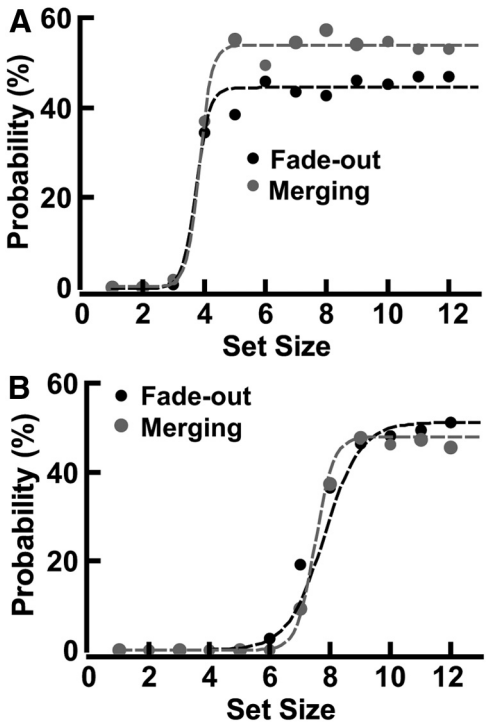

Figure 6. Dependence of fade-out and merging of mnemonic activity bumps on set size. The fraction of mnemonic activity bumps that fade out (black) or merge with each other (gray) during a 9 s delay is plotted as a function of set size. When set size is below WM capacity (3 in ***A*** with wide connectivity, 6 in ***B*** with narrow connectivity), activity bumps seldom fade out or merge. For a set size above WM capacity, the probabilities for merging and fade-out increase sharply. With a sufficiently large set size, a plateau is reached where the sum of the fade-out and merging probabilities is ~100%, hence an activity bump either fades out or merges with another bump.

mance in the detection task. The previous research exhibits either that the sample stage plays a bottleneck role of limiting the number of the encoded items in memory in a bottom-up manner via attention, and thus the working memory capacity (Zhang and Luck, 2008; Buschman et al., 2011), or that retrieval process affects the detection accuracy through an inhibitory reciprocal network (Johnson et al., 2009), nevertheless, little has been unfolded from WM retention process per se. The previous behavioral observation (Basile and Hampton, 2011) demonstrates that the psychometric curves from CDTs mimic that from DRTs; one can therefore predict that change-detection performance would decrease with increasing the set size (Luck and Vogel, 1997; Vogel et al., 2001; Wilken and Ma, 2004; Basile and Hampton, 2011; Elmore et al., 2011). To test this hypothesis, we performed a change-detection task with different set size (Fig. 8).

In this CDT, the set size of the cue array is the same as that of the test array. The network reports that the test array is the same (match) as or different (nonmatch) from the cue array. In half of the trials, we used the test arrays which are identical to the cue arrays, namely same trials (the amplitude of change is 0°; Fig. 8*A*, top), while in the other half of the trials, one color in the cue array is changed to a color with an amplitude from 10° to 90° away from its original value, namely diff trials (nonmatch; Fig. 8*A*, bottom). The probability to report match is obtained from a downstream match-nonmatch decision neural circuit (Fig. 8*B*; Engel and Wang, 2011). In simulations, we found that hit rate decreases and false-alarm rate increases as a function of set size (Fig. 8*C*), which is consistent with the behavioral observations (Wilken and Ma, 2004, their Fig. 4). Psychometric curves shows the probability to respond to diff as a function of the amplitude of change, $|\theta_{in} - \theta_{test}|$, for different set sizes (Fig. 8*D*). Of note, when the amplitude of change is small, e.g., $|\theta_{in} - \theta_{test}| = 10°$, the change-detection performance is improved as the set size increases, implying that "similarity" could improve the change-detection performance in some parameter regime. When the amplitude is large, e.g., $|\theta_{in} - \theta_{test}| > 50°$, the psychometric curves are saturated, and the performance curve from CDT mimics that from DRT (Fig. 8*E*) (Wilken and Ma, 2004; Basile and Hampton, 2011; see also the comparison between the change-detection tasks with different amplitudes of change by Fougnie et al., 2010). Overall, our simulations demonstrate that the change-detection performance would decrease when the set size increases, which agrees with the predicted performance in DRTs in Figure 5, *B* and *C*.

Similarity effect on working memory performance

Our model exhibits two distinct mechanisms underlying the decrease of memory precision with the increase of WM loads or delay duration in DRT: fade-out (complete loss of stored information), and merging (more quantitative blurring of stored information). However, a misreporting error from merging can

easily be overlooked in analysis of a DRT with a minimum distance ≥24°, using binned data (Bays et al., 2010). We thus proposed two testable tasks to investigate merging or similarity effect on the WM performance.

First, we investigated the merging process using 2-item cues with different similarity and found that (1) merging can take place given a long delay (~8 s) when the items are of the weak similarity (100°; Fig. 9*A*); (2) merging leads the memory traces bias to the convergent side (Fig. 9*B*), where similarity of items gets enhanced (still 2 items), and network could thus confuse one another in a CDT using test arrays which are similar to the cue arrays (Fig. 8), e.g., purple to blue (Elmore et al., 2011).

We then conducted a Lin and Luck (2009) CDT (Materials and Methods), using three types of the cue arrays, namely *far* (low-similarity), *close* (high-similarity), *far+close*, and three types of the test arrays, namely *same*, *diff1* (nonconvergent side), *diff2* (convergent side; Fig. 9*C*). In simulations, merging occurs only between the high-similarity items. Consistent with 2-item-cue result, the memory traces of high-similarity items converge to an intermediate level, while that of a low-similarity item drifts around the cue (Fig. 9*D*). As a result, the distribution of response offset of the high-similarity items biases to the convergent side, implying an increase (decrease) of the distance between the cue and test arrays for *diff1* (*diff2*, respectively) trials, while that of the low-similarity items is centered at zero. One can thus argue that the similarity would show a great effect on the tests of *diff1* and *diff2*, but little on that of *same*. To test this, we assessed the probability of choosing same for each trial using a downstream match-non-match decision circuit (Fig. 8*B*). Figure 9*E* demonstrates that all three types of trials exhibit similar performance in the *same* test; trials with high similarity show a better performance in the *diff1* test, which resembles the behavioral observation that similarity improves the performance in a Lin and Luck (2009) task, whereas the similarity deteriorates the performance in the *diff2* test.

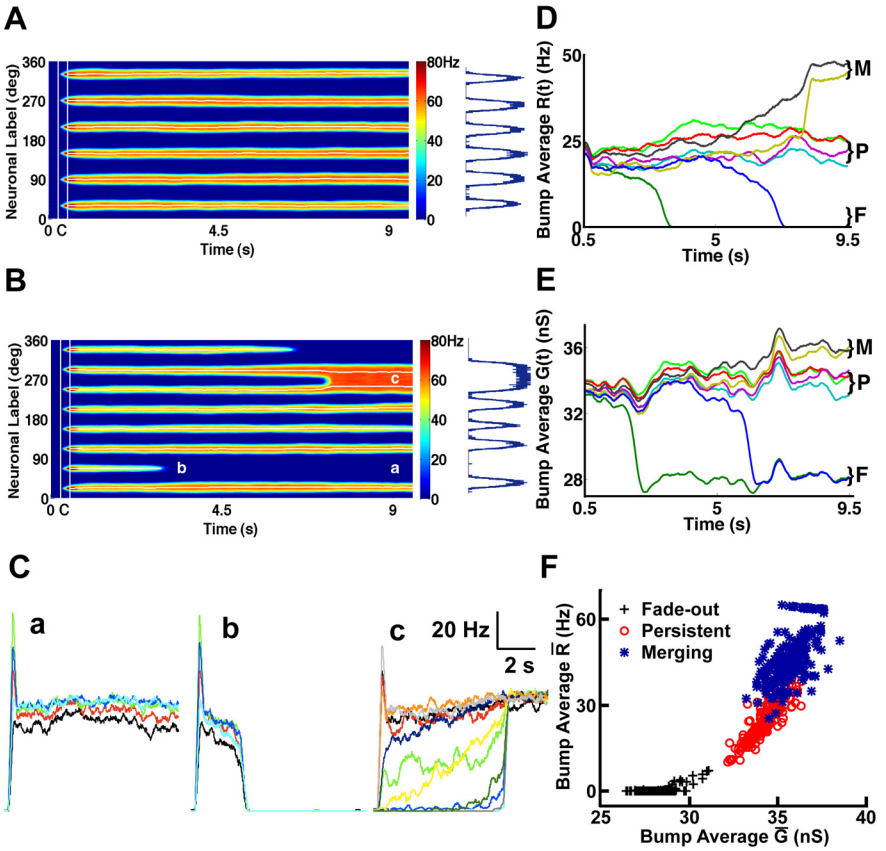

Figure 7. Dynamics of persistent, fade-out, and merging activity bumps in WM delay. The network has a narrow connectivity. ***A***, Spatiotemporal activity in response to a brief stimulus of uniform cue arrays with 6 items (capacity). All activity bumps persist throughout the delay. White lines are memory traces. The spatial distribution of pyramidal cells' firing rate (last 1 s in the delay) shows a bell-shaped profile (activity bump). ***B***, Same as in ***A*** except for a set size above WM capacity. Two bumps (67.5° and 337.5°) fade out, and two bumps (247.5° and 292.5°) merge into one wide bump. The other bumps persist throughout the delay. The activity profile shows a wide plateau for the merging bumps and comparatively sharp peaks for the persistent bumps. ***C***, Firing activity of single neurons marked as ***a***, ***b***, and ***c*** in ***B***. Left, Neurons near the peak of a persistent bump (22.5°) spike at a high rate in the cueing stage and show persistent activity at ~50 Hz during the delay. Middle, Neurons in a fade-out bump (67.5°) abruptly drop their activities in the middle of the delay, implying a sudden death of the corresponding item. Right, Firing activity of neurons (from 247.5° to 292.5°) within and between two bumps that eventually merge into one. Neurons within bumps (e.g., red) behave in a manner similar to those in the left panel. Neurons at the edge of bumps (e.g., yellow) are not boosted by the cue stimulus, but their firing rates gradually ramp up to a stable level during the delay. Neurons in the middle of two bumps (e.g., dark green, ~270°) spike at a low rate in the early phase and suddenly jump to persistent activity in the late delay. ***D***, ***E***, The feedback dynamics between neural firing and recurrent excitatory drive (data from ***B***). ***D***, Instantaneous average firing rate, $R(t)$, of each bump as a function of time. M, P, and F denote merging, persistent, and fade-out bumps, respectively. $R(t)$ values of fade-out bumps suddenly drop to 0 Hz at a random time in WM delay as an all-or-none process. $R(t)$ values of persistent bumps stay at ~25 Hz with small fluctuations. $R(t)$ values of merging bumps gradually increase (black) or jump (yellow) to ~45 Hz after merging. ***E***, Instantaneous average excitatory synaptic conductance, $G(t)$, of each activity bump as a function of time. $G(t)$ values of fade-out bumps quickly decay to ~28 nS preceding the sudden decreases of $R(t)$; $G(t)$ values of merging bumps increase above 35 nS (larger than the maximum value of $G(t)$ of persistent bumps) preceding the merging process. ***F***, The average firing rate $\bar{R}$ plotted against the average excitatory synaptic conductance $\bar{G}$ for different activity bumps (from 100 simulations using the same network and cue arrays as those in ***B***). Three activity groups can be clearly discerned: merging bumps (blue) have high $\bar{R}$ and $\bar{G}$, while fade-out bumps (black) have low $\bar{R}$ and $\bar{G}$.

To conclude, we proposed testable tasks to detect the merging in WM delay, and showed that similarity in cue arrays can either improve (Johnson et al., 2009; Lin and Luck, 2009) or impair (Elmore et al., 2011) the detection accuracy, mainly relying on the post-WM comparison process (compare also performance curves in DRT and CDT in Fig. 8*E*; Hollingworth, 2003; Mitroff et al., 2004).

Discussion

In this work, we carried out a systematic study of WM capacity using a spiking network. We found that the model actively maintains the multiple objects with an analog feature using concurrent activity bumps and reproduces the salient characteristics of performance in visual WM tasks (Bays and Husain, 2008; Zhang and Luck, 2008, 2009; Anderson et al., 2011). The spatial extent (σ) and the strength (J^+) of recurrent synaptic excitation greatly affect WM capacity (Wang et al., 2011), in contrast to or complement with previous work that the spatial extent of lateral inhibition determines WM capacity (Macoveanu et al., 2006; Edin et al., 2009). We also identify two distinct dynamical effects limiting WM capacity, namely excessive (respectively insufficient) recurrent excitation leads to a merging (respectively fade-out) of the activity bumps.

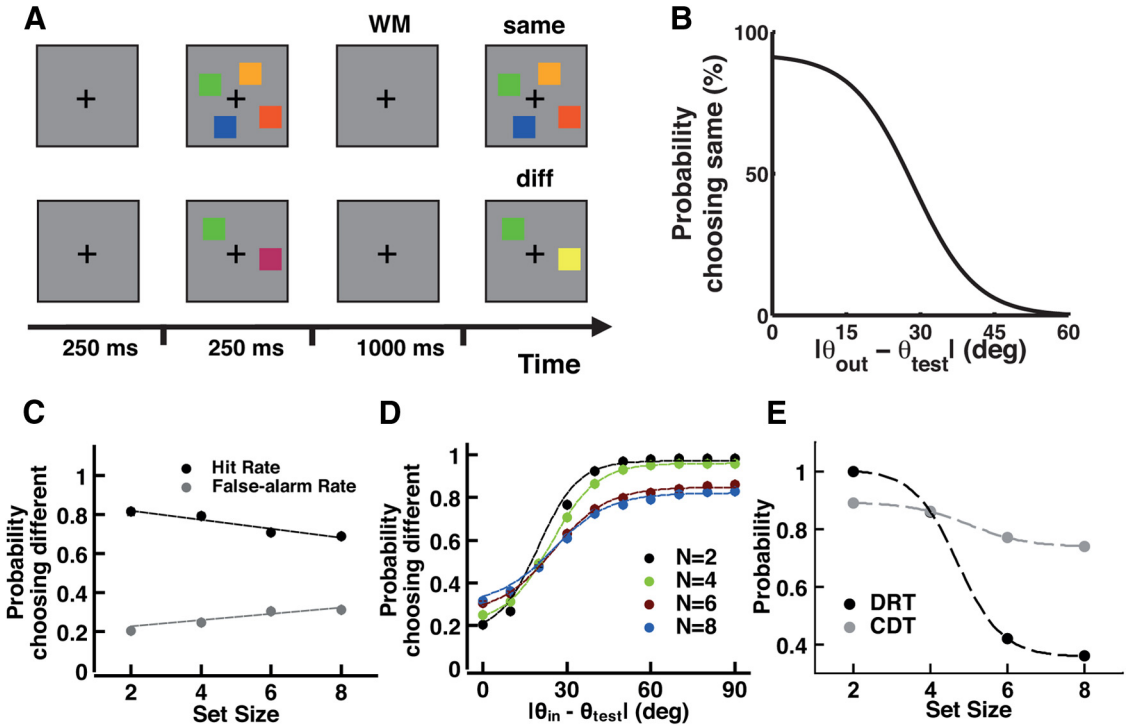

Figure 8. Performance as a function of size set in a change-detection task. ***A***, Experimental scheme of a change-detection task. In each trial, network views a cue array (with 2, 4, 6 or 8 colors) and a test array (with the same set size as the cue), separated by a 1 s delay, and identifies whether they are the same. In half of the trials, the test arrays are identical to the cue arrays, namely same trials, where the amplitude of change is 0°, while in the other half of the trials, one color in the cue array is changed to a color with an amplitude from 10° to 90° away from its value, namely diff trials. ***B***, A downstream match-nonmatch neural circuit underlies the probability of responding to same. ***C***, Hit rate decreases and false-alarm rate increases as a function of set size. ***D***, Psychometric curves show the probability to respond to different as a function of the amplitude of change, $|\theta_{in} - \theta_{test}|$, for different set sizes. ***E***, Performance curve from CDT mimics that from DRT, showing that the change-detection performance (the probability of correct response) would decrease when the set size increases.

Reconciling discrete-slot and shared-resource models in a neural circuit model

Two models have been proposed to understand WM capacity based on the psychophysical observations, i.e., discrete-slot model, wherein the capacity originates from the number of discrete memory slots (Luck and Vogel, 1997; Zhang and Luck, 2008) and shared-resource model, wherein the capacity is conceptually limited by a constant memory resource (Wilken and Ma, 2004; Bays and Husain, 2008). For discrete-slot model, Lisman and Idiart (1995) and Raffone and Wolters (2001) suggested that items are maintained by oscillatory activity across populations. It has been observed that neuronal activity about items in WM is enhanced at specific phases of gamma cycle (Siegel et al., 2009). However, the non-overlapping enhanced phases may result from the sequential presentation of the items. Furthermore, if the memorized items were encoded as the different non-overlapping phases, the interference between them would not be observed in experiments. Alternatively, our model provides the neural mechanism underlying discrete-slot and shared-resource models without considering phase code.

Behaviorally, our model offers a unifying explanation for seemingly incompatible features from the two contrast models. In the psychological studies, the discrete-slot model predicts a hard limit of WM, where memory resolution decreases as a bilinear function of set size, while shared-resource model predicts a monotonic decline (Anderson et al., 2011). Our model exhibits the hard limit of WM capacity in a broad parameter region. However, for the network with narrow connectivity, this hard limit could increase to a large number, therefore the shared-resource-like behavior is observed, using random cue arrays. Furthermore, when the set size exceeds WM capacity, some bumps fade out suddenly, rather than a gradual exponential decay during WM delay. The fade-out implies the "sudden death" of WM in human experiments, which strongly supports the discrete-slot model (Zhang and Luck, 2009) but is hardly accounted for by the shared-resource model (Huang, 2010). While, with randomly distributed cues, correct rates and WM resolutions smoothly decrease, which was taken as strong evidence for shared-resource model (Bays and Husain, 2008). Finally, we found the interference between similar items, which only supports shared-resource model (Wilken and Ma, 2004; Elmore et al., 2011), but would be hardly incorporated in the discrete-slot model. Of note, in the broad parameter region, these two behavioral features, sudden death and "interference," could coexist in our model; the probability of sudden death (respectively interference) increases in a network with narrow (respectively wide) connectivity. Therefore, our model provides a hybrid view for WM capacity, that the cue items are memorized into different chunks ("activity bumps"); items within the same chunk shows shared-resource-like behaviors ("merging bumps"), while chunks behave like discrete-slots ("fade-out" from through global inhibition; Buschman et al., 2011; Machizawa and Driver, 2011).

Besides the behavioral observations, we also found neural evidence for reconciliation of these two models. First, the overall activity of memory neurons remains nearly constant despite the fade-out and merging of bumps as increasing WM loads, agreeing with the neurophysiological observation that the average firing rate of the prefrontal cortex neurons of behaving monkey is roughly identical using different number of cues during WM maintenance (Siegel et al., 2009). Second, a limited WM resource

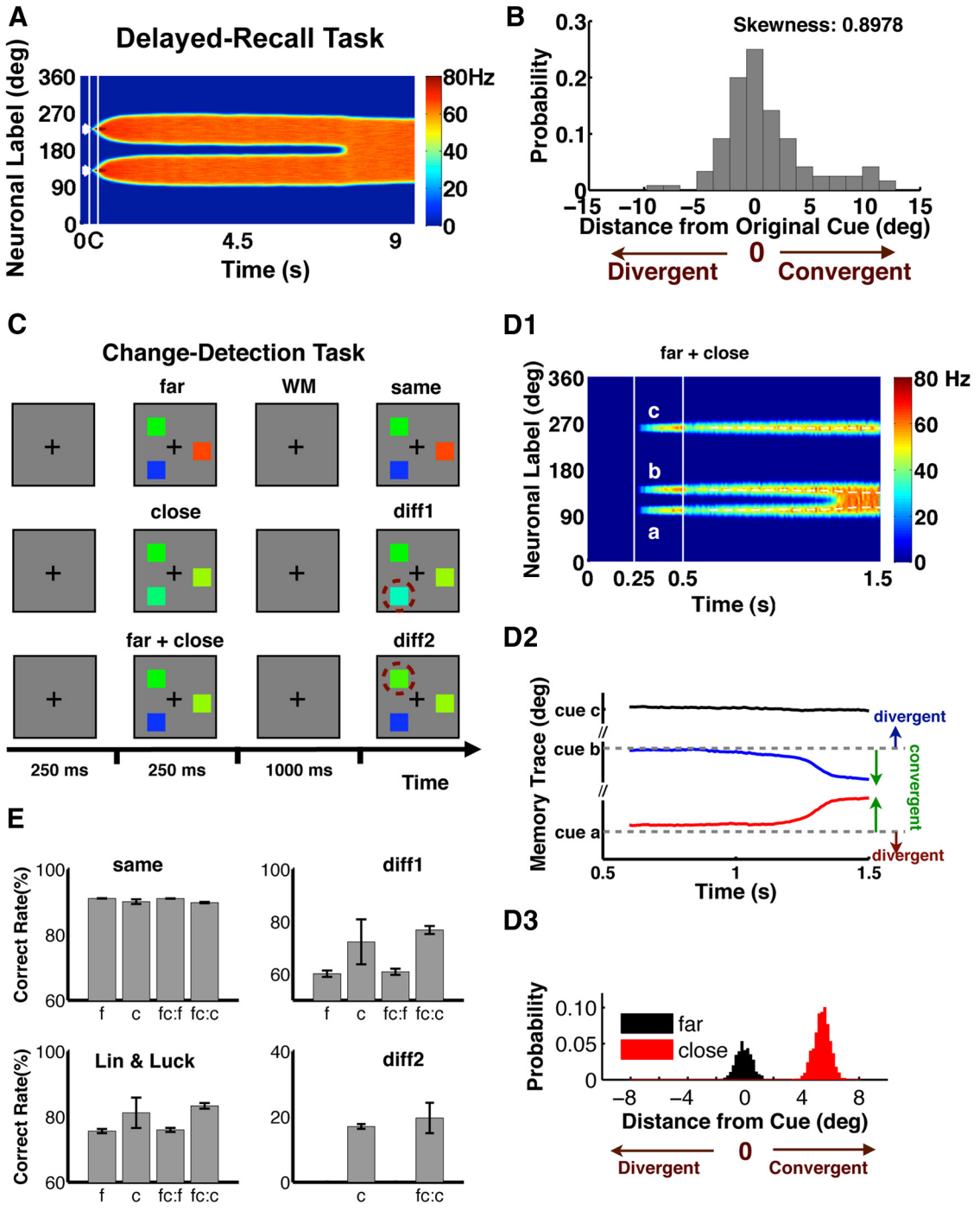

Figure 9. Behavioral manifestation of merging in the change-detection task and free-recall task. ***A***, Merging happens even when two cues are sufficiently separated from each other: $\theta_{in,1} = 130°$ and $\theta_{in,2} = 230°$ (white arrows) in DRTs. Two stimuli evoke two activity bumps, which eventually merge into a single wide bump. Therefore, the reports bias to the center of $\theta_{in,1}$ and $\theta_{in,2}$. ***B***, The distribution of the difference between the report and the original cue, $\theta_{in,2} - \theta_{out,2}$ or $\theta_{out,1} - \theta_{in,1}$ ($\theta_{in,1} < \theta_{in,2}$), across 100 trials. A positive (respectively, negative) distance from the original cue implies convergence (respectively, divergence) of two activity bumps. The distribution skews significantly to the positive side, indicating that merging happens in a large amount of trials. Such a skewed distribution of reports can be tested in behavioral experiments. ***C***, Experimental scheme of a change-detection task to test merging. In each trial, network views a cue array (3 colors) and a test array, separated by a 1 s delay, and identifies whether they are the same. Three types of cues, i.e., *far* (low-similarity), *close* (high-similarity), and *far*+*close*, and three types of tests, i.e., *same*, *diff1* (divergent side), and *diff2* (convergent side; changed colors in *diff1* and *diff2* are circled) are applied in the task; the Lin and Luck (2009) task is a mixture of *same* (50%) and *diff1* (50%) tests. ***D1***, ***D2***, A sample from *far*+*close* trials (2 greens + 1 blue) exhibits a merging process between 2 greens (***a***, ***b***), the memory traces of which converge to an intermediate level (still greens); the memory trace of the blue (***c***) only drifts around its initial cue. ***D3***, Distributions of the response offset in different trials. That with low similarity (black bars) centers at 0°; that with high similarity (red bars) shows a strong bias to the convergent side (>0°). ***E***, Performance for each test. Low- and high-similarity trials show similar performance in the *same* test (upper left). High-similarity trials show a better performance in the *diff1* test (upper right), indicating that similarity of the cue array improves the change-detection performance in the Lin and Luck (2009) task (lower left), whereas it also shows that the similarity could deteriorate the performance in the *diff2* test (lower right).

is shared by all activity bumps and can be reallocated during WM delay (Bays and Husain, 2008). When the set size exceeds WM capacity, local excitation within a bump may be insufficient, and some activity bumps fade out, which leads to the reallocation of its memory resources to other bumps; this may result in excessive local excitation of some activity bumps and merging between them. The merging and fade-out phenomena are correlated, as a result of the "overload effect," and thus a limited number of activity bumps persist separately. Consequently, a continuous recurrent (attractor) neural network endowed with normalization exhibits a rich repertoire of dynamical effects compatible with the discrete-slot and shared-resource models. Furthermore, the normalization of neural activity is a general principle for sensory information processing (Treue et al., 2000; Reynolds and Heeger, 2009). Here we suggest that it is also a desirable property of WM circuits (Buschman et al., 2011).

Role of recurrent excitatory connection in limited WM capacity

Using a neural network with uniform connections onto and from interneurons, we differentially assessed the impact of recurrent excitatory connections on WM capacity. First, increasing J^+ enhances local iso-directional excitation, decreases long-range cross-directional excitation, and thus monotonically boosts WM capacity. While, an intermediate value of σ can maximize the WM capacity. A systematically analysis of the parameter space of J^+ and σ indicates that WM capacity is constrained between 2 and 7, consistent with human studies (Xu and Chun, 2006). Furthermore, the E-E connections strongly affect the amount of memory resources, as measured by the total width of activity bumps and the mean population firing rate of pyramidal cells. Using randomly versus uniformly distributed cues, we found that the normalization is independent of the configuration of external inputs. Therefore, for a given network connectivity, the total amount of memory resources is roughly fixed, and different external inputs lead to a different dynamical allocation of resources. Previously, Edin et al. (2009) showed that WM capacity is limited by lateral inhibition, and top-down excitation could rescue a fade-out activity bump. Our work is complementary, suggesting that the recurrent synaptic excitation greatly affects, perhaps even predominantly controls, the limited capacity of a WM circuit.

Similarity effect on change-detection tasks

In our model, a network could show confusable memory slots, using random cue arrays for a set size below capacity, which would result from merging of neural subpopulations storing different items. Merging skews the response offset distribution to the convergent side, and maintains high-similarity objects with poor precision. Consequently, merging causes low P_c as increasing set size in CDTs (Luck and Vogel, 1997; Wilken and Ma, 2004). However, for a given set size of cue arrays, with the different similarities, we found a counterintuitive phenomenon that similarity improves the performance, when the test is placed on the non-convergent side of merging (Lin and Luck, 2009). Johnson et al. (2009) provided a population firing-rate model leading to the same prediction, which had a specialized and tuned network scheme for the retrieval process; they applied the model to behavioral experiments where the discriminability index d' is larger with similar stimuli than disparate ones. Comparisons between two models are worthwhile. First, the prediction of their model originates from the proposed mechanism of the match-nonmatch decision circuit, rather than the WM retention per se. Second, fade-out is the exclusive mechanism for WM capacity in their model. When fade-out occurs, their model responded to nonmatch, whereas a more reasonable alternative is to respond randomly (since no memory trace is available to guide the response). Furthermore, our biologically-based spiking network (rather than an abstract population rate model) is required to elucidate the detailed circuit dynamics underlying the limited memory capacity during a retention delay.

To conclude, this study focused on delay-period dynamics of a WM circuit, which limits storage capacity for a single feature; the model can potentially be extended to a multi-feature version and used to study the resource allocation over different features (Fougnie et al., 2010). Although we did not explicitly investigate the influence of the delay duration on CDTs, from the result of DRTs, we could predict that the performance would decay as increasing the delay duration (Magnussen et al., 1996; Magnussen, 2000), e.g., in a sudden-death manner (Regan, 1985; Bennett and Cortese, 1996). Other factors may also contribute to determine the WM capacity, such as the role of selective attention during encoding stimulus items, i.e., bottleneck effect (Awh and Jonides, 2001), interactions of the distributed network perspectives of WM (Pessoa et al., 2002), overlaps of neural representation for different items (Warden and Miller, 2007) or synchronous oscillations (Siegel et al., 2009). Regardless, this work revealed and highlighted a rich repertoire of dynamical behaviors that unfold in time and underlie the limited capacity of a WM circuit. It shows that a shared-resource mechanism, using population coding in a continuous network, can nevertheless capture behavioral characteristics predicted by the discrete-slot model. Our work therefore contributes to resolving a major debate in the field, and shed new insights into the neurodynamical mechanism of WM capacity.

References

Amari S, Nakahara H (2005) Difficulty of singularity in population coding. Neural Comput 17:839–858.

Anderson DE, Vogel EK, Awh E (2011) Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. J Neurosci 31:1128–1138.

Awh E, Jonides J (2001) Overlapping mechanisms of attention and spatial working memory. Trends Cogn Sci 5:119–126.

Baddeley A (1992) Working memory. Science 255:556–559.

Basile BM, Hampton RR (2011) Monkeys recall and reproduce simple shapes from memory. Curr Biol 21:774–778.

Bays PM, Husain M (2008) Dynamic shifts of limited working memory resources in human vision. Science 321:851–854.

Bays PM, Catalao RFG, Husain M (2009) The precision of visual working memory is set by allocation of a shared resource. J Vis 9:7.1–7.11.

Bays PM, Singh-Curry V, Gorgoraptis N, Driver J, Husain M (2010) Integration of goal- and stimulus-related visual signals revealed by damage to human parietal cortex. J Neurosci 30:5968–5978.

Bennett PJ, Cortese F (1996) Masking of spatial frequency in visual memory depends on distal, not retinal, frequency. Vis Res 36:233–238.

Buschman TJ, Siegel M, Roy JE, Miller EK (2011) Neural substrates of cognitive capacity limitations. Proc Natl Acad Sci U S A 108:11252–11255.

Compte A, Brunel N, Goldman-Rakic PS, Wang XJ (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb Cortex 10:910–923.

Constantinidis C, Franowicz MN, Goldman-Rakic PS (2001) Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. J Neurosci 21:3646–3655.

Conway BR, Tsao DY (2009) Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. Proc Natl Acad Sci U S A 106:18034–18039.

Cowan N (2005) Working memory capacity. New York: Psychology.

Deneve S, Latham PE, Pouget A (1999) Reading population codes: a neural implementation of ideal observers. Nat Neurosci 2:740–745.

Edin F, Klingberg T, Johansson P, McNab F, Tegnér J, Compte A (2009) Mechanism for top-down control of working memory capacity. Proc Natl Acad Sci U S A 106:6802–6807.

Elmore LC, Ji Ma WJ, Magnotti JF, Leising KJ, Passaro AD, Katz JS, Wright AA (2011) Visual short-term memory compared in rhesus monkeys and humans. Curr Biol 21:975–979.

Engel TA, Wang XJ (2011) Same or different? A neural circuit mechanism of similarity-based pattern match decision making. J Neurosci 31:6982–6996.

Fisher N (1993) Statistical analysis of circular data. Cambridge, UK: Cambridge UP.

Fougnie D, Asplund CL, Marois R (2010) What are the units of storage in visual working memory? J Vis 10:27.

Fukuda K, Awh E, Vogel EK (2010) Discrete capacity limits in visual working memory. Curr Opin Neurobiol 20:177–182.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331–349.

Georgopoulos AP, Lurito JT, Petrides M, Schwartz AB, Massey JT (1989) Mental rotation of the neuronal population vector. Science 243:234–236.

Goldman-Rakic PS (1995) Cellular basis of working memory. Neuron 14:477–485.

Hollingworth A (2003) Failures of retrieval and comparison constrain change detection in natural scenes. J Exp Psychol Hum Percept Perform 29:388–403.

Huang L (2010) Visual working memory is better characterized as a distributed resource rather than discrete slots. J Vis 10:8.

Jahr CE, Stevens CF (1990) Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. J Neurosci 10:3178–3182.

Johnson JS, Spencer JP, Luck SJ, Schöner G (2009) A dynamic neural field model of visual working memory and change detection. Psychol Sci 20:568–577.

Lin PH, Luck SJ (2009) The influence of similarity on visual working memory representations. Vis Cogn 17:356–372.

Lisman JE, Idiart MA (1995) Storage of 7+/− 2 short-term memories in oscillatory subcycles. Science 267:1512–1515.

Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. Nature 390:279–281.

Machizawa MG, Driver J (2011) Principal component analysis of behavioural individual differences suggests that particular aspects of visual working memory may relate to specific aspects of attention. Neuropsychologia 49:1518–1526.

Macoveanu J, Klingberg T, Tegnér J (2006) A biophysical model of multiple-item working memory: a computational and neuroimaging study. Neuroscience 141:1611–1618.

Magnussen S (2000) Low-level memory processes in vision. Trends Neurosci 23:247–251.

Magnussen S, Greenlee MW, Thomas JP (1996) Parallel processing in visual short-term memory. J Exp Psychol Hum Percept Perform 22:202–212.

Miller GA (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol Rev 63:81–97.

Mitroff SR, Simons DJ, Levin DT (2004) Nothing compares 2 views: change blindness can occur despite preserved access to the changed information. Percept Psychophys 66:1268–1281.

Pashler H (1988) Familiarity and visual change detection. Percept Psychophys 44:369–378.

Pessoa L, McKenna M, Gutierrez E, Ungerleider LG (2002) Neural processing of emotional faces requires attention. Proc Natl Acad Sci U S A 99:11458–11463.

Pouget A, Dayan P, Zemel R (2000) Information processing with population codes. Nat Rev Neurosci 1:125–132.

Raffone A, Wolters G (2001) A cortical mechanism for binding in visual working memory. J Cogn Neurosci 13:766–785.

Rao SG, Williams GV, Goldman-Rakic PS (1999) Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization in PFC. J Neurophysiol 81:1903–1916.

Regan D (1985) Storage of spatial-frequency information and spatial frequency discrimination. J Opt Soc Am A 2:619–621.

Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168–185.

Siegel M, Warden MR, Miller EK (2009) Phase-dependent neuronal coding of objects in short term memory. Proc Natl Acad Sci U S A 106:21341–21346.

Treue S, Hol K, Rauber HJ (2000) Seeing multiple directions of motion-physiology and psychophysics. Nat Neurosci 3:270–276.

Troyer TW, Miller KD (1997) Physiological gain leads to high ISI variability in a simple model of a cortical regular spiking cell. Neural Comput 9:971–983.

Tuckwell HC (1988) Introduction to theoretical neurobiology Cambridge, UK: Cambridge UP.

Vogel EK, Woodman GF, Luck SJ (2001) Storage of features, conjunctions and objects in visual working memory. J Exp Psychol Hum Percept Perform 27:92–114.

Wang M, Gamo NJ, Yang Y, Jin LE, Wang XJ, Laubach M, Mazer JA, Lee D, Arnsten AFT (2011) Neuronal basis of age-related working memory decline. Nature 476:210–213.

Wang XJ (1999) Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. J Neurosci 19:9587–9603.

Warden MR, Miller EK (2007) The representation of multiple objects in prefrontal neuronal delay activity. Cereb Cortex 17:i41–50.

Wilken P, Ma WJ (2004) A detection theory account of change detection. J Vis 4:1120–1135.

Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 440:91–95.

Zemel RS, Dayan P, Pouget A (1998) Probabilistic interpretation of population codes. Neural Comput 10:403–430.

Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. Nature 453:233–235.

Zhang W, Luck SJ (2009) Sudden death and gradual decay in visual working memory. Psychol Sci 20:423–428.