Hierarchical recruitment of competition alleviates working memory overload in a fronto-parietal model

Dominic Standage, Martin Paré and Gunnar Blohm

Centre for Neuroscience Studies, Queen's University

Correspondence

Dr. Dominic Standage

Queen's University

Botterell Hall, Room 230

Kingston, Ontario, K7L 3N6, Canada

standage@queensu.ca

Abstract

Working memory capacity has long been the subject of intense research interest, but few studies have systematically investigated the dependence of capacity on memory load that exceeds our retention abilities. Under this real-world scenario, WM performance typically declines beyond a critical load among low-capacity subjects, a phenomenon known as working memory overload. We used a fronto-parietal cortical model to test the hypothesis that high-capacity subjects select a manageable number of items for storage, thereby avoiding overload. The model accounts for behavioural and electrophysiological data from high-capacity subjects in a parameter regime where competitive encoding in its prefrontal network selects items for storage, inter-aerial projections sustain their representations after stimulus offset, and weak dynamics in its parietal network limit their mutual interference. Violation of these principles accounts for these data among low-capacity subjects, implying that poor working memory perfor-

1

mance reflects poor control over fronto-parietal circuitry, and making testable predictions for experiments.

1 Introduction

The transient retention and manipulation of information is central to cognition and is known as working memory. The modular nature of working memory has long been recognized [see Baddeley (2012)], with visual working memory (WM) receiving considerable attention for several decades. Much of this research has focused on the limited capacity of WM, typically three or four items in healthy young adults [see Luck and Vogel (2013)]. Because capacity is a reliable predictor of cognitive ability more generally [see Unsworth, Fukuda, Awh, and Vogel (2014)], understanding its neural basis is a fundamental goal of cognitive neuroscience.

In the laboratory, WM capacity is estimated by varying the number of items for retention over a memory delay (WM load), but little emphasis has been given to the dependence of capacity on load that exceeds our retention abilities. This real-world scenario has long been of concern to instructional designers (Sweller, 1988; Merrienboër & Sweller, 2005), who consider the avoidance of 'overload' a fundamental principle of effective design. Consistent with this concern, recent WM studies have shown a decrease in capacity beyond a critical load (Chee & Chuah, 2007; Xu, 2007; Cusack, Lehmann, Veldsman, & Mitchell, 2009; Linke, Vicente-Grabovetsky, Mitchell, & Cusack, 2011; Matsuyoshi, Osaka, & Osaka, 2014; Fukuda, Woodman, & Vogel, 2015), referred to as WM overload (Matsuyoshi et al., 2014).

It is widely believed that WM storage is supported by 'attractor states' in neocortex, where regenerative excitation sustains neural firing after stimulus offset, kept in check by feedback inhibition [see Wang (2001)]. In models of this kind, overload is a consequence of the competition imposed by inhibition (Edin et al., 2009), but not all subjects show overload [*e.g.* Fukuda et al. (2015)] and among those who do, overload is not typically as pronounced as in these models [*e.g.* Wei, Wang, and Wang (2012); Section 3.1 here]. The occurrence of overload in attractor models, however, assumes that all items in a stimulus array are encoded for retention. Thus, a viable strategy for managing overload is to limit the number of items selected for storage (Cusack et al., 2009). We hypothesize that this selection process is implemented by strong competitive dynamics during stimulus encoding.

2

To test our hypothesis, we simulated a multiple-item WM task with bio-physical models of posterior parietal cortex (PPC) and lateral prefrontal cortex (PFC), both of which are extensively correlated with WM [see Curtis (2006); Funahashi (2013)]. PPC is hypothesized to be the hub of distributed WM storage (Palva, Monto, Kulashekhar, & Palva, 2010; Christophel, Hebart, & Haynes, 2012; Salazar, Dotson, Bressler, & Gray, 2012) and is well characterized by neural data from visual tasks [see Goldberg, Bisley, Powell, and Gottlieb (2006); Serences and Yantis (2006)], so we first sought to determine whether competitive encoding in a local-circuit PPC model could alleviate overload in a manner consistent with behavioural data from high-performing WM subjects, and with neural data from visual tasks. We reasoned that any inconsistencies between the model and these data might point to the role of PPC in distributed storage, and by extension, to the roles of its bidirectional connectivity with PFC. Next, we did the same thing with a hierarchical model of PPC and PFC, reasoning that the values of biophysical parameters required to account for the data might identify specific computational principles for distributed storage. If so, then violation of these principles should account for behavioural data from low-performing subjects. Thus, we evaluated our hierarchical model for its ability to account for capacity and overload among subject groups distinguished according to these measures. Finally, we sought to make predictions to test our hypothesis experimentally, approximating electroencephalogram (EEG) recordings over PPC and PFC. We reasoned that if this approximation could account for the different EEG profiles of high- and how-performing subject groups during the storage of memoranda (Fukuda et al., 2015), then its profile during stimulus encoding would be a testable prediction for our hypothesis.

2   Methods

Our local-circuit PPC model (the PPC-only model, Figure 1A) is a network of simulated pyramidal neurons and inhibitory interneurons, connected by AMPA, NMDA and GABA receptor conductance synapses (AMPAR, NMDAR and GABAR). Our hierarchical model (the PPC-PFC model, Figure 1D) is comprised of two such networks, bidirectionally connected. In both models, intrinsic synaptic connectivity within and between classes of neuron was structured according to *in vitro* data, as was the connectivity between networks in the PPC-PFC model. Our chosen parameters and their values

3

are justified in Section 2.6.

We ran simulations of a visuospatial WM task with both models, where the number of items for retention ranged from 1 to 8. On each trial of the task, a stimulus interval was preceded by a pre-trial interval and followed by a delay of 1s. In both models, the items were provided to the PPC network during the stimulus interval and their accurate retention (or otherwise) was determined from its activity at the end of the delay (Sections 2.4 and 2.5).

2.1 The network model

Each local circuit is a fully connected network of leaky integrate-and-fire neurons (Tuckwell, 1988), comprised of $N^p = 400$ simulated pyramidal neurons and $N^i = 100$ fast-spiking inhibitory interneurons. Each model neuron is described by

$$C_m^{\{p,i\}} \frac{dV}{dt} = -g_L^{\{p,i\}} (V - E_L^{\{p,i\}}) - I, \tag{1}$$

where C_m is the membrane capacitance of the neuron, g_L is the leakage conductance, V is the membrane potential, E_L is the equilibrium potential, and I is the total input current. When V reaches a threshold ϑ_v, it is reset to V_{res}, after which it is unresponsive to its input for an absolute refractory period of τ_{ref}. Here and below, superscripts p and i refer to pyramidal neurons and interneurons respectively, indicating that parameter values are assigned separately to each class of neuron.

The total input current at each neuron is given by

$$I = I^{sel} + I^{rec} + I^{hier} + I^{back}, \tag{2}$$

where I^{sel} is stimulus-selective synaptic current (set to 0 for all neurons in the PFC network and for interneurons in the PPC network), I^{rec} is recurrent (intrinsic) synaptic current, I^{hier} is hierarchical (inter-aerial) synaptic current projected to PPC from PFC and *vice versa* (set to 0 in single-circuit simulations) and I^{back} is background current. Of these currents, I^{sel}, I^{rec} and I^{hier} are comprised of synaptic currents, and I^{back} is comprised of synaptic current and injected current. Synaptic currents driven by pyramidal neuron spiking are mediated by simulated AMPA receptor (AMPAR) and/or NMDA receptor (NMDAR) conductances, and synaptic currents driven by interneuron spiking are mediated by simulated GABA receptor (GABAR)

4
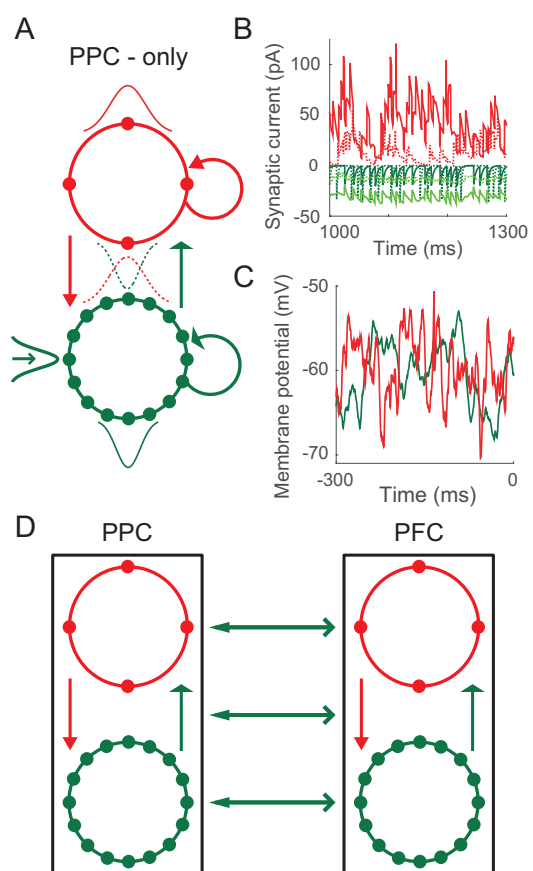
conductances. For AMPAR and GABAR currents, synaptic activation (the proportion of open channels) is defined by

$$\frac{dg^a_{AMPA}}{dt} = -\frac{g^a_{AMPA}}{\tau^{\{p,i\}}_{AMPA}} + \delta(t - t_f)$$

$$\frac{dg^a_{GABA}}{dt} = -\frac{g^a_{GABA}}{\tau^{\{p,i\}}_{GABA}} + \delta(t - t_f),$$

(3)

where τ_{AMPA} and τ_{GABA} are the time constants of AMPAR and GABAR deactivation respectively, δ is the Dirac delta function, t_f is the time of firing of a pre-synaptic neuron and superscript a indicates that synapses are activated by different sources of spiking activity (selective, recurrent, hierarchical and background). NMDAR activation has a slower rise and decay and is described by

$$\frac{dg^a_{NMDA}}{dt} = -\frac{g^a_{NMDA}}{\tau^{\{p,i\}}_{NMDA}} + \alpha_{NMDA} \cdot \omega_{NMDA}(1 - g^a_{NMDA}),$$

(4)

Figure 1 *(preceding page)*: (A) Schematic of the PPC-only model. Solid circles depict pyramidal neurons (green) and inhibitory interneurons (red), arranged periodically by their connectivity structures. The 4-to-1 ratio of pyramidal neurons to interneurons preserves their population sizes in the model. Arced and straight arrows depict synaptic connectivity within and between classes of neuron respectively. Thin Gaussian curves depict the structure of this connectivity (within, solid; between, dotted). The Gaussian curve on the left depicts the RF of a pyramidal neuron. Red, open green and wide green arrows depict GABAR, AMPAR-only, and AMPAR-NMDAR synapses respectively. (B) Synaptic currents onto a pyramidal neuron (solid) and an interneuron (dotted) during the delay interval of the 1-item memory task. Red, light green and dark green curves show GABAR, AMPAR and NMDAR currents respectively. (C) Membrane potential of a pyramidal neuron and an interneuron during the pre-trial interval. (D) Schematic of the PPC-PFC model. The PPC network is identical to the PPC-only model. The PFC network differs only in the strength of GABAR conductance onto pyramidal neurons. Open and thin arrows depict topographically aligned feed-forward and feedback projections, mediated by AMPARs and NMDARs respectively. See Section 2 for description of the model.

6

where τ_{NMDA} is the time constant of receptor deactivation and α_{NMDA} controls the saturation of NMDAR channels at high pre-synaptic spike frequencies. The slower opening of NMDAR channels is captured by

$$\frac{d\omega_{NMDA}}{dt} = -\frac{\omega_{NMDA}}{\tau_\omega} + \delta(t - t_f), \tag{5}$$

where τ_ω and η determine the rate of channel opening and the voltage-dependence of NMDARs respectively.

Recurrent synaptic current to each neuron j is defined by

$$I_j^{rec} = I_{AMPA,j}^{rec} + I_{NMDA,j}^{rec} + I_{GABA,j}^{rec}$$

$$I_{AMPA,j}^{rec} = \sum_k 1/\gamma_g^{\{ppc,pfc\}} \cdot G_{AMPA}^{\{p,i\}} \cdot g_{AMPA,k}^{rec}(V_j - V_E) \cdot W_{j,k}^{rec|pp,ip}$$

$$I_{NMDA,j}^{rec} = \sum_k 1/\gamma_g^{\{ppc,pfc\}} \cdot G_{NMDA}^{\{p,i\}} \cdot g_{NMDA,k}^{rec}(V_j - V_E) \cdot W_{j,k}^{rec|pp,ip} \cdot \eta_j$$

$$I_{GABA,j}^{rec} = \sum_k 1/\gamma_g^{\{ppc,pfc\}} \cdot G_{GABA}^{\{p,i\}} \cdot g_{GABA,k}^{rec}(V_j - V_I) \cdot W_{j,k}^{rec|pi,ii}, \tag{6}$$

where $\gamma_g^{\{ppc,pfc\}}$ is a scale factor controlling the relative strength of extrinsic and intrinsic synaptic conductance (subscripts ppc and pfc indicate that its value is assigned separately to each network, see Section 2.6); G_{AMPA}, G_{NMDA} and G_{GABA} are the respective strengths of AMPAR, NMDAR and GABAR conductance; V_E is the reversal potential for AMPARs and NMDARs, and V_I is the reversal potential for GABARs; $g_{AMPA,k}^{rec}$, $g_{NMDA,k}^{rec}$ and $g_{GABA,k}^{rec}$ are the activation of AMPAR, NMDAR and GABAR receptors respectively by pre-synaptic neurons k; and matrices $W^{rec|pp,ip}$ and $W^{rec|pi,ii}$ scale conductance strength or *weight* according to the connectivity structure of the network. This structure depends on the class of neuron receiving and projecting spiking activity, where superscripts pp, ip, pi and ii denote connections to pyramidal neurons from pyramidal neurons, to interneurons from pyramidal neurons, to pyramidal neurons from interneurons, and to interneurons from interneurons respectively. For each of these structures $s \in \{pp, ip, pi, ii\}$, $W^{rec|s}$ is a Gaussian function of the distance between periodically-arranged neurons, where the weight $W_{j,k}^{rec|s}$ to neuron j from neuron k is given by

$$W_{j,k}^{rec|s} = e^{-d^2/2\sigma_{rec|s}^2} \cdot (1 - \zeta_{rec|s}) + \zeta_{rec|s}. \tag{7}$$

7

The distance between neurons is defined by $d = min(|j-k|\Delta x, 2\pi - |j-k|\Delta x)$ for $W^{rec|pp}$ and $W^{rec|ii}$, and by $d = min(|j-z|\Delta x, 2\pi - |j-z|\Delta x)$ for $W^{rec|ip}$ and $W^{rec|pi}$, where $z = N^p/N^i \cdot k$ for $W^{rec|ip}$ and $z = N^i/N^p \cdot k$ for $W^{rec|pi}$. $\Delta x = 2\pi/N^{\{p,i\}}$ is a scale factor and $\sigma_{rec|s}$ determines the spatial extent of connectivity. Parameter $\zeta_{rec|s}$ allows the inclusion of a baseline weight, with the function normalized to a maximum of 1 $(0 \leq \zeta_{rec|s} < 1)$.

2.2 Background activity

For each neuron, *in vivo* cortical background activity is simulated by current I^{back}, defined by

$$I^{back} = I^{back,syn} + I^{back,inj}, \tag{8}$$

where $I^{back,syn}$ is driven by synaptic bombardment and $I^{back,inj}$ is noisy current injection. The former is generated by AMPAR synaptic activation, where independent, homogeneous Poisson spike trains are provided to all neurons at rate μ_{back}. $I^{back,syn}$ is therefore defined by

$$I^{back,syn} = \gamma_g^{\{ppc,pfc\}} \cdot \gamma_g^{ext} \cdot G_{AMPA}^{\{p,i\}} \cdot g_{AMPA}^{back}(V - V_E), \tag{9}$$

where γ_g^{ext} is a scale factor and g_{AMPA}^{back} is given in Equation 3.

For $I^{back,inj}$, we used the point-conductance model by (Destexhe, Rudolph, Fellous, & Sejnowski, 2001):

$$I^{back,inj} = g_e(t)(V - V_E) + g_i(t)(V - V_I). \tag{10}$$

The time-dependent excitatory and inhibitory conductances $g_e(t)$ and $g_i(t)$ are updated at each timestep Δt according to

$$g_e(t + \Delta t) = g0_e + [g_e(t) - g0_e] \cdot e^{-\Delta t/\tau_e} + A_e \Upsilon \tag{11}$$

and

$$g_i(t + \Delta t) = g0_i + [g_i(t) - g0_i] \cdot e^{-\Delta t/\tau_i} + A_i \Upsilon \tag{12}$$

respectively, where $g0_e$ and $g0_i$ are average conductances, τ_e and τ_i are time constants, and Υ is normally distributed random noise with 0 mean and unit standard deviation. Amplitude coefficients A_e and A_i are defined by

8

$$A_e = \sqrt{\frac{D_e \tau_e}{2} \left[1 - exp \left(\frac{-2\Delta t}{\tau_e} \right) \right]} \qquad (13)$$

and

$$A_i = \sqrt{\frac{D_i \tau_i}{2} \left[1 - exp \left(\frac{-2\Delta t}{\tau_i} \right) \right]} \qquad (14)$$

respectively, where $D_e = 2\sigma_e^2/\tau_e$ and $D_i = 2\sigma_i^2/\tau_i$ are noise 'diffusion' coefficients. See (Destexhe et al., 2001) for the derivation of these equations.

2.3 The PPC-PFC model

In the hierarchical model, inter-aerial projections mediate synaptic currents $I^{hier} \in \{I^{ff}, I^{fb}\}$, where superscript ff (fb) refers to feedforward (feedback) currents onto neurons in the PFC (PPC) network from neurons in the PFC (PPC) network. Only pyramidal neurons make inter-aerial projections, where feedforward projections are mediated by AMPARs (onto pyramidal neurons only) and feedback projections are mediated by NMDARs (onto pyramidal neurons and interneurons). Feed-forward currents at each pyramidal neuron j in the PFC network are defined by

$$I_j^{ff} = \sum_k \gamma_g^{ext} \cdot G_{AMPA}^p \cdot g_{AMPA,k}^{ff}(V_j - V_E) \cdot W_{j,k}^{ff} \qquad (15)$$

where γ_g^{ff} is a scale factor, $g_{AMPA,k}^{ff}$ is the activation of AMPAR receptors by pre-synaptic pyramidal neurons k in PPC, and matrix $W_{j,k}^{ff} = exp(-d^2/2\sigma_{ff}^2)$ scales conductance strength according to the structure of FF connectivity. Constant d is given above for recurrent synaptic structure $W^{rec|pp}$, where the two networks are considered to be topographically aligned, *i.e.* the lateral distance between neuron j in PFC and neuron k in PPC is the same as that between neurons j and k within either network.

Feedback currents at each neuron j in the PPC network are defined by

$$I_j^{fb} = \sum_k \gamma_g^{fb} \cdot G_{NMDA}^{\{p,i\}} \cdot g_{NMDA,k}^{fb}(V_j - V_E) \cdot W_{j,k}^{fb|pp,ip} \qquad (16)$$

where γ_g^{fb} is a scale factor, $g_{NMDA,k}^{fb}$ is the activation of NMDAR receptors by pre-synaptic pyramidal neurons k in PFC, and matrices $W^{fb|pp,ip}$ scale conductance strength according to the structure of FB connectivity. Each

of these structures $s \in \{pp, ip\}$ is defined by $W_{j,k}^{fb|s} = e^{-d^2/2\sigma_{fb|s}^2}$, where d is defined for $W^{fb|pp}$ and $W^{fb|ip}$ in the same way as for $W^{rec|pp}$ and $Wrec|ip$ respectively above.

2.4 Simulated working memory task

We simulated the stimulus array by providing independent, homogeneous Poisson spike trains to all pyramidal neurons j in the PPC network, where spike rates were drawn from a normal distribution with mean μ_{sel} corresponding to the centre of a Gaussian response field (RF) defined by $W_{j,k}^{rf} = exp(-d^2/2\sigma_{rf}^2)$. Constant d is given above for recurrent synaptic structure $W^{rec|pp}$, σ_{rf} determines the width of the RF and subscript k indexes the neuron at the RF centre. Spike response adaptation by upstream visually responsive neurons was modelled by a step-and-decay function

$$\mu_{sel}(t) = \begin{cases} (\mu_{init} - \mu_{init}/\mu_{div}) \ e^{-(t-t_{vrd})/\tau_\mu} + \mu_{init}/\mu_{div} & \text{for } t > t_{vrd} \\ 0 & \text{for } t \le t_{vrd} \end{cases} \quad (17)$$

where μ_{init} determines the initial spike rate, μ_{div} determines the asymptotic rate, τ_μ determines the rate of upstream response adaptation, and t_{vrd} is a visual response delay. These selective spike trains were provided for 300ms, following the 300ms pre-trial interval and followed by a 1000ms delay (*e.g.* Figure 2A). The stimuli were mediated by AMPARs only, so for all pyramidal neurons j in the PPC network,

$$I_j^{sel} = \gamma_g^{ppc} \cdot \gamma_g^{ext} \cdot G_{AMPA}^p \cdot g_{AMPA,j}^{sel}(V_j - V_E) \cdot W_{j,k}^{rf}. \quad (18)$$

All simulations were run with the standard implementation of Euler's forward method, where the timestep was $\Delta t = 0.25$ms.

2.5 Determining working memory performance

We ran 100 trials with $1 - 8$ stimuli (henceforth the n-item memory task; $1 \le n \le 8$). To determine WM performance on each trial, spike density functions (SDFs) were calculated for all pyramidal neurons in the network by convolving their spike trains with a rise-and-decay function

$$\frac{(1 - e^{-t/\tau_f}) \cdot e^{-t/\tau_d}}{\frac{\tau_d^2}{\tau_r + \tau_d}} \quad (19)$$

10

where t is the time following stimulus onset and $\tau_r = 1$ms and $\tau_d = 20$ms are the time constants of rise and decay respectively (Thompson, Hanes, Bichot, & Schall, 1996; Standage & Paré, 2011). On each n-item trial, we calculated the mean of the SDFs over the last 300ms of the delay, obtaining the average activity over the network, and then partitioned the network into n equal regions. The location of each item was centred within each region. We then fit the mean activity in each region with a Gaussian function with four parameters: the height of the peak, the position of the peak, the standard deviation (controlling width), and the height that is approached asymptotically from the peak. An item was considered accurately stored if the fitted Gaussian satisfied three criteria: the height parameter h exceeded 30Hz, the difference between h and the fitted asymptote on both sides of the peak exceeded $h/2$Hz, and the position parameter was within $\Delta c = 10$ degrees of the centre of the RF for that item. For the first criterion, we chose 30Hz because in electrophysiological experiments with macaque monkeys (Johnston et al, SfN abstracts, 2009), memory trials were discarded if the recorded PPC neuron did not fire at least 10 spikes during the last 300ms of the delay. The second criterion dictates that items are only considered accurately stored if the population response is discriminable. The third criterion ensures that the memory of the location of the item is close to the actual location.

2.6 Parameter values

In setting parameter values in the two models, our aim was to justify every value by anatomical and physiological data, thus constraining our choices as much as possible, and then to use control parameters to explore the models' performance on simulated WM tasks. In the PPC-only model, our control parameter was γ_g^{ppc} (Equations 6, 9 and 18), governing the relative strengths of extrinsic and intrinsic synaptic conductance and therefore the strength of recurrent processing (see Section 3.1). In the PPC-PFC model, the PPC network was identical to the PPC-only model, and the PFC network was identical to the PPC network, except for the strength of GABAR conductance onto pyramidal neurons. Parameter values for inter-aerial projections in the PPC-PFC model are described below (Section 2.6.1). Control parameters for the PPC-PFC model are described in Section 3.2.

For cellular parameters, we used standard values for integrate-and-fire neurons in cortical simulations (Compte, Brunel, Goldman-Rakic, & Wang, 2000), justified by electrophysiological data in earlier, related work (Troyer &

11

Miller, 1997; Wang, 1999). These values are $C_m^p = 0.5$nF, $g_L^p = 25$nS, $E_L^p = -70$mV, $\vartheta_v^p = -50$mV, $V_{res}^p = -60$mV and $\tau_{ref}^p = 2$ms; and $C_m^i = 0.2$nF, $g_L^i = 20$nS, $E_L^i = -70$mV, $\vartheta_v^i = -50$mV, $V_{res}^i = -60$mV and $\tau_{ref}^i = 1$ms. Likewise, synaptic reversal potentials are $V_E = 0$mV and $V_I = -70$mV, and the parameters governing the opening and saturation of NMDARs are $\tau_\omega = 2$ms and $\alpha_{NMDA} = 0.5$kHz respectively (Compte et al., 2000). The voltage-dependence of NMDARs is given by $\eta = 1/[1 + Mg \cdot exp(-0.062 \cdot V)/3.57]$, where $Mg = 1$mM is the extracellular Magnesium concentration and V is measured in millivolts (Jahr & Stevens, 1990).

In setting parameters for the conductance strengths and time constants of decay of AMPARs and NMDARs, we followed Standage, You, Wang, and Dorris (2013), emphasising fast inhibitory recruitment in response to slower excitation [see Povysheva et al. (2006) for discussion]. For AMPARs, $G_{AMPA}^p = 0.2$nS, $G_{AMPA}^i = 2 \cdot G_{AMPA}^p$, $\tau_{AMPA}^p = 4$ms and $\tau_{AMPA}^i = \tau_{AMPA}^p/2$; and for NMDARs, $G_{NMDA}^p = 4$nS, $G_{NMDA}^i = G_{NMDA}^p/2$, $\tau_{NMDA}^p = 100$ms and $\tau_{NMDA}^i = \tau_{NMDA}^p/2$. These values produce fast-decaying AMPAR currents on the order of 10pA (Angulo, Rossier, & Audinat, 1999; Desai, Cudmore, Nelson, & Turrigiano, 2002) that are stronger and shorter-lived onto inhibitory interneurons than onto pyramidal neurons (Hestrin, 1993; J.McBain & Fisahn, 2001; Hull, Isaacson, & Scanziani, 2009), and slow-decaying NMDAR currents on the order of 10pA (Berretta & Jones, 1996; Angulo et al., 1999) that are stronger and longer-lived at synapses onto pyramidal neurons than onto inhibitory interneurons (Hull et al., 2009). For GABARs, $G_{GABA}^p = 1.5$nS and $G_{GABA}^i = G_{GABA}^p/2$, producing GABAR currents several times stronger than the above excitatory currents, where the stronger conductance at synapses onto pyramidal neurons captures their greater prevalence of GABARs (Markram et al., 2004). GABAR time constants were set to $\tau_{GABA}^p = \tau_{GABA}^i = 10$ms (Salin & Prince, 1996; Xiang, Huguenard, & Prince, 1998). Example synaptic currents are shown in Figure 1D.

The connectivity structures $W^{rec|pp,ip,pi,ii}$ capture the probability of lateral synaptic contact within and between classes of neurons in local cortical circuitry (Wilson & Cowan, 1973; Somers, Nelson, & Sur, 1995). A considerable volume of data indicates that the probability of lateral synaptic contact between cortical pyramidal neurons is normally distributed with mean 0 and half-width of ~ 0.25mm (Hellwig, 2000; Berger, Perin, Silberberg, & Markram, 2009; Voges, Schüz, Aertsen, & Rotter, 2010). Thus, $\sigma_{rec|pp}$ corresponds to 0.25mm, determining the size of the cortical region being modelled, and $\zeta_{rec|pp} = 0$. We are unaware of any data suggesting that the lateral pro-

12

jections of pyramidal neurons target basket cells differently than they target other pyramidal neurons, so we set $\sigma_{rec|ip} = \sigma_{rec|pp}$ and $\zeta_{rec|ip} = \zeta_{rec|pp}$. Arguably, $\sigma_{rec|ip}$ should be narrower than $\sigma_{rec|pp}$, since the dendritic trees of basket cells are less extensive than those of pyramidal neurons, but setting these parameters to equal values supported more stable network dynamics, *i.e.* it furnished sufficient local-circuit inhibition for the model to simulate the experimental tasks without modifications to other parameter values.

For connectivity structures $W^{rec|pi,ii}$, values for $\sigma_{rec|pi,ii}$ and $\zeta_{rec|pi,ii}$ are justified by four premises: firstly, we assume that basket cells are a major source of lateral inhibition (Krimer & Goldman-Rakic, 2001) and we limit our focus to this class of inhibitory interneuron; secondly, basket cells synapse onto the somatic and perisomatic regions of their targets [see Markram et al. (2004)]; thirdly, the axons of basket cells contact their targets indiscriminately throughout the range of their ramifications (Packer & Yuste, 2011); and fourthly, the basket cell population can be divided into small (local arbour), medium (medium arbour) and large (wide arbour) cells in equal proportion, *i.e.* one third each (Krimer et al., 2005). Under the first and second premises, we do not need to consider the dendritic morphology of the targets of inhibitory interneurons, so we set $\sigma_{rec|pi} = \sigma_{rec|pp}$. Under the second and third premises, we assume a uniform synaptic distribution for inhibitory targets, where the axonal ramifications of small, medium and large basket cells cover progressively larger areas (Krimer & Goldman-Rakic, 2001; Krimer et al., 2005), with large basket cells (LBC) covering the entire local circuit (Kisvárday, Beaulieu, & Eysel, 1993; Markram et al., 2004). We therefore approximate this connectivity structure by setting $\sigma_{rec|pi} = \sigma_{rec|ii} = 2 \cdot \sigma_{rec|pp}$ and $\zeta_{rec|pi} = \zeta_{rec|ii} = 1/3$, where the former corresponds to a half-width of ~ 0.5mm [*cf.* Kisvárday et al. (1993); Krimer and Goldman-Rakic (2001); Krimer et al. (2005)] and the latter refers to the $1/3$ proportion of LBCs. This approach to determining inhibitory connectivity parameters is depicted in Figure 1B. We set $\sigma_{rec|pp} = 0.2$ because this value supported the simultaneous representation of 5 simulated visual stimuli, corresponding to the upper limit on human WM capacity, *i.e.* 4 ± 1 items (Luck & Vogel, 1997; Cowan, 2001). Finally, we set $\sigma_{rec|ii} = \sigma_{rec|pi}$ because LBCs make extensive contacts onto one another over the full range of their axonal ramifications (Kisvárday et al., 1993). Note that we do not attribute biological significance to the spatial periodicity of the network. Rather, this arrangement allows the implementation of $W^{rec|pp,ip,pi,ii}$ with all-to-all connectivity without biases due to asymmetric lateral interactions between neurons, and further

13

captures the topographic mapping of spatially periodic stimuli in many visual [*e.g.* Thomas and Paré (2007)] and WM [*e.g.* Funahashi, Bruce, and Goldman-Rakic (1989); Matsuyoshi et al. (2014)] tasks. In Results section 'Feedback inhibition underlies slot-like capacity and resource-like coding', we eliminated broad inhibition by setting $\zeta_{rec|pi}$ and $\zeta_{rec|ii}$ to 0, and we increased the strength of local feedback inhibition by setting G^p_{GABA} and G^i_{GABA} to 3.5nS and 1.75nS respectively.

In setting parameter values for background activity in each network, we initially omitted background synaptic input $I^{back,syn}$ and followed the data by Fellous, Rudolph, Destexhe, and Sejnowski (2003) to produce $I^{back,inj}$, where $g0_e = 5$nS and $g0_i = 25$nS, $\tau_e = 2.5$ms, $\tau_i = 10$ms, $\sigma_e = 5$nS and $\sigma_i = 12.5$nS. Because the average inhibitory background conductance $g0_i$ is five times the average excitatory background conductance $g0_e$ [see Destexhe (2010)], our simulated pyramidal neurons did not respond adequately to selective stimuli under these parameter values. We therefore reduced the average conductances by a factor of two, setting $g0_e = 2.5$nS, retaining the ratio of inhibitory to excitatory conductance strength $g0_i = 5 \cdot g0_e = 12.5$nS, and simulating the 'other half' of upstream cortical background activity by providing independent, homogeneous Poisson spike trains to all neurons in the network. As such, we assumed that each neuron forms $\sim 10,000$ synapses with upstream cortical neurons (Douglas, Markram, & Martin, 2004), and that by dividing $g0_e$ and $g0_i$ by two, we were effectively omitting $\sim 5,000$ background inputs. We therefore approximated 5000 upstream cortical neurons firing at 1Hz each by setting the rate of background Poisson spike trains to $\mu_{back} = 500$Hz and setting the extrinsic synaptic scale factor to $\lambda = 10$, trading temporal summation for spatial summation (Prescott & De Koninck, 2003; Standage et al., 2013). As noted above, background spike trains were provided to all pyramidal neurons and interneurons in each network, mediated by AMPARs on the assumption that spike trains converging on PPC (an association cortical area) are predominantly ascending. Evidence for AMPAR-mediated ascending activity is provided by Self, Kooijmans, Supèr, Lamme, and Roelfsema (2012). This approach simultaneously released the network model from the overly-strong background inhibitory currents and implemented an established, biologically plausible form of gain modulation [balanced background inputs (Chance, Abbott, & Reyes, 2002)], rendering the PPC network responsive to simulated visual stimuli. Note that our parameter values for background current injection ($g0_e$, $g0_i$, τ_e, τ_i, σ_e and σ_i) were based on recordings from pyramidal neurons (Fellous et al., 2003), but

14

since we are unaware of any data to guide these parameters for inhibitory interneurons, we assigned them the same values for all neurons. The effect of this background activity on the membrane potential of a pyramidal neuron and an interneuron is shown in Figure 1E.

For the target stimuli, the width of RFs was determined by $\sigma_{rf} = \sigma_{pp}/2$. This narrow width captures the less-extensive dendritic branching in cortical (input) layer 4 compared to layers 2/3 and 5 (see above for justification of lateral connectivity in the model). The initial spike rate at the RF centre was $\mu_{init} = 10,000/\gamma_g$Hz, which (for $\gamma_g = 1$) can be equated with *e.g.* 100 upstream, visually-responsive neurons firing at 100Hz each, given our use of homogeneous, independent Poisson spike trains. Note, however, that the synaptic scale factor $\lambda = 10$ probably renders this spike rate unrealistically high, since it implies *e.g.* 1000 upstream neurons firing at 100Hz. Nonetheless, the high initial spike rate ensured a rapid-onset, high-rate visual response in the network for all processing regimes furnished by control parameter γ_g, as observed experimentally [*e.g.* Paré and Wurtz (1997); Thomas and Paré (2007); Churchland, Kiani, and Shadlen (2008)]. Upstream, visual response adaptation was simulated by $\mu_{div} = 10$ and $\tau_\mu = 50$ms. The former is somewhat extreme, but allowed the rate of the initial population response in PPC to exceed the steady state response on the visual task for all values of γ_g [*e.g.* Paré and Wurtz (1997); Churchland et al. (2008)]. Our use of γ_g as a denominator in determining μ_{init} (Equation 17) supported stronger selective inputs when the network had stronger recurrent processing (smaller γ_g), allowing the rapid-onset, high-rate visual response described above. For larger γ_g, the network more readily gives way to its inputs, so a weaker input is sufficient to elicit a similar response. The visual response delay was $t_{vrd} = 50$ms (Thomas & Paré, 2007).

2.6.1 Parameter values for the PPC-PFC model

As noted above, the PPC network was identical in the PPC-only and PPC-PFC models. The PFC network was identical to the PPC network, except that GABAR conductance onto pyramidal neurons was twice as strong ($G_{GABA}^p = 3$nS), facilitating strong competition during stimulus encoding. We limited the number of parameters in the PPC-PFC model by making three simplifying assumptions. Firstly, we assumed that only pyramidal neurons make inter-aerial projections [see Jones (1984); White (1989); Tomioka and Rockland (2007)] and we therefore omitted inter-aerial projections from

15

inhibitory interneurons. Secondly, based on evidence that feedforward and feedback excitation are predominantly mediated by AMPARs and NMDARs respectively (Self et al., 2012), we omitted feedforward NMDARs and feedback AMPARs altogether. Thirdly, feedforward and feedback projections were each assumed to be topographically organized, based on evidence that the proportion of feedforward supragranular projections and feedback infragranular projections increase with hierarchical distance between cortical areas. Since supragranular and infragranular projections are believed to be topographic and diffuse respectively, and since PPC and lateral prefrontal cortex are separated by a single hierarchical layer [see Goldman-Rakic (1988)], we assumed that the proportion of topographic supragranular projections and diffuse infragranular projections between these areas is approximately equal in both directions. See Markov and Kennedy (2013) for a more thorough description of these macrocircuit principles. Given the diffuse activity already simulated by background current I^{back} (Equations 9 and 10), we ignored the effect of infragranular feedforward and feedback projections between the PPC and PFC networks, and simulated topographic supragranular connectivity only. The width of feedforward RFs (PFC) was the same as for stimulus-selective RFs in the PPC network ($\sigma_{ff} = \sigma_{rf}$, Equations 15 and 18), based on evidence that differences in dendritic branching are minimal between hierarchically adjacent cortical areas (Elston, 2002). Feedback RFs (PPC) were wider than feedforward RFs ($\sigma_{fb} = 1.5 \cdot \sigma_{ff}$, Equations 16 and 15) because dendritic branching in supragranular layers is more extensive than in (feedforward input) layer 4. For simplicity, we omitted FF projections onto PFC interneurons, having omitted stimulus-selective inputs to PPC interneurons [a common approach with local-circuit models of this class, *e.g.* Compte et al. (2000); Wei et al. (2012)].

3 Results

3.1 Simulations with the PPC-only model

In the PPC-only model, we used control parameter γ_g^{ppc} (Equations 6, 9 and 18 in Methods) to modulate network dynamics, scaling the strength of all extrinsic synapses inversely with that of all intrinsic synapses. By applying this scale factor to excitatory and inhibitory synapses onto pyramidal neurons and interneurons, we were able to maintain excitatory/inhibitory
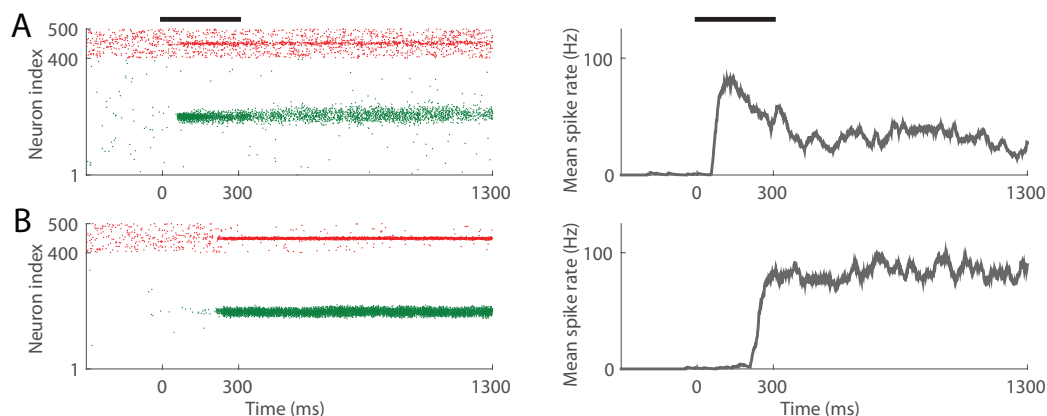
16

Figure 2: A single trial of the 1-item memory task with the PPC-only model for the lowest and highest values of control parameter $\gamma_{rec}^{ppc} = (1/\gamma_g^{ppc}$, furnishing weak (A, $\gamma_{rec}^{ppc} = 1.33$) and strong (B, $\gamma_{rec}^{ppc} = 4$) recurrent dynamics respectively. In raster plots, pyramidal neurons and interneurons are indexed from $1 - 400$ and $401 - 500$ respectively. Each right-side panel shows the mean rate of the item-encoding pyramidal population. Thick horizontal bars show the timing of the stimulus.

balance, controlling retention ability by recurrent excitation and competitive interactions by lateral inhibition. We determined the floor and ceiling on γ_g^{ppc} by stipulating that the model must perform the 1-item memory task with 90% accuracy. To this end, we ran 100 trials of the task for a range of values of γ_g^{ppc} (increments of 0.05), finding that our criterion was satisfied for $0.25 \leq \gamma_g^{ppc} \leq 0.75$. Figure 2 shows example trials of the 1-item task for the highest and lowest values of this parameter. Because lower values support stronger recurrent dynamics (intrinsic synaptic conductance values are divided by γ_g^{ppc}, Equation 6), it is intuitive to define parameter $\gamma_{rec}^{ppc} = 1/\gamma_g^{ppc}$, so that higher values support stronger recurrent dynamics. We use the latter term in the description of our results below.

We ran 100 trials for each value of γ_{rec}^{ppc} in each load condition of the multiple-item memory task ($2 \leq n \leq 8$, where n is the number of items). See Figure 3 for example trials of the 8-item task. For each load condition, we calculated the mean number of items accurately encoded during the stimulus interval, referred to as the effective load $E(n)$, and the mean number of items accurately retained over the memory delay, referred to as capacity $K(n)$. We refer to the maximum value of $K(n)$ as peak capacity $\hat{k}$ and we define WM

17

overload as $\Theta = 1 - K(8)/\max[K(n)]$, *i.e.* overload refers to a decrease in capacity on the 8-item task, relative to peak capacity.

Our results with the PPC-only model distinguished between two subsets of our control parameter. For $\gamma_{rec}^{ppc} < 2.5$, peak capacity increased with the strength of recurrent dynamics (Figure 4A), almost all stimuli were encoded (Figure 4B), and overload was catastrophic [$K(8)$ was close to 0, Figure 4C]. For $\gamma_{rec}^{ppc} > 2.5$, peak capacity, effective load and overload all decreased with stronger recurrent dynamics. Thus, $\gamma_{rec}^{ppc} = 2.5$ separated two qualitatively different regimes, one in which weaker recurrent dynamics offered no advantages (lower capacity and catastrophic overload) and one in which stronger recurrent dynamics imposed more intense competition during stimulus encoding, limiting effective load and thereby reducing overload. Consequently, we limit further consideration of the PPC-only model to $\gamma_{rec}^{ppc} \geq 2.5$. In this regime, peak capacity was 2 or 3 items for all but the strongest recurrent dynamics, consistent with that of human and monkey subjects [*e.g.* Luck and Vogel (1997); Heyselaar, Johnston, and Paré (2011)], but the alleviation of overload came at the expense of peak capacity (significant positive correlation, Figure 4D). The available data show the opposite trend: overload is typically more pronounced among lower-capacity subjects (Linke et al., 2011; Matsuyoshi et al., 2014; Fukuda et al., 2015), on lower-capacity tasks (Xu, 2007) and in lower-capacity conditions of the same task (Chee & Chuah, 2007). Furthermore, recurrent dynamics strong enough to eliminate overload altogether furnished a peak capacity of $\hat{k} \approx 1$ item, unrealistically low for healthy adult subjects [1]. Thus, the PPC-only model provides proof of concept for competitive encoding, but is qualitatively inconsistent with prominent trends in behavioural data. Furthermore, spiking activity in the model conflicts with neural data from PPC, which typically show a response to all items in a stimulus array, prior to the selection of task-relevant items [*e.g.* Thomas and Paré (2007)]. This discrepancy is implicit in Figure 4B and explicit in Figure 3A-C, where the effective load on the 8-item task is three, two and one items respectively. We therefore turned to the PPC-PFC

---

[1] The effective load for the strongest recurrent dynamics in the PPC-only model (highest value of γ_{rec}^{ppc}, darkest red curves in Figure 4) is less than capacity in some load conditions, *i.e.* $E(n) < K(n)$ (Figure 4A and B). This seeming anomaly reflects the later onset of stimulus-selective activity with such strong recurrent dynamics, where the rate of item-encoding activity (calculated over the full stimulus interval) is too low to satisfy our criteria for encoded items. This effect can be seen in Figure 2, where the onset of item-encoding activity occurs later with higher γ_{rec}^{ppc}.
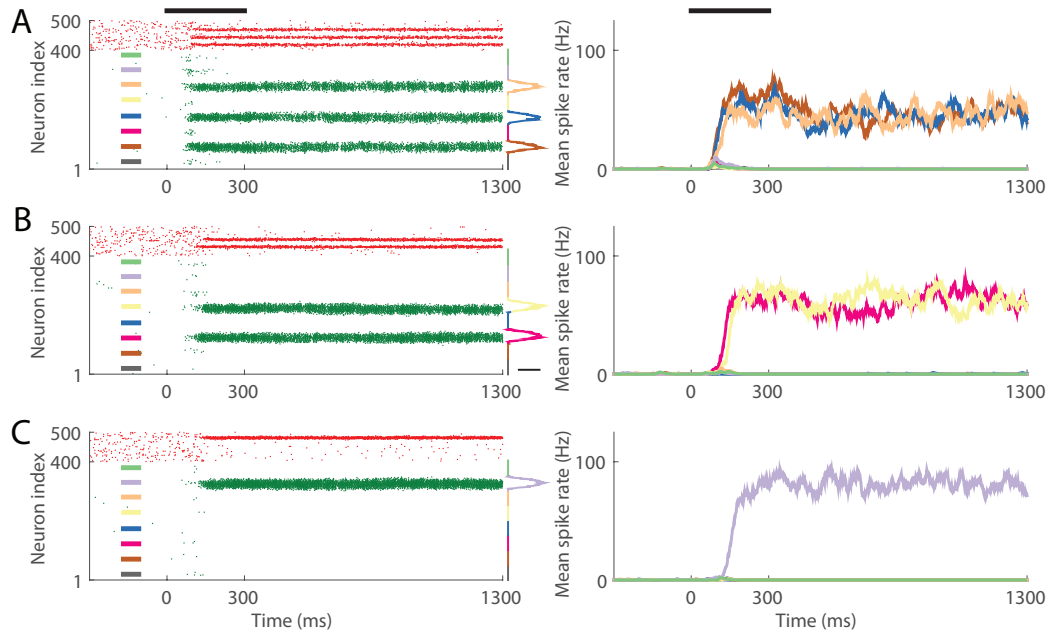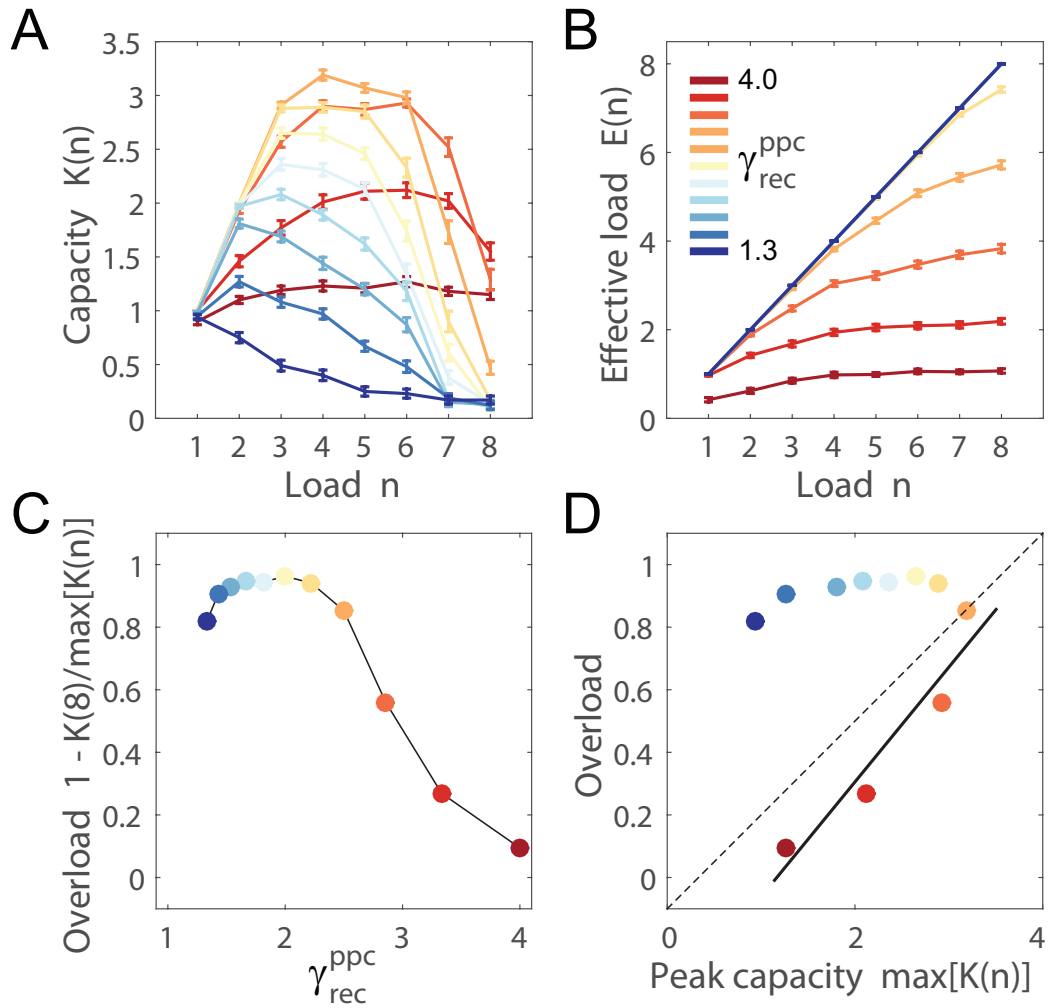
Figure 3: A single trial of the 8-item memory task with the PPC-only model for each of three values of control parameter γ_g^{ppc}, on which the network accurately retained three (A, $\gamma_g^{ppc} = 0.35$), two (B, $\gamma_g^{ppc} = 0.3$) and one (C, $\gamma_g^{ppc} = 0.25$) item(s). Raster plots (left) and mean spike rates (right) are described in Figure 2. Here, colours show the correspondence between item-encoding populations in the left and right panels. Competition during the stimulus interval reduces the load to 3 (A), 2 (B) and 1 (C) item(s) by the onset of the memory delay.

model, investigating its ability to account for these and other data.

3.2   Hierarchical simulations with the PPC-PFC model

We used three control parameters with the PPC-PFC model. Once again, we controlled recurrent dynamics in the PPC network (henceforth PPC) with parameter γ_{rec}^{ppc}. We controlled recurrent dynamics in the PFC network (henceforth PFC) in the same way with parameter $\gamma_{rec}^{pfc} = 1/\gamma_g^{pfc}$ (Equations 6, 9 and 18) and we controlled the strength of feedback projections to PPC from PFC with parameter γ_g^{fb} (Equation 16). We began by assigning values to these parameters that explicitly capture the computational princi-

ples identified above: weak recurrent dynamics in PPC, allowing the initial encoding of all items, and strong recurrent dynamics in PFC, supporting strong competition during stimulus encoding. Under this approach, persistent mnemonic activity was supported by strong feedback connectivity to PPC from PFC. An example of the 8-item memory task is shown in Figure 5, where our chosen parameter values are provided in the figure caption. The model had a peak capacity of $\hat{k} \approx 3$ items, overcame overload almost entirely ($\Theta \approx 0.075$) and allowed the initial encoding of all items in PPC (Figure 6). To emphasize the effectiveness of the proposed mechanism, we compared these results to the best possible performance by the PPC-only model, selecting the highest capacity for each load condition across all values of γ_{rec}^{ppc}. As shown in Figure 6A, the PPC-only model cannot reduce overload without a cost to peak capacity, even if we assume perfect load-dependent modulation of network dynamics by our control parameter from trial to trial.

3.2.1   Simulated cognitive control with the PPC-PFC model

Having determined that hierarchical recruitment of competition alleviates overload in the PPC-PFC model, we sought to determine whether less over-

---

Figure 4 *(preceding page)*: Competitive encoding alleviates overload in the PPC-only model, but at a cost to peak capacity (see text). (A) Capacity $K(n)$ in each load condition (number of items n) for each value of control parameter $\gamma_{rec}^{ppc} = 1/\gamma_g^{ppc}$. Error bars show standard error. Colour coding interpolates between weakest (dark blue) and strongest (dark red) recurrent dynamics (see legend in panel B). Peak capacity is highest for $\gamma_g^{ppc} = 2.5$ and $K(8)$ is close to 0 for γ_{rec}^{ppc} below this value $[K(8) < 0.15]$. Thus, there is no advantage to $\gamma_{rec}^{ppc} < 2.5$. (B) The mean number of items encoded during the stimulus interval for each value of γ_{rec}^{ppc}, referred to as the effective load $E(n)$. For $\gamma_{rec}^{ppc} < 2.5$, $E(n) > 0.9 \cdot n$ in all load conditions. For γ_{rec}^{ppc} above this value, encoding is increasingly competitive (bottom four curves). (C) WM overload for each value of γ_{rec}^{ppc}, calculated as the relative change from peak capacity to $K(8)$ (see text). For $\gamma_{rec}^{ppc} > 2.5$, overload is greatly reduced. (D) Overload over peak capacity. For $\gamma_{rec}^{ppc} \leq 2.5$, the linear fit (solid line) shows a significant positive correlation between these measures ($r > 0.95$, $P < 0.05$). The dashed line depicts two qualitatively different regimes in the model, separated by $\gamma_{rec}^{ppc} = 2.5$.
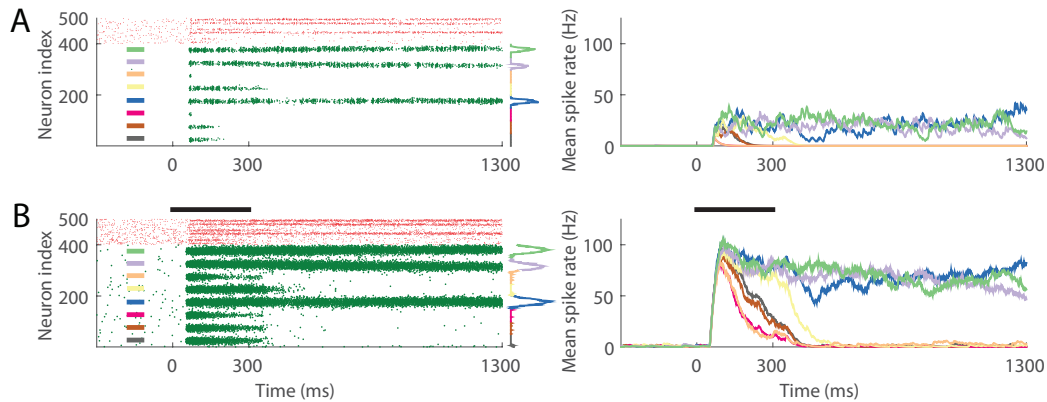
21

Figure 5: A single trial of the 8-item memory task with the PPC-PFC model, on which three items were accurately retained. Raster plots (right), mean spike rates (left) and corresponding colour schemes are described in Figures 2 and 3. Here, spiking activity is shown in the PPC network (PPC, bottom) and the PFC network (PFC, top). Weak recurrent dynamics in PPC ($\gamma_{rec}^{ppc} = 0.67$) support the initial encoding of all items in the network, but the effective load is reduced by competitive encoding in PFC ($\gamma_{rec}^{pfc} = 4$). Persistent activity is supported by inter-aerial projections ($\gamma_{g}^{fb} = 3$).
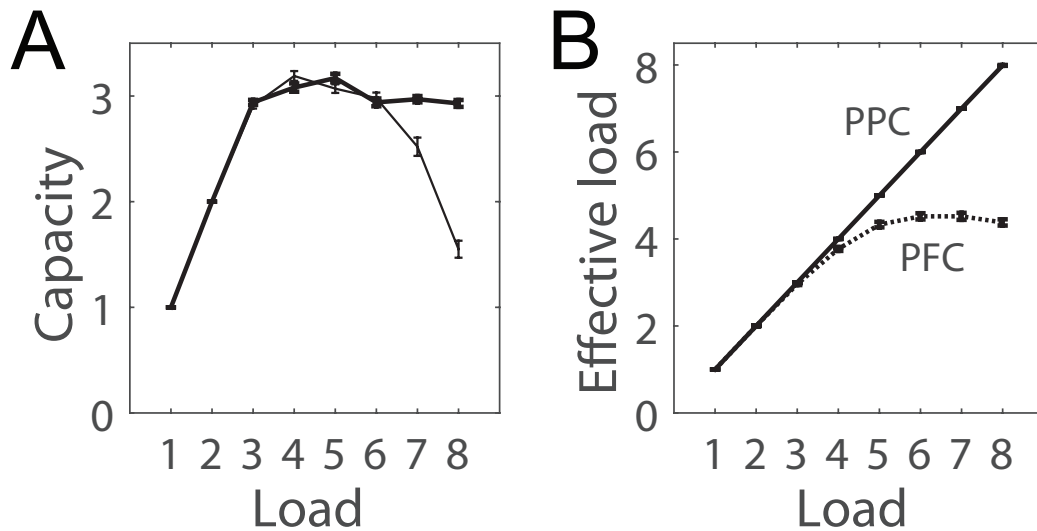
22

Figure 6: Competitive encoding in PFC alleviates overload in the PPC-PFC model, where weak recurrent dynamics in PPC permit the initial encoding of all items. (A) Capacity $K(n)$ for the PPC-PFC model (thicker curve) and the 'best of' $K(n)$ in the PPC-only model [thin curve; highest $K(n)$ over all γ_{rec}^{ppc}, see text]. (B) Effective encoding $E(n)$ in the PPC-PFC model, where solid and dashed curves correspond to PPC and PFC respectively. Error bars show standard error. Control parameters are $\gamma_{rec}^{ppc} = 0.67$, $\gamma_{rec}^{pfc} = 4$ and $\gamma_g^{fb} = 5$.

23

load would be seen under parameters supporting higher peak capacity, per the available data (Chee & Chuah, 2007; Cusack et al., 2009; Matsuyoshi et al., 2014; Fukuda et al., 2015). To this end, we ran a block of trials (100 trials for memory loads $1:8$) for a range of values of each control parameter, calculating peak capacity $\hat{k}$ and overload Θ for each combination of values (each configuration). We then determined the correlation between peak capacity and overload $r_{\hat{k},\Theta}$ with monotonic changes to our parameter values, simulating cognitive control by modulation of PPC dynamics (varying γ_{rec}^{ppc}), PFC dynamics (varying γ_{rec}^{pfc}) and the functional connectivity between PPC and PFC (varying γ_{g}^{fb}). Overall, we ran blocks for 500 configurations, spanning 10 values of γ_{rec}^{ppc} ($0.67:0.1:1.67$), five values of γ_{rec}^{pfc} ($2.5:0.05:5$) and ten values of γ_{g}^{fb} ($1:1:10$). This range of values was sufficiently broad and fine-grained for monotonic trends in $r_{\hat{k},\Theta}$ to saturate and for non-monotonic trends to clearly reverse direction (see below).

We first considered our control parameters in isolation from one another, calculating $r_{\hat{k},\Theta}$ over all values of a given parameter for every combination of the remaining two parameters. PPC modulation yielded a strong negative correlation over a broad region of the γ_{rec}^{pfc} x γ_{g}^{fb} parameter space ($r < -0.5$ for 14 out of 50 combinations of γ_{rec}^{pfc} and γ_{g}^{fb}), where moderate to strong recurrent dynamics in PFC coincided with moderate to strong feedback projections to PPC (Figure 7A). Thus, a negative correlation was seen when competitive encoding was strong enough to select a manageable number of items and feedback projections were strong enough to sustain their neural representations over the delay. Notably, peak capacity increased and overload decreased as recurrent dynamics in PPC were *weakened* (lower γ_{rec}^{ppc}, Figure 8A-C), suggesting that if PFC selects memoranda and inter-aerial projections sustain them, then the best thing for PPC to do is give way to its inputs, thereby limiting competition during the delay.

PFC modulation yielded positive correlations over the full γ_{rec}^{ppc} x γ_{g}^{fb} parameter space (Figure 7B). This result was predictable, since limiting the number of items available for storage limits peak capacity. In other words, γ_{rec}^{pfc} embodies a trade-off between peak capacity and overload, as seen in the PPC-only model (Figure 4). Stronger competitive encoding therefore had a monotonic effect on correlations (*e.g.* Figure 8F).

Modulation of the connectivity between PPC and PFC (varying γ_{g}^{fb}, inter-aerial modulation) yielded a strong negative correlation ($r < -0.5$ for 6 out of 50 combinations of γ_{rec}^{ppc} and γ_{g}^{fb}) in the corner of the γ_{rec}^{ppc} x γ_{rec}^{pfc} parameter space where stronger recurrent dynamics in PPC coincide with weaker
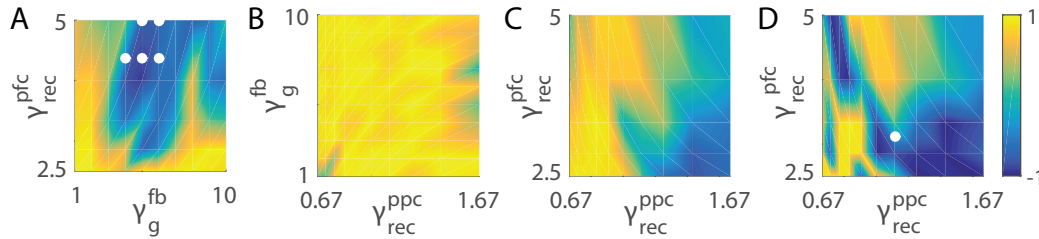
24

Figure 7: Correlation between peak capacity and overload in the PPC-PFC model under PPC modulation (varying γ_{rec}^{ppc}, A), PFC modulation (varying γ_{rec}^{pfc}, B) and modulation of PPC-PFC connectivity (inter-aerial modulation, varying γ_g^{fb}, C-D). Each heat map shows the Pearson correlation coefficient $r_{\hat{k},\Theta}$ for every combination of the other two parameters. Correlations were calculated over the full range of γ_g^{fb} (C) and the largest continuous subset of values showing a reduction in overload (D, see text). White dots in panels A and D show configurations with a significant negative correlation ($p < 0.05$), where peak capacity and overload are consistent with behavioral data from WM tasks, and effective encoding in PPC is consistent with neural data from visual tasks (see text, Section 3.2.1).

recurrent dynamics PFC (Figure 7C). Thus, a negative correlation was seen when PFC was insufficiently selective and PPC was too competitive to be overcome by weak (inter-aerial) feedback projections. Stronger feedback projections therefore increased peak capacity and decreased overload by providing more support to the subset of items selected by PFC. Notably, increasing the strength of PPC-PFC connectivity rarely had a monotonic effect on correlations, so we calculated the correlation between peak capacity and overload for a continuous subset of γ_g^{fb}, where this subset ranged from $\gamma_g^{fb} = 1$ to the value of γ_g^{fb} producing the least overload for a given combination of γ_{rec}^{ppc} and γ_{rec}^{pfc} (requiring at least three values, Figure 8I). A strong negative correlation ($r < -0.5$) occurred for 18 out of 50 combinations of γ_{rec}^{ppc} and γ_{rec}^{pfc}, most of which resulted from a broadening of the parameter region identified above (strong and weak recurrent dynamics in PPC and PFC respectively; compare panels C and D in Figure 7). The rest were spurious, where the magnitude of overload to be overcome was negligible (*e.g.* overload was always below $\Theta = 0.1$ for the negative correlation in Figure 8J-L) or where overload was considerably reduced, but remained high (*e.g.* overload never crosses $\Theta = 0.1$ in Figure 8M-O).

25
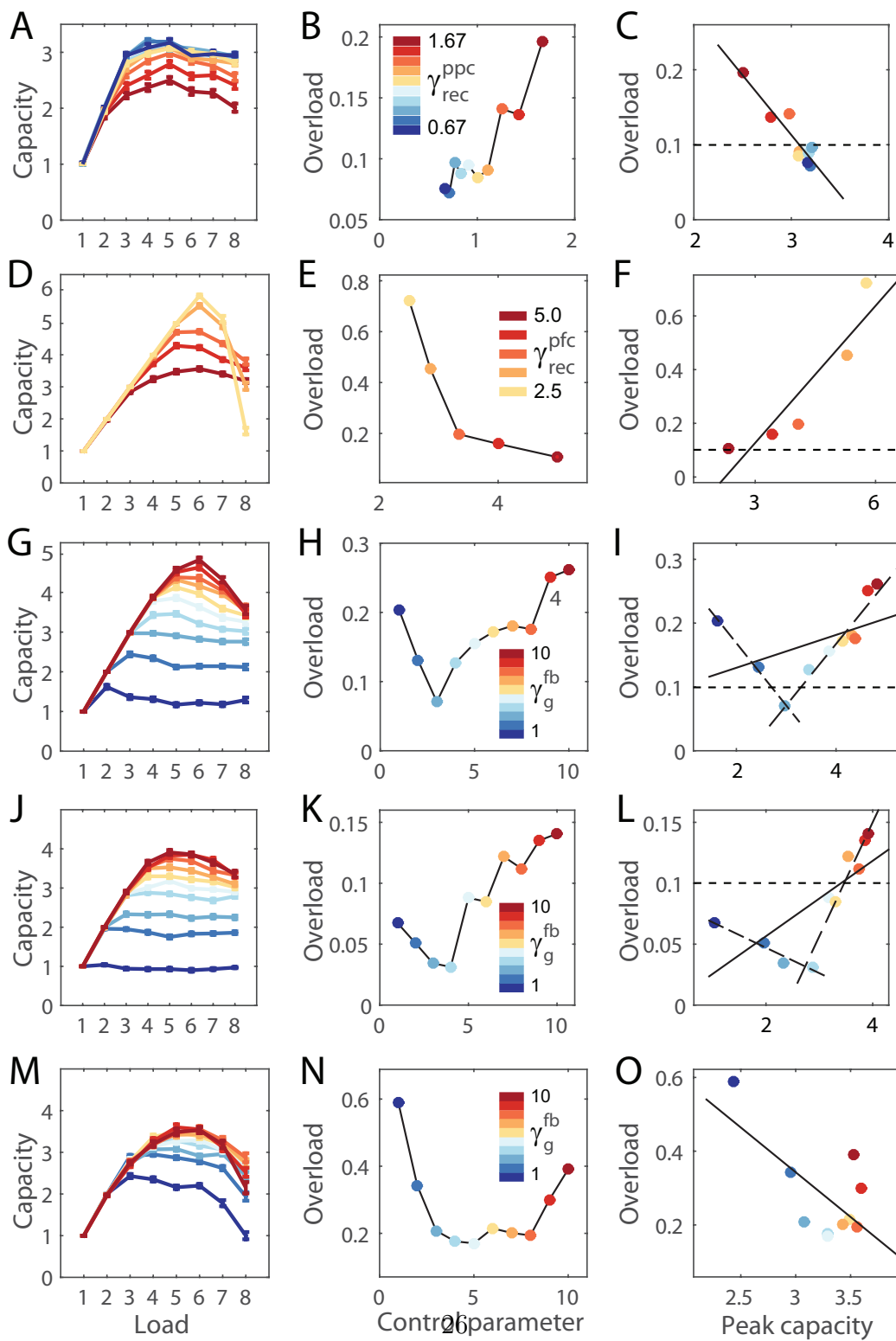
Next, we sought to determine whether peak capacity and the reduction

Figure 8 *(preceding page)*: Examples of PPC modulation (first row), PFC modulation (second row) and inter-aerial modulation (third to fifth rows) under a single configuration of the remaining two parameters. First, second and third columns show capacity as a function of load, overload as a function of control parameter (given by legend, inset) and overload as a function of peak capacity respectively. (A-C) PPC modulation with $\gamma_{rec}^{pfc} = 4$ and $\gamma_g^{fb} = 5$. Weaker PPC dynamics increase peak capacity (A) and reduce overload (B), producing a negative correlation between these measures ($r < -0.93$, $p = 0.0001$, C). The line in panel C shows the best linear fit (least squares). (D-F) PFC modulation with $\gamma_{rec}^{ppc} = 0.67$ and $\gamma_g^{fb} = 10$. Stronger PFC dynamics reduce overload (E) at the expense of peak capacity (D), producing a positive correlation between these measures ($r > 0.91$, $p < 0.03$, F), as in the PPC-only model. (G-H) Inter-aerial modulation with $\gamma_{rec}^{ppc} = 1.00$ and $\gamma_{rec}^{pfc} = 2.86$. Stronger feedback projections to PPC from PFC increase capacity (G) and decrease overload (H) over a continuous subset of γ_g^{fb} ($1 \leq \gamma_g^{fb} \leq 3$), producing a negative correlation ($r < -0.99$, $p < 0.05$, short-dashed line in I). Stronger feedback projections increase capacity and overload, producing a positive correlation over another continuous subset ($r > 0.96$, $p = 0.0001$, long-dashed line in I). The solid line in panel I shows the linear fit to the full range of γ_g^{fb}. The horizontal dotted line shows $\Theta = 0.1$, chosen as the threshold for alleviating overload (see text). (J-L) Inter-aerial modulation with $\gamma_{rec}^{ppc} = 0.83$ and $\gamma_{rec}^{pfc} = 4$. Stronger feedback projections increase capacity (J) and decrease overload (K) over a continuous subset of γ_g^{fb} ($1 \leq \gamma_g^{fb} \leq 4$), producing a negative correlation ($r < -0.96$, $p < 0.04$, short-dashed line in L) but the magnitude of overload to be overcome is negligible ($\Theta < 0.07$ for all four values fit by the short-dashed line in L). Stronger feedback projections increase capacity and overload over another continuous subset, producing a positive correlation ($r > 0.94$, $p < 0.01$, long-dashed line in L). (M-O) Inter-aerial modulation with $\gamma_{rec}^{ppc} = 1.67$ and $\gamma_{rec}^{pfc} = 2.5$. Stronger feedback projections increase capacity (M) and decrease overload (N) over a continuous subset of γ_g^{fb} ($1 \leq \gamma_g^{fb} \leq 4$), before overload reverses direction with further increases in γ_g^{fb}. Overload is greatly reduced in magnitude, but never reaches a value consistent with high-capacity subjects and conditions ($\Theta > 0.17$ for all γ_g^{fb}).

27

in overload were quantitatively consistent with behavioural data in the parameter regions showing a significant negative correlation ($p < 0.05$), and whether PPC was consistent with neural data. In this regard, peak capacity should exceed around three items [see Cowan (2001); Luck and Vogel (2013)], overload should be reduced to close to zero (Chee & Chuah, 2007; Linke et al., 2011; Fukuda et al., 2015) and PPC should initially encode all stimuli (Thomas & Paré, 2007). For the last of these requirements, we defined E' as the minimum ratio of the effective load to the actual load over all load conditions, that is, $E' = min[E(n)/n]$ (in practice, $E' = E(8)/8$ under all configurations). Allowing a tolerance of 10%, we therefore searched for parameter regions in which $\hat{k} > 2.7$, $min(\Theta) < 0.1$ and $E' > 0.9$. We included the additional constraint that the magnitude of overload to be overcome $[max(\Theta)]$ must be greater than 0.1. Under PPC modulation, these criteria were satisfied by five contiguous locations in the parameter region showing a negative correlation (white dots in Figure 7A). Under modulation of PPC-PFC connectivity, they were not satisfied in any parameter region, but they were satisfied in one location of the $\gamma_{rec}^{ppc}x\gamma_{rec}^{pfc}$ parameter space when we searched the continuous subsets of γ_g^{fb} described above (white dot in Figure 7D). These results were unchanged when we lowered our criterion for peak capacity to 2, again with a tolerance of 10% ($\hat{k} > 1.8$). Thus, in isolation, PPC modulation and inter-aerial modulation were both able to increase capacity and reduce overload in a manner consistent with neural and behavioural data from multiple-item visual and WM tasks, where weaker PPC dynamics and stronger feedback projections supported better task performance. PPC modulation was the more robust mechanism, as functional connectivity was only able to account for these data under a single combination of the other two parameters.

3.2.2 Hierarchical recruitment of competition during stimulus encoding: computational principles for distributed WM storage

Finally, we considered our control parameters in combination with one another, aiming to provide a better quantitative account of the magnitude of overload shown by low-capacity subjects, and in low-capacity tasks and conditions (henceforth low-capacity performance). Under PPC modulation and inter-aerial modulation, the most extreme cases of overload in the parameter regions satisfying our criteria were $\Theta \approx 0.2$ (Figure 8A-C and G-I). While

28

this degree of overload is consistent with some data [*e.g.* Cusack et al. (2009); Linke et al. (2011)], overload has approached 50% of peak capacity in several studies [*e.g.* Chee and Chuah (2007); Xu (2007); Fukuda et al. (2015)] and has even approached 100% among elderly, low-capacity subjects in extreme load conditions (Matsuyoshi et al., 2014). We therefore investigated whether simultaneous variation of our control parameters could alleviate the more extreme cases of overload that occur with low-capacity performance, resulting in peak capacity consistent with high-capacity performance in the same studies. High-capacity performance is explained by the computational principles identified above: strong recurrent dynamics in PFC, supporting competitive encoding; strong feedback projections to PPC from PFC, supporting persistent activity; and weak recurrent dynamics in PPC, limiting competition during the memory delay. These principles not only capture a tangible strategy for WM storage on multiple-item tasks (selection of a subset of items for storage), but offer a specific set of mechanisms for their implementation in fronto-parietal circuitry. Thus, we reasoned that low-capacity performance would be explained by *non*-compliance with these principles, investigating the PPC-PFC model under the lowest value of γ_{rec}^{pfc}, the lowest value of γ_{g}^{fb} and the highest value of γ_{rec}^{ppc} (the low-capacity configuration). In other words, the low-capacity configuration had the weakest recurrent dynamics in PFC, the weakest feedback projections to PPC and the strongest recurrent dynamics in PPC.

The low-capacity configuration had a peak capacity of $\hat{k} \approx 2$ items, overload of $\Theta > 0.5$ and allowed all items to be encoded by PPC (Figure 9A-C, black). For a range of parameter values, increasing the strength of competition in PFC (increasing γ_{rec}^{pfc}), increasing the strength of feedback projections to PPC (increasing γ_{g}^{fb}) and decreasing the strength of competition in PPC (decreasing γ_{rec}^{ppc}) produced very similar results to those shown in Figure 6 (high-capacity configurations), raising peak capacity to $\hat{k} \sim 3$ items, reducing overload to close to $\Theta = 0$ and retaining the encoding of all items in PPC. These results are strikingly similar to peak capacity and overload among low- and high-capacity subjects [*cf.* Fukuda et al. (2015)] and task conditions [*cf.* Chee and Chuah (2007)]. We do not propose a specific trajectory through the parameter space from one extreme to the other, but rather, we emphasize that multiple trajectories show robustness of the computational principles identified by our simulations.

Surprisingly, under the high-capacity configurations satisfying the criteria described in the previous section ($\hat{k} > 2.7$, $min(\Theta) < 0.1$ and $E' > 0.9$),

modulation of all three parameters (henceforth 3-parameter modulation) produced an effective load in PFC that was greater than or equal to the effective load under the low-parameter configuration (Figure 9B), that is, the low-capacity configuration did actually implement selective encoding in PFC. This result supports the hypothesis that poor performance by low-capacity subjects does not reflect poor task strategy, *e.g.* a failure to select a subset of items for storage, but rather, reflects poor control over fronto-parietal circuitry. We therefore determined whether the simultaneous modulation of any two parameters (2-parameter modulation, holding the third parameter fixed) could satisfy the above criteria, overcome a degree of overload comparable to that of the low-capacity configuration ($\Theta \approx 0.5$), and show a lower effective load in PFC with high capacity than with the low capacity. If so, it would support the hypothesis that low-capacity subjects are encoding too many memoranda and would allow us to consider predictions under 3-parameter and 2-parameter modulation that might distinguish between these two hypotheses (poor cognitive control *vs.* poor task strategy). To this end, we simultaneously decreased and increased γ_{rec}^{ppc} and γ_{rec}^{pfc} from their highest and lowest values respectively, while holding γ_g^{fb} fixed (at all possible values); simultaneously decreased and increased γ_{rec}^{ppc} and γ_g^{fb} from their highest and lowest values respectively, while holding γ_{rec}^{pfc} fixed; and simultaneously increased γ_{rec}^{pfc} and γ_g^{fb} from their lowest values, while holding γ_{rec}^{ppc} fixed. The first two approaches were unable to satisfy our constraints, but the third approach was able to do so with a moderate increase in the strength of competition in PFC, a small increase in the strength of feedback projections, and with PPC dynamics fixed at a moderate level (maximum overload was $\Theta \approx 0.4$).

3.2.3 Predictions for experimental testing

As detailed above, the PPC-PFC model implements hierarchical recruitment of competition during stimulus encoding, accounting for WM performance when memory load exceeds subjects' retention abilities [compare Figure 9A-C with Figure 4B in Fukuda et al. (2015)]. To make predictions to test this hypothesis experimentally, we approximated high-density EEG recordings over PPC and lateral PFC, simulating low- and high-capacity performance with the low- and high-capacity parameter configurations identified above. To approximate the EEG signal, we followed the approach by McCarthy, Brown, and Kopell (2008), summing all excitatory currents
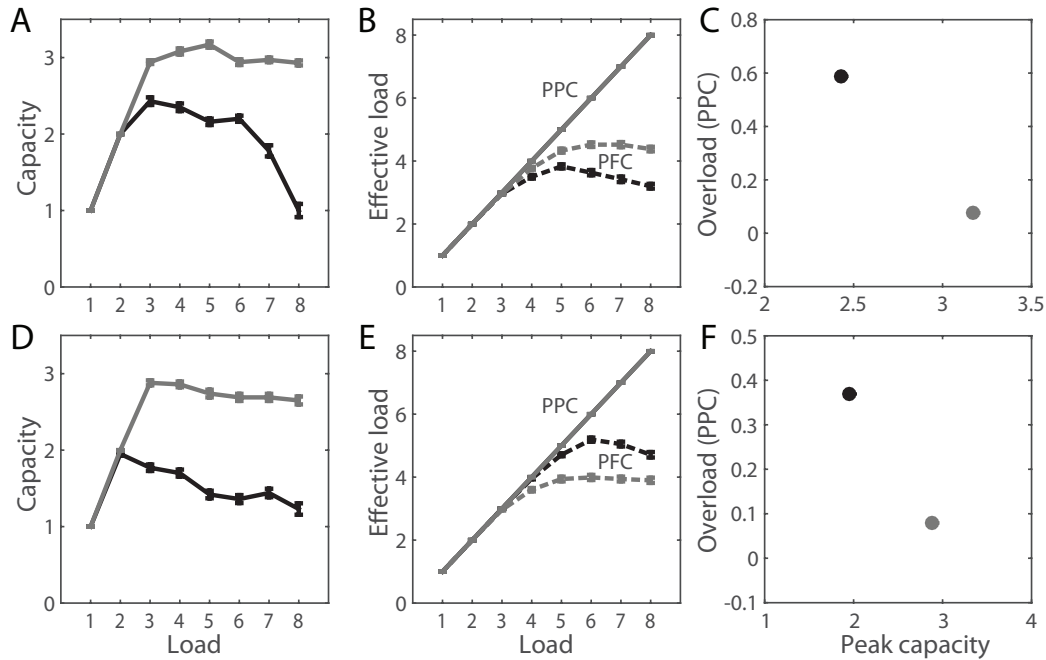
Figure 9: (A-C) Simultaneously increasing the strength of competition in PFC, increasing the strength of feedback projections to PPC from PFC, and decreasing the strength of competition in PPC (3-parameter modulation) raised peak capacity (A) and reduced overload (C) in a manner consistent with neural and behavioural data (see text, Section 3.2.1). Effective load in PFC (lower curves in panel B) was higher under the high-capacity configuration (grey, parameter values provided in Figure 6) than the low-capacity configuration (black; $\gamma_{rec}^{ppc} = 1.67$, $\gamma_{rec}^{pfc} = 2.5$, $\gamma_{g}^{fb} = 1$. 3-parameter modulation was therefore inconsistent with the hypothesis that low capacity and pronounced overload result from the encoding of too many items in PFC. (D-F) Simultaneously increasing the strength of competition in PFC, increasing the strength of feedback projections to PPC, and fixing PPC dynamics at a moderate level (2-parameter modulation) similarly raised peak capacity and reduced overload, but the effective load in PFC was higher under the low-capacity configuration than under the high-capacity configuration (E). 2-parameter modulation was therefore consistent with the hypothesis that low capacity and pronounced overload result from the encoding of too many items in PFC.

31

onto pyramidal neurons in each network. Thus, we simulated the instantaneous source amplitude of EEG over PPC by $EEG^{ppc} = \sum_j I_j^{back,syn} + \sum_j I_j^{sel} + \sum_j I_{AMPA,j}^{rec} + \sum_j I_{NMDA,j}^{rec} + \sum I_j^{fb}$ and over lateral PFC by $EEG^{pfc} = \sum_j I_j^{back,syn} + \sum_j I_{AMPA,j}^{rec} + \sum_j I_{NMDA,j}^{rec} + \sum I_j^{ff}$, where index j refers to pyramidal neurons, currents $I_{AMPA,j}^{rec}$ and $I_{NMDA,j}^{rec}$ are given by Equation 6, and currents $I_j^{back,syn}$, I_j^{sel}, I_j^{fb} and I_j^{ff} are given by Equations 9, 18, 16 and 15 respectively. We summed this instantaneous signal over the 300ms stimulus interval and the last 300ms of the memory delay to obtain the total amplitude $EEG_s^{\{ppc,pfc\}}$ in each task epoch. Because competitive dynamics in PPC and PFC may be modulated by different mechanisms than the scaling of synaptic conductances used here (γ_{rec}^{ppc} and γ_{rec}^{pfc}) and because our EEG approximation sums the resulting synaptic currents, we normalized the EEG approximation by its minimum and maximum values ($EEG_\eta = [EEG_s - min(EEG_s)]/[max(EEG_s) - min(EEG_s)]$), predicting the qualitative form of EEG amplitude as a function of memory load, rather than absolute amplitude.

During the stimulus interval, EEG_η showed greater concavity (concave down) as a function of memory load for PFC than PPC (Figure 10A and C) under the low- and high-capacity configurations alike. In this regard, concavity serves an index of the timing of selective encoding, $i.e.$ greater concavity reveals earlier selection of memoranda. Notably, this finding was the case for both 3-parameter and 2-parameter modulation, providing a robust signature of hierarchical recruitment of competition. Thus the PPC-PFC model makes a specific, testable prediction for our hypothesis: EEG amplitude will show greater concavity over memory load when recorded over lateral PFC than when recorded over PPC during the stimulus interval of multiple-item WM tasks.

During the delay interval, EEG_η was qualitatively indistinguishable for PPC and PFC under the low- and high-capacity configurations, as expected (together, the two networks support persistent activity via inter-aerial projections). EEG_η was similarly indistinguishable under the high-capacity configurations under 3-parameter and 2-parameter modulation (grey curves, Figure 10B and D), where simulated EEG amplitude was bi-linear over memory load. This bi-linearity is strikingly similar to the bi-linear EEG amplitude over PPC shown by Vogel and Machizawa (2004) for all subjects and by Fukuda et al. (2015) for low- and high-capacity subject groups.

Finally, EEG_η was qualitatively distinct during the delay period under 3-

32

parameter and 2-parameter modulation, allowing us to address the question posed by our results in the previous section: is low-capacity WM performance attributable to poor cognitive control or poor task strategy? Under 3-parameter modulation, EEG_η was bi-linear over memory load for both networks under low- and high-capacity configurations, but the 'second line' of the bi-linear curve had a negative slope under the low-capacity configuration and was approximately flat under the high-capacity configuration (Figure 10B). These curves are strikingly similar to the bilinear EEG amplitude shown by Fukuda et al. (2015) over PPC for low- and high-capacity subjects. Under 2-parameter modulation, EEG_η was tri-linear for both networks (Figure 10D) and therefore did not account for the available EEG data. As such, these findings support the hypothesis that low-capacity subjects are indeed selectively encoding memoranda for storage, but they have poor control over their fronto-parietal circuitry.

4   Discussion

The storage limitations of WM have been the subject of intense research interest for several decades [see Luck and Vogel (2013)], but although several studies have reported a reduction in WM capacity with high memory load [*e.g.* Xu (2007); Chee and Chuah (2007)], WM overload has only been the focus of a handful of behavioural experiments (Cusack et al., 2009; Linke et al., 2011; Matsuyoshi et al., 2014; Fukuda et al., 2015). We investigated the neural basis of overload with the PPC and PPC-PFC models, finding that overload could be reduced in both models by strong competitive dynamics during the stimulus interval of simulated WM tasks. The PPC-only model, however, showed a positive correlation between peak capacity and overload (Figure 4), in opposition to available data (Chee & Chuah, 2007; Xu, 2007; Linke et al., 2011; Matsuyoshi et al., 2014; Fukuda et al., 2015). The PPC-PFC model accounted for these data in a parameter regime where selective encoding was supported by strong competitive dynamics in PFC, persistent activity was supported by inter-aerial projections, and weak dynamics in PPC limited competition during the memory delay (Figure 6). As such, the model implemented hierarchical recruitment of competition during stimulus encoding and identified a set of computational principles for WM storage in distributed circuitry. Under these principles, all WM items were encoded by PPC (Figure 6B), consistent with single-cell electrophysiological recordings

33
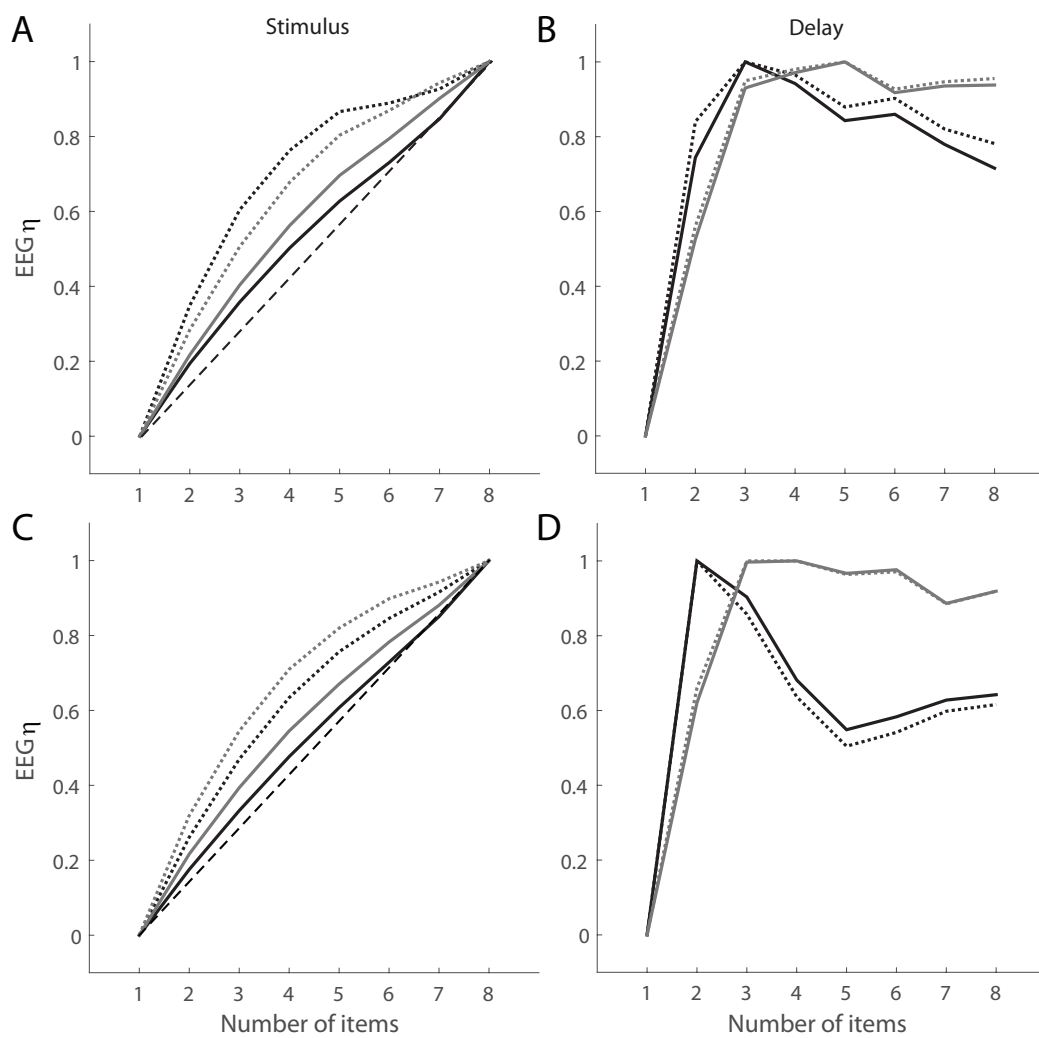
from PPC (Thomas & Paré, 2007); simulated EEG amplitude was bi-linear over memory load during the delay period (Figure 10B), consistent with EEG recordings over PPC (Vogel & Machizawa, 2004; Fukuda et al., 2015); and peak capacity was around three items, consistent with behavioural data from numerous WM tasks (Figure 6B) [see Cowan (2001); Luck and Vogel (2013)]. When we violated the identified principles (increased competition in PPC, decreased strength of feedback projections and decreased competition in PFC), peak capacity was reduced to just over two items (Figure 9A), overload was greater than 50% of peak capacity (Figure 9C) and the 'second line' of the bi-linearity of simulated EEG amplitude showed a negative slope (Figure 10B). These results are strikingly consistent with behavioural and EEG data from low-capacity subjects in the study by Fukuda et al. (2015). To our surprise, the model implemented selective encoding in this low-capacity regime (Figure 9B). Thus, it captured a strategy for WM storage under high load and offered a set of neural mechanisms for its implementation in hierarchical circuitry, but it predicted that low-capacity subjects are indeed attempting this strategy and that their performance reflects poor control of fronto-parietal processing. Our hypothesis is testable by the prediction that EEG amplitude over memory load will show greater concavity over lateral

Figure 10 (preceding page): Predictions by the PPC-PFC model for normalized EEG amplitude (EEG_η, see text) over PPC (solid) and lateral PFC (dotted) for low-capacity (black) and high-capacity (grey) WM performance. (A-B) EEG_η as a function of memory load during the stimulus interval (A) and during the last 300ms of the memory delay (B) under 3-parameter modulation. During the stimulus interval (A), hierarchical recruitment of competition during stimulus encoding predicts greater concavity over PFC, indicating earlier selection of memoranda than in PPC. The dashed unity line highlights concavity of the curves. During the memory delay (B), the model predicts bilinear amplitude over PPC and lateral PFC in low- and high-capacity conditions, where the slope of the 'second line' will be negative in low-capacity conditions [cf. Vogel and Machizawa (2004); Fukuda et al (2015)]. (C-D) Under 2-parameter modulation (PPC dynamics fixed at a moderate level, see text), the PPC-PFC model again predicts greater concavity over PFC during the stimulus interval, and again predicts bi-linear EEG amplitude over PPC and PFC in high-capacity conditions, but EEG_η is tri-linear over both networks.

PFC than over PPC during the stimulus interval of WM tasks (Figure 10A), providing a neural signature of early selection.

Our models have limitations, of course. While there is growing support for the hypotheses that PPC is the hub of distributed WM circuitry (Palva et al., 2010; Christophel et al., 2012; Salazar et al., 2012) and that fronto-parietal interactions play a central role in multiple-item WM storage (Edin et al., 2009; Palva et al., 2010), other brain regions play important roles in WM [see Sreenivasan, Curtis, and DEsposito (2014); D'Esposito and Postle (2015)]. Thus, we do not claim that the PPC-PFC model and the computational principles it identifies should explain WM overload under all possible conditions. For example, the negative correlation between peak capacity and overload that guided our investigations is common (Chee & Chuah, 2007; Xu, 2007; Cusack et al., 2009; Linke et al., 2011; Matsuyoshi et al., 2014; Fukuda et al., 2015), but not ubiquitous. These two measures have been shown to increase together with the duration of stimulus encoding [see Figure 4 by Cusack et al. (2009)], as well as during childhood development [from $6-7$ year-old children to college students in the study by Cowan, Morey, AuBuchon, Zwilling, and Gilchrist (2010)].

Another limitation of our models is that they only consider the spatial location of memoranda, ignoring other features and their conjunctions. In effect, our simulations assume that everything encoded by PPC satisfies a set of rules for selection, *e.g.* red squares or blue circles. This approach is common among neural models of WM storage [*e.g.* Compte et al. (2000); Tanaka (2002); Macoveanu, Klingberg, and Tegnér (2006); Edin et al. (2009); Wei et al. (2012)] and is reasonable for studies focused on capacity. While the PPC-PFC model takes an important step toward the understanding of WM storage in distributed circuitry, an understanding of feature-bound memoranda will likely require hierarchical models with converging feature maps. See Raffone and Wolters (2001) for a binding mechanism for sequentially presented memoranda [related models are described by Lisman and Idiart (1995); Jensen and Lisman (1996)].

Finally, our focus on WM overload inherently limits our study to the investigation of capacity, but we do not suggest that capacity provides the only limitation on WM storage. There is ample evidence that the precision of memoranda is load-dependent [*e.g.* Zhang and Luck (2008); Bays, Catalao, and Husain (2009); van den Berg, Shin, Chou, George, and Ma (2012)]. Historically, capacity and precision have been presented as evidence for conflicting hypotheses on the nature of WM storage [see Luck and Vogel (2013);

Ma, Husain, and Bays (2014)], but neural modelling studies have begun to focus on their relationship and its neural basis (Wei et al., 2012; Roggeman, Klingberg, Feenstra, Compte, & Almeida, 2013; Okimura, Tanaka, Maeda, Kato, & Mimura, 2015). We are unaware of studies showing anything resembling overload in relation to precision (*e.g.* unchanging precision up to a critical load, followed by a decrease), but future experiments should investigate the dependence of precision on supra-capacity memory load.

Local-circuit attractor models (such as the PPC-only model) have been invaluable to our understanding of the neural basis of persistent activity on single-item tasks (Wang, 1999; Compte et al., 2000) and capacity limitations on multiple-item tasks with memory loads similar to (or less than) capacity (Tanaka, 2002; Macoveanu et al., 2006; Edin et al., 2009). Assuming that all items are encoded for storage, these models necessarily produce overload when the number of memoranda sufficiently exceeds capacity, due to the competition between simulated neural populations [see Edin et al. (2009) for analysis]. This finding reveals a limitation of local-circuit models of WM storage, since not all experimental tasks, conditions and subjects show overload [*e.g.* Xu (2007); Chee and Chuah (2007); Cusack et al. (2009); Fukuda et al. (2015)]. The same can be said of hierarchical models in which a top-down control signal modulates the recurrent dynamics of a downstream network (Roggeman et al., 2013; Almeida, Barbosa, & Compte, 2015), since persistent activity is supported by attractor dynamics in the network receiving the control signal (Edin et al., 2009). The PPC-PFC model builds on this work, taking a step toward an understanding of the roles played by local circuits in distributed WM storage. An important next step is to simulate inter-aerial cortical pathways in more detail, since these pathways systematically differ according to layer, hierarchical distance and (presumably) function. The structural and mechanistic differences between the PPC-PFC model and the model by Edin et al. (2009) are instructive in this regard. Our model emphasizes the role of topographic inter-aerial pathways, which run bidirectionally in supra-granular layers between hierarchically adjacent cortical areas (such as PPC and PFC), but which are increasingly dominated by feed-forward (ascending) projections with greater hierarchical distance. In the model by (Edin et al., 2009) [and Roggeman et al. (2013) and Almeida et al. (2015)], the top-down control signal is spatially non-selective (diffuse), an established form of gain modulation in local-circuit models of this class (Salinas & Abbott, 1996; Furman & Wang, 2008; Standage et al., 2013). Diffuse pathways run bidirectionally in infra-granular layers between adja-

cent cortical areas, but are increasingly dominated by feedback (descending) projections with greater hierarchical distance, *i.e.* the opposite arrangement to topographic pathways [see Markov and Kennedy (2013)]. Thus, our different approaches capture fundamentally different mechanisms for control of WM storage: bottom-up recruitment by (and of) topographic pathways and top-down control by diffuse pathways respectively. It seems likely that both mechanisms are involved in WM storage. Future work should test the predictions of our respective models, aiming to identify the roles of different cortical areas and their functional interactions in support of WM.

The real-world phenomenon of WM overload has long been of concern to educators (Sweller, 1988), who have identified the need for a stronger scientific foundation for pedagogic strategies aiming to prevent its occurrence in the classroom (Schnotz & Kurschner, 2007; de Jong, 2010). Nonetheless, despite intense research interest in the storage limitations of WM more generally [see Luck and Vogel (2013)], only a handful of studies have specifically investigated overload (Cusack et al., 2009; Linke et al., 2011; Matsuyoshi et al., 2014; Fukuda et al., 2015) and to the best of our knowledge, no previous study has investigated its mechanistic basis. Our findings point to cognitive control as the source of differential WM performance across subject groups, rather than capacity *per se*. This finding is consistent with recent experimental work emphasizing strategic ability as the source of high performance on WM tasks and on tests of cognitive ability more generally (Cusack et al., 2009; Linke et al., 2011). Given the strong correlation between capacity and scores on intelligence tests [see Unsworth et al. (2014)], we believe this message is a positive one, though our findings do not suggest that capacity can necessarily be improved by simple strategic adjustments (Section 3.2.2). Rather, they suggest that individuals with better control of distributed cortical processing are better positioned to implement effective strategies. Significant research investment will be required to identify ways to improve this control. Our predictions for experimental testing (Section 3.2.3) are a step in this direction.

References

Almeida, R., Barbosa, J., & Compte, A. (2015). Neural circuit basis of visuo-spatial working memory precision: a computational and behavioral study. *Journal of Neurophysiology*, *114*, 1806–1818.

Angulo, M. C., Rossier, J., & Audinat, E. (1999). Postsynaptic glutamate receptors and integrative properties of fast-spiking interneurons in the rat neocortex. *The Journal of Neuroscience*, *82*, 1295–1302.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(11), 1–11.

Berger, T. K., Perin, R., Silberberg, G., & Markram, H. (2009). Frequency-dependent disynaptic inhibition in the pyramidal network: a ubiquitous pathway in the developing rat neocortex. *The Journal of Physiology*, *587*(22), 5411–5425.

Berretta, N., & Jones, R. S. G. (1996). A comparison of spontaneous epscs in layer ii and layer iv-v neurons of the rat entorhinal cortex in vitro. *Journal of Neurophysiology*, *76*(2), 1089–1100.

Chance, F. S., Abbott, L. F., & Reyes, A. D. (2002). Gain modulation from background synaptic input. *Neuron*, *35*, 773–782.

Chee, M. W. L., & Chuah, Y. M. L. (2007). Functional neuroimaging and behavioral correlates of capacity decline in visual short-term memory after sleep deprivation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(22), 9487–9492.

Christophel, T. B., Hebart, M. N., & Haynes, J.-D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *The Journal of Neuroscience*, *32*(38), 12983–12989.

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*(6), 693–702.

Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, *10*, 910–923.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E., & Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science*, *13*(1), 120–133.

Curtis, C. E. (2006). Prefrontal and parietal contributions to spatial working memory. *Neuroscience*, *139*, 173–180.

Cusack, R., Lehmann, M., Veldsman, M., & Mitchell, D. J. (2009). Encoding strategy and not visual working memory capacity correlates with intelligence. *Psychonomic Bulletin and Review*, *16*(4), 641–647.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, *38*, 105–134.

Desai, N. S., Cudmore, R. H., Nelson, S. B., & Turrigiano, G. G. (2002). Critical periods for experience-dependent synaptic scaling in visual cortex. *Nature Neuroscience*, *5*(8), 783–789.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142.

Destexhe, A. (2010). Inhibitory "noise". *Frontiers in Cellular Neuroscience*, *4*(9), 1–7.

Destexhe, A., Rudolph, M., Fellous, J.-M., & Sejnowski, T. (2001). Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons. *Neuroscience*, *107*(1), 13–24.

Douglas, R., Markram, H., & Martin, K. (2004). Neocortex. In G. M. Shepherd (Ed.), *The synaptic organization of the brain* (pp. 499–558). Oxford University Press.

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegnér, J., & Compte, A. (2009). Mechanism for top-down control of working memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(16), 6802-6807.

Elston, G. N. (2002). Cortical heterogeneity: Implications for visual processing and polysensory integration. *Journal of Neurocytology*, *31*, 317–335.

Fellous, J.-M., Rudolph, M., Destexhe, A., & Sejnowski, T. (2003). Synaptic background noise controls the input/ouput characteristics of single cells in an in vitro model of in vivo activity. *Neuroscience*, *122*, 811–829.

Fukuda, K., Woodman, G. F., & Vogel, E. K. (2015). Individual differences in visual working memory capacity: Contributions of attentional control to storage. In P. Jolicoeur, C. Lefebvre, & J. Martinez-Trujillo (Eds.), *Mechanisms of sensory working memory: Attention and performance xxv* (pp. 105–119). Elsevier.

Funahashi, S. (2013). Space representation in the prefrontal cortex. *Progress in Neurobiology*, *103*, 131–155.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex.

Journal of Neurophysiology, *61*(2), 331–349.

Furman, M., & Wang, X.-J. (2008). Similarity effect and optimal control of multiple-choice decision making. *Neuron*, *60*, 1153–1168.

Goldberg, M. E., Bisley, J. W., Powell, K. D., & Gottlieb, J. (2006). Saccades, salience and attention: the role of the lateral intraparietal area in visual behavior. *Progress in Brain Research*, *155B*, 157–175.

Goldman-Rakic, P. S. (1988). Topography of cognition: parallel distributed networks in primate association cortex. *Annual Review of Neuroscience*, *11*, 137–156.

Hellwig, B. (2000). A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biological Cybernetics*, *82*, 111–121.

Hestrin, S. (1993). Different glutamate receptor channels mediate fast excitatory synaptic currents in inhibitory and excitatory cortical neurons. *Neuron*, *11*, 1083–1091.

Heyselaar, E., Johnston, K., & Paré, M. (2011). A change detection approach to study visual working memory of the macaque monkey. *Journal of Vision*, *11*(3), 1–10.

Hull, C., Isaacson, J. S., & Scanziani, M. (2009). Postsynaptic mechanisms govern the differential excitation of cortical neurons by thalamic inputs. *The Journal of Neuroscience*, *29*(28), 9127–9136.

Jahr, C. E., & Stevens, C. F. (1990). Voltage dependence of nmda-activated macroscopic conductances predicted by single-channel kinetics. *The Journal of Neuroscience*, *10*, 3178–3182.

Jensen, O., & Lisman, J. E. (1996). Theta/gamma networl(s with slow nmda channels learn sequences and encode episodic memory: Role of nmda channels in recall. *Learning and Memory*, *3*, 264–278.

J.McBain, C., & Fisahn, A. (2001). Interneurons unbound. *Nature Reviews Neuroscience*, *2*, 11–23.

Jones, E. G. (1984). Laminar distribution of cortical efferent cells. In A. Peters & E. G. Jones (Eds.), *The cerebral cortex, vol 1, cellular components of the cerebral cortex* (pp. 521–554). Plenum Press: New York.

Kisvárday, Z., Beaulieu, C., & Eysel, U. T. (1993). Network of gabaergic large basket cells in cat visual cortex (area 18): Implication for lateral disinhibition. *The Journal of Comparative Neurology*, *327*, 398–415.

Krimer, L. S., & Goldman-Rakic, P. S. (2001). Prefrontal microcircuits: Membrane properties and excitatory input of local, medium, and wide

arbor interneurons. *The Journal of Neuroscience*, *21*(11), 3788–3796.

Krimer, L. S., Zaitsev, A. V., Czanner, G., Kröner, S., González-Burgos, G., Povysheva, N. V., . . . Lewis, D. A. (2005). Cluster analysis-based physiological classification and morphological properties of inhibitory neurons in layers 2-3 of monkey dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *94*, 3009–3022.

Linke, A., Vicente-Grabovetsky, A., Mitchell, D., & Cusack, R. (2011). Encoding strategy accounts for individual differences in change detection measures of vstm. *Neuropsychologia*, *49*, 1476–1486.

Lisman, J. E., & Idiart, M. A. P. (1995). Storage of 7+-2 short-term memories in oscillatory subcycles. *Science*, *267*, 1512–1515.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400.

Ma, W., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356.

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

Markov, N. T., & Kennedy, H. (2013). The importance of being hierarchical. *Current Opinion in Neurobiology*, *23*, 187–194.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience*, *5*, 793–807.

Matsuyoshi, D., Osaka, M., & Osaka, N. (2014). Age and individual differences in visual working memory deficit induced by overload. *Frontiers in Psychology*, *5*, 1–7.

McCarthy, M. M., Brown, E. N., & Kopell, N. (2008). Potential network mechanisms mediating electroencephalographic beta rhythm changes during propofol-induced paradoxical excitation. *Journal of Neuroscience*, *50*, 13488–13504.

Merrienboër, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: recent development and future directions. *Educational Psychology Review*, *17*(2), 148–177.

Okimura, T., Tanaka, S., Maeda, T., Kato, M., & Mimura, M. (2015). Simulation of the capacity and precision of working memory in the hy-

podopaminergic state: relevance to schizophrenia. *Neuroscience*, *295*, 80–89.

Packer, A. M., & Yuste, R. (2011). Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: A canonical microcircuit for inhibition? *Journal of Neuroscience*, *31*(37), 13260–13271.

Palva, J. M., Monto, S., Kulashekhar, S., & Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(16), 7580–7585.

Paré, M., & Wurtz, R. H. (1997). Monkey posterior parietal cortex neurons antidromically activated from superior colliculus. *Journal of Neurophysiology*, *78*, 3493–3497.

Povysheva, N. V., González-Burgos, G., Zaitsev, A. V., Kröner, S., Barrionuevo, G., Lewis, D. A., & Krimer, L. S. (2006). Properties of excitatory synaptic responses in fast-spiking interneurons and pyramidal cells from monkey and rat prefrontal cortex. *Cerebral Cortex*, *16*, 541–552.

Prescott, S. A., & De Koninck, Y. (2003). Gain control of firing rate by shunting inhibition: Roles of synaptic noise and dendritic saturation. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(4), 2076–2081.

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience*, *13*(6), 766–785.

Roggeman, C., Klingberg, T., Feenstra, H. E. M., Compte, A., & Almeida, R. (2013). Trade-off between capacity and precision in visuospatial working memory. *Journal of Cognitive Neuroscience*, *26*(2), 211–222.

Salazar, R. F., Dotson, N. M., Bressler, S. L., & Gray, C. M. (2012). Content-specific fronto-parietal synchronization during visual working memory. *Science*, *338*(6110), 1097–1100.

Salin, P. A., & Prince, D. A. (1996). Spontaneous GABA$_A$ receptor-mediated inhibitory currents in adult rat somatosensory cortex. *Journal of Neurophysiology*, *75*(4), 1573–1588.

Salinas, E., & Abbott, L. F. (1996). A model of multiplicative neural responses in parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 11956–11961.

Schnotz, W., & Kurschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*, 469–508.

Self, M. W., Kooijmans, R. N., Supèr, H., Lamme, V. A., & Roelfsema, P. R.

(2012). Different glutamate receptors convey feedforward and recurrent processing in macaque v1. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(27), 11031–11036.

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*(1), 38–45.

Somers, D. C., Nelson, S. B., & Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *The Journal of Neuroscience*, *15*(8), 5448–5465.

Sreenivasan, K. K., Curtis, C. E., & DEsposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89.

Standage, D., & Paré, M. (2011). Persistent storage capability impairs decision making in a biophysical network model. *Neural Networks*, *24*, 1062–1073.

Standage, D., You, H., Wang, D.-H., & Dorris, M. C. (2013). Trading speed and accuracy by coding time: A coupled-circuit cortical model. *PLOS Computational Biology*, *9*, e100321.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science*, *12*, 257–285.

Tanaka, S. (2002). Multi-directional representation of spatial working memory in a model prefrontal cortical circuit. *Neurocomputing*, *44-46*, 1001–1008.

Thomas, N. W. D., & Paré, M. (2007). Temporal processing of saccade targets in parietal cortex area LIP during visual search. *Journal of Neurophysiology*, *97*, 942–947.

Thompson, K. G., Hanes, D. P., Bichot, N. P., & Schall, J. D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field. *Journal of Neurophysiology*, *76*, 440–455.

Tomioka, R., & Rockland, K. S. (2007). Long-distance corticocortical gabaergic neurons in the adult monkey white and gray matter. *The Journal of Comparative Neurology*, *505*, 526–538.

Troyer, T. W., & Miller, K. D. (1997). Physiological gain leads to high isi variability in a simple model of a cortical regular spiking cell. *Neural Computation*, *9*, 971–983.

Tuckwell, H. (1988). *Introduction to theoretical neurobiology*. Cambridge: Cambridge University Press.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary

memory retrieval. *Cognitive Psychology*, *71*, 1–26.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.

Voges, N., Schüz, A., Aertsen, A., & Rotter, S. (2010). A modeler's view on the spatial structure of intrinsic horizontal connectivity in the neocortex. *Progress in Neurobiology*, *92*, 277–292.

Wang, X.-J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *The Journal of Neuroscience*, *19*, 9587–9603.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences*, *24*(8), 87–185.

Wei, Z., Wang, X.-J., & Wang, D.-H. (2012). From distributed resources to limited slots in multiple-item working memory: A spiking network model with normalization. *Journal of Neuroscience*, *32*(33), 11228–11240.

White, E. L. (1989). *Cortical circuits: Synaptic organization of the cerebral cortexstructure, function and theory.* Boston: Birkhauser.

Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, *13*, 55–80.

Xiang, Z., Huguenard, J. R., & Prince, D. A. (1998). $GABA_A$ receptor-mediated currents in interneurons and pyramidal cells of rat visual cortex. *Journal of Physiology*, *506*(3), 715–730.

Xu, Y. (2007). The role of the superior intraparietal sulcus in supporting visual short-term memory for multifeature objects. *Journal of Neuroscience*, *27*(43), 11676–11686.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(8), 233–236.