eNeuro

Research Article: New Research

Cognition and Behavior

An Indexing Theory for Working Memory Based on Fast Hebbian Plasticity

Florian Fiebig,[1] Pawel Herman,[1] and Anders Lansner[1,2]

https://doi.org/10.1523/ENEURO.0374-19.2020

[1]Lansner Laboratory, Department of Computational Science and Technology, Royal Institute of Technology, 10044 Stockholm, Sweden and [2]Department of Mathematics, Stockholm University, 10691 Stockholm, Sweden

Abstract

Working memory (WM) is a key component of human memory and cognition. Computational models have been used to study the underlying neural mechanisms, but neglected the important role of short-term memory (STM) and long-term memory (LTM) interactions for WM. Here, we investigate these using a novel multiarea spiking neural network model of prefrontal cortex (PFC) and two parietotemporal cortical areas based on macaque data. We propose a WM indexing theory that explains how PFC could associate, maintain, and update multimodal LTM representations. Our simulations demonstrate how simultaneous, brief multimodal memory cues could build a temporary joint memory representation as an "index" in PFC by means of fast Hebbian synaptic plasticity. This index can then reactivate spontaneously and thereby also the associated LTM representations. Cueing one LTM item rapidly pattern completes the associated uncued item via PFC. The PFC–STM network updates flexibly as new stimuli arrive, thereby gradually overwriting older representations.

Key words: computational model; long-term memory; short-term memory; spiking neural network; synaptic plasticity; working memory

Significance Statement

Most, if not all, computational working memory (WM) models have focused on short-term memory (STM) aspects. However, from the cognitive perspective the interaction of STM with long-term memory (LTM) bears particular relevance since the WM-activated LTM representations are considered central to flexible cognition. Here we present a large-scale biologically detailed spiking neural network model accounting for three connected cortical areas to study dynamic STM–LTM interactions that reflect the underlying theoretical concept of memory indexing, adapted to support distributed cortical WM. Our cortex model is constrained by relevant experimental data about cortical neurons, synapses, modularity, and connectivity. It demonstrates encoding, maintenance, and flexible updating of multiple items in WM as no single model has done before. It thereby bridges microscopic synaptic effects with macroscopic memory dynamics, and reproduces several key neural phenomena reported in WM experiments.

Introduction

By working memory (WM) we typically understand a flexible but volatile kind of memory capable of holding a small number of items over short time spans, allowing us to act beyond the immediate here and now. WM is thus a key component in cognition and is often affected early on in neurologic and psychiatric conditions (e.g., Alzheimer's disease and schizophrenia; Slifstein et al., 2015). Although prefrontal cortex (PFC) has consistently been implicated as a key neural substrate for WM in humans and nonhuman primates (Fuster, 2009; D'Esposito and

Received September 17, 2019; accepted January 27, 2020; First published February 28, 2020.

Author contributions: F.F. and A.L. designed research; F.F. performed research; F.F. and P.H. analyzed data; F.F., P.H., and A.L. wrote the paper.

This work was supported by the EuroSPIN Erasmus Mundus doctoral program, SeRC (Swedish e-science Research Center), and StratNeuro (Strategic Area Neuroscience at Karolinska Institutet, Umeå University and KTH Royal Institute of Technology). The simulations were performed using computing resources provided by the Swedish National Infrastructure for Computing (SNIC) at PDC Centre for High Performance Computing.

Postle, 2015), there is accumulated evidence for the involvement of other cortical regions, particularly parietotemporal networks associated with long-term memory (LTM) correlates. Consequently, there is growing understanding that WM function emerges from the interactions between dynamically coupled short-term memory (STM) and LTM systems (Eriksson et al., 2015; Sreenivasan and D'Esposito, 2019), which enable activation or the "bringing online" of a small set of task-relevant LTM representations (Eriksson et al., 2015). This prominent effect is envisaged to underlie complex cognitive phenomena, which are reported in experiments on humans as well as animals. Nevertheless, since there is limited availability of multiarea mesoscopic recordings of neural activity during WM, the neural mechanisms involved remain elusive. Furthermore, computational models of WM have so far focused solely on its short-term memory aspects, explained either by means of persistent activity (Funahashi et al., 1989; Goldman-Rakic, 1995; Camperi and Wang, 1998; Compte et al., 2000) or, more recently, fast synaptic plasticity (Mongillo et al., 2008; Lundqvist et al., 2011; Fiebig and Lansner, 2017), and there are no detailed hypotheses about neural underpinnings of the operational STM–LTM interplay in the service of WM.

To address this gap and draw attention to the wider cognitive perspective of WM accounting for more than STM correlates in PFC, we present a large-scale multiarea spiking neural network model of WM and focus on investigating the neural mechanisms behind the fundamental STM–LTM interactions critical to WM function. Our model comprises a subsampled PFC network model of STM that is reciprocally connected with two LTM component networks representing different sensory modalities (e.g., visual and auditory) in parietotemporal cortical areas. This new model exploits the architecture of a recent PFC-dependent STM model of human word-list learning (Fiebig and Lansner, 2017), shown to reproduce a range of patterns of mesoscopic neural activity observed in WM experiments. It uses the same fast Hebbian plasticity as a key neural mechanism, intrinsically within PFC but also in PFC backprojections that target parietotemporal LTM stores. The core idea of our theory rests on the concept of cell assemblies formed in the PFC, as STM correlates, by means of fast Hebbian plasticity that serve as "indices" linking LTM representations. The associative plasticity in this functional context has to be induced and expressed on a timescale of a few hundred milliseconds. Recent experiments have demonstrated the existence of fast forms of Hebbian synaptic plasticity (e.g., short-term potentiation or labile LTP; Erickson et al., 2010; Park et al., 2014; Pradier et al., 2018), which lends credibility to this type of WM mechanism.

Acknowledgments: We thank Drs. Jeanette Hellgren Kotaleski and Arvind Kumar for helpful comments and suggestions.

Correspondence should be addressed to Anders Lansner at ala@kth.se.

https://doi.org/10.1523/ENEURO.0374-19.2020

The proposed concept of distributed WM resting on the dynamical STM–LTM interactions, mediated by fast synaptic plasticity, draws inspiration from the hippocampal memory indexing theory (Teyler and DiScenna, 1986), originally proposed to account for the role of hippocampus in storing episodic memories (Teyler and Rudy, 2007). Binding and indexing of neural representations have been a common recurring theme in memory research, in particular in relation to the role of hippocampus and surrounding structures (Squire, 1992; O'Reilly and Frank, 2006; Teyler and Rudy, 2007). We therefore adapt this theoretical notion and formulate a cortical indexing theory of WM, thereby reflecting a more general computational principle of indexing that supports multiarea memory phenomena. Our main novel contribution here is to show that a neurobiologically constrained large-scale spiking neural network model of interacting cortical areas via biologically realistic sparse connectivity can function as a robust and flexible multi-item and cross-modal WM. This includes its important role of bringing relevant LTM representations temporarily online by means of "indexing," and thus to computationally validate the proposed concept of distributed WM. In addition, the model replicates many experimentally observed effects in terms of oscillations, coherence, and latency within and between cortical regions, and offers new macroscopic predictions about large-scale internetwork dynamics as a neural correlate of WM operations. Interestingly, it can also explain the so far poorly understood cognitive phenomenon of variable binding or object–name association, which is one key ingredient in human reasoning and planning (Cer and O'Reily, 2012; Pinkas et al., 2013; van der Velde and de Kamps, 2015).

Materials and Methods

Neuron model

We use an integrate-and-fire point neuron model with spike–frequency adaptation (Brette and Gerstner, 2005), which was modified by Tully et al. (2014) for compatibility with a custom-made Bayesian Confidence Propagation Neural Network (BCPNN) synapse model in NEST (see Simulation environment) through the addition of the intrinsic excitability current I_{β_j}. The model was simplified by excluding the subthreshold adaptation dynamics. Membrane potential (V_m) and adaptation current are described by the following equations:

$$-C_m \frac{dv_m}{dt} = -g_L(V_m - E_L) + g_L \Delta_T e^{\frac{v_m - v_t}{\Delta_T}} - I_w(t) - I_{tot}(t) + I_{\beta_j} + I_{ext} \quad (1)$$

$$\frac{dI_w(t)}{dt} = \frac{-I_w(t)}{\tau_{I_w}} + b\delta(t - t_{sp}). \quad (2)$$

The membrane voltage changes through incoming currents over the membrane capacitance (C_m). A leak reversal potential (E_L) drives a leak current through the conductance (g_L), and an upstroke slope factor (Δ_T) determines the sharpness of the spike threshold (V_t). Spikes are followed by a reset of membrane potential to V_r. Each

spike increments the adaptation current by b, which decays with time constant τ_{I_w}. Simulated basket cells feature neither the intrinsic excitability current I_{β_j} nor this spike-triggered adaptation.

In addition to external input I_{ext} (see Stimulation protocol), neurons receive a number of different synaptic currents from their presynaptic neurons in the network (AMPA, NMDA, and GABA), which are summed at the membrane accordingly:

$$I_{tot_j}(t) = \sum_{syn}\sum_{i} g_{ij}^{syn}(t)(V_{m_j} - E_{ij}^{sym}) = I_j^{AMPA}(t) + I_j^{NMDA}(t) + I_j^{GABA}(t). \quad (3)$$

Synapse model

Excitatory AMPA and NMDA synapses have a reversal potential $E^{AMPA} = E^{NMDA}$, while inhibitory synapses drive the membrane potential toward E^{GABA}. Every presynaptic input spike (at t_{sp}^i with transmission delay t_{ij}) evokes a transient synaptic current through a change in synaptic conductance that follows an exponential decay with time constants τ^{syn} depending on the synapse type ($\tau^{\text{AMPA}} \ll \tau^{\text{NMBA}}$), as follows:

$$g_{ij}^{syn}(t) = x_{ij}^{dep}(t) w_{ij}^{syn} e^{-\frac{t-t^t-t_{ij}}{\tau^{syn}}} H(t - t_{sp}^i - t_{ij}). \quad (4)$$

$H(\cdot)$ is the Heaviside step function. w_{ij}^{syn} is the peak amplitude of the conductance transient, learned by the spike-based BCPNN learning rule (next section). Plastic synapses are also subject to synaptic depression (vesicle depletion) according to the Tsodyks–Markram formalism (Tsodyks and Markram, 1997), modeling the transmission-dependent depletion of available synaptic resources x_{ij}^{dep} by a utilization factor U, and a depression/reuptake time constant τ_{rec}, as follows:

$$\frac{dx_{ij}^{dep}}{dt} = \frac{1 - x_{ij}^{dep}}{\tau_{rec}} - U x_{ij}^{dep} \sum_{sp} \delta(t - t_{sp}^i - t_{ij}). \quad (5)$$

Spike-based BCPNN learning rule

Plastic AMPA and NMDA synapses are modeled to mimic NMDA-dependent Hebbian short-term potentiation (Erickson et al., 2010) with a spike-based version of the BCPNN learning rule (Wahlgren and Lansner, 2001; Tully et al., 2014). For a full derivation from Bayes rule, deeper biological motivation, and proof of concept, see Tully et al. (2014) and an earlier STM model implementation by Fiebig and Lansner (2017).

Briefly, the BCPNN learning rule makes use of biophysically plausible local traces to estimate normalized presynaptic and postsynaptic firing rates, as well as coactivation, which can be combined to implement Bayesian inference because connection strengths and neural unit activations have a statistical interpretation (Sandberg et al., 2002; Fiebig and Lansner, 2014; Tully et al., 2014). Crucial parameters include the synaptic activation trace Z, which is computed from spike trains via presynaptic and postsynaptic time constants $\tau_{z_i}^{syn}$, $\tau_{z_j}^{syn}$, which are the same here but differ between AMPA and NMDA synapses, as follows:

$$\tau_{z_i}^{AMPA} = \tau_{z_j}^{AMPA} = 5ms, \quad \tau_{z_i}^{NMDA} = \tau_{z_j}^{NMDA} = 100ms. \quad (6)$$

The larger NMDA time constant reflects the slower closing dynamics of NMDA receptor-gated channels. All excitatory connections are drawn as AMPA and NMDA pairs, such that they feature both components. Further filtering of the Z traces leads to rapidly expressing memory traces (referred to as P-traces) that estimate activation and coactivation as follows:

$$\tau_p \frac{dP_i}{dt} = \kappa(Z_i - P_i), \quad \tau_p \frac{dP_j}{dt} = \kappa(Z_j - P_j), \quad \tau_p \frac{dP_{ij}}{dt} = \kappa(z_i z_j - P_{ij}). \quad (7)$$

These traces constitute memory itself and decay in a palimpsest fashion. Short-term potentiation decay is known to take place on timescales that are highly variable and activity dependent (Volianskis et al., 2015; see Discussion, The case for Hebbian plasticity).

We make use of the learning rule parameter κ (Eq. 7), which may reflect the action of endogenous neuromodulators [e.g., dopamine (DA) acting on D_1 receptors (D1Rs)] that signal relevance and thus modulate learning efficacy). It can be dynamically modulated to switch off learning to fixate the network or temporarily increase plasticity (κ_{encoding}, κ_{normal}; Table 1). In particular, we trigger a transient increase of plasticity concurrent with external stimulation.

Tully et al. (2014) showed that Bayesian inference can be recast and implemented in a network using the spike-based BCPNN learning rule. Prior activation levels are realized as an intrinsic excitability of each postsynaptic neuron, which is derived from the postsynaptic firing rate estimate p_j and implemented in the NEST neural simulator (Gewaltig and Diesmann, 2007) as an individual neural current I_{β_j} with scaling constant β_{gain}:

$$I_{\beta_j} = \beta_{\text{gain}} \log(P_j). \quad (8)$$

I_{β_j} is thus an activity-dependent intrinsic membrane current to the neurons, similar to the A-type potassium channel (Hoffman et al., 1997) or TRP channel (Petersson et al., 2011). Synaptic weights are modeled as peak amplitudes of the conductance transient (Eq. 4) and determined from the logarithmic BCPNN weight, as derived from the P-traces with a synaptic scaling constant w_{gain}^{syn}, as follows:

$$w_{ij}^{syn} = w_{gain}^{syn} \log \frac{p_{ij}}{p_i p_j}. \quad (9)$$

In this model, AMPA and NMDA synapses make use of w_{gain}^{AMPA} and w_{gain}^{NMDA}, respectively. The logarithm in Equations 8 and 9 is motivated by the Bayesian underpinnings of the learning rule and means that synaptic weights w_{ij}^{syn} multiplex both the learning of excitatory and disynaptic inhibitory interaction. The positive weight component is here interpreted as the conductance of a monosynaptic

Table 1: Neurons, synapses, and plasticity

Adaptation current	b	86 pA	Depression time constant	τ_{rec}	500 ms	BCPNN AMPA gain	w_{gain}^{AMPA}	3.93 nS
Adaptation time constant	τ_{I_w}	500 ms	AMPA synaptic time constant	τ^{AMPA}	5 ms	BCPNN NMDA gain	w_{gain}^{NMDA}	0.21 nS
Membrane capacity	C_m	280 pF	NMDA synaptic time constant	τ^{NMDA}	100 ms	BCPNN bias current gain	β_{gain}	90 *pA*
Leak reversal potential	E_L	−70 mV	GABA synaptic time constant	τ^{GABA}	5 ms	BCPNN lowest rate	f_{min}	0.2 Hz
Leak conductance	g_L	14 pS	AMPA reversal potential	E^{AMPA}	0 mV	BCPNN highest rate	f_{max}	20 Hz
Upstroke slope factor	Δ_T	3 mV	NMDA reversal potential	E^{NMDA}	0 mV	BCPNN lowest probability	ε	0.01
Spike threshold	V_t	−55 mV	GABA reversal potential	E^{GABA}	−75 mV	BCPNN Spike event duration	Δt	1 ms
Spike reset potential	V_r	−80 mV	Dopaminergic modulation	$\kappa_{encoding}$	6.0	P-trace time constant	τ_p	5 s
Utilization factor	U	0.33	Regular plasticity	κ_{normal}	1.0			

excitatory pyramidal to pyramidal synapse [Fig. 1, plastic connection to the coactivated minicolumn (MC)], while the negative component (Fig. 1, plastic connection to the competing MC) is interpreted as disynaptic via a dendritic targeting and vertically projecting inhibitory interneuron like a double bouquet and/or bipolar cell (Tucker and Katz, 2003; Kapfer et al., 2007; Ren et al., 2007; Silberberg and Markram, 2007). Accordingly, BCPNN connections with a negative weight use a GABAergic reversal potential instead, as in previously published models of this kind (Tully et al., 2014, 2016; Fiebig and Lansner, 2017). Model networks with negative synaptic weights have been shown to be functionally equivalent to those with both excitatory and inhibitory neurons with only positive weights (Parisien et al., 2008). In the context of this particular model microcircuit and learning rule, this was explicitly and conclusively demonstrated by the addition of double bouquet cells (Chrysanthidis et al., 2019).

Code for the NEST implementation of the BCPNN synapse is openly available (see Code accessibility).

Axonal conduction delays

We compute axonal delays t_{ij} between presynaptic neuron i and postsynaptic neuron j, based on a constant conduction velocity V and the Euclidean distance between respective columns. Conduction delays were randomly drawn from a normal distribution with mean according to the connection distance divided by conduction speed

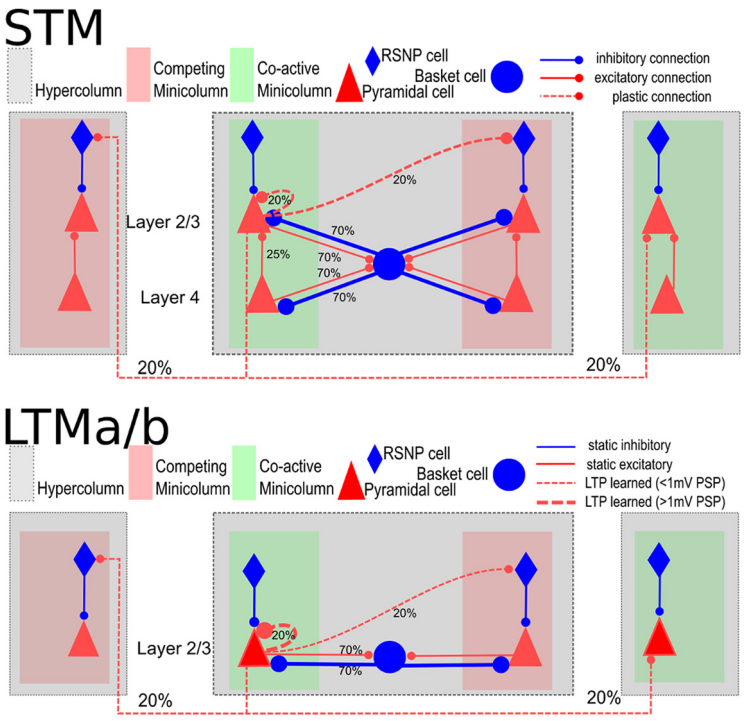

Figure 1. Local columnar connectivity within STM and LTM. Connection probabilities are given by the percentages; further details are in Tables 1, 2, and 3. The strength of plastic connections develops according to the synaptic learning rule described in the spike-based BCPNN learning rule. Initial weights are low and distributed by a noise-based initialization procedure (see Stimulation protocol). However, dashed connections are not plastic in LTM (besides the synaptic depression of Eq. 5), but already encode memory patterns previously learned through an LTP protocol, and loaded before the simulation using receptor-specific weights found in Table 2.

Table 2: Network size, conduction delay, stimulation, and LTM preload BCPNN weights

STM patch size	17×17 mm		Initialization input rate layer 2/3	r^{L23}_{bg-low}	550 Hz
Simulated HCs	n^{STM}_{HC}	25	Background activity rate layer 2/3	r^{L23}_{bg}	625 Hz
Simulated MC per HC	n^{STM}_{MC}	12	Background activity rate layer 4	r^{24}_{bg}	300 Hz
LTM patch size	25×25 mm		High Background activity rate layer 2/3 (e.g., STM maintenance)	$r^{L23}_{bg-high}$	950 Hz
Simulated HCs	n^{LTM}_{HC}	16			
Simulated MC per HC	n^{LTM}_{MC}	9	Background conductance	g_{bg}	$\pm$1.5 nS
Axonal conduction speed	V	$2\frac{m}{s}$			
Minimal conduction delay	t^{syn}_{min}	1.5 ms	Cue stimulus duration	t_{cue}	50 ms
STM–LTM distance	$d_{STM\text{-}LTM}$	40 mm	Stimulation rate	r_{cue}	650 Hz
Hypercolumn diameter	d_{HC}	0.64 mm	Cue stimulus conductance	g_{cue}	+1.5 nS
Layer 2 pyramidal per MC	n^{PYR-L2}_{MC}	20	LTM intra-HC–intra-MC weight	$w^{IntraHC}_{IntraMC}$	3.36 w^{syn}_{gain}
Layer 3A pyramidal per MC	$n^{PYR-L3A}_{MC}$	5	LTM intra-HC–inter-MC weight	$w^{IntraHC}_{InterMC}$	$-$4.82 w^{syn}_{gain}
Layer 3B pyramidal per MC	$n^{PYR-L3B}_{MC}$	5			
Layer 4 pyramidal per MC	n^{PYR-L4}_{MC}	30	LTM inter-HC–coactive MC weight	$w^{InterHC}_{CoactiveMC}$	3.08 w^{syn}_{gain}
Basket cells per MC	n^{basket}_{MC}	4	LTM inter-HC–competing MC weight	$w^{Inter\,HC}_{Competin\,gMC}$	$-$4.28 w^{syn}_{gain}

Layer 4 not simulated in LTM.

and with a relative SD of 15% of the mean in order to account for individual arborization differences and varying conduction speeds as a result of axonal thickness/myelination. Further, we add a minimal conduction delay t^{syn}_{min} of 1.5 ms to reflect not directly modeled delays, such as diffusion of transmitter over the synaptic cleft, dendritic branching, thickness of the cortical sheet, and the spatial extent of columns, as follows:

$$\overline{t_{ij}} = \frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{V} + t^{syn}_{mn} ms \quad t_{ij} \sim N(\overline{t_{ij}}, .15\overline{t_{ij}}). \tag{10}$$

STM network architecture

The model organizes cells in the three simulated cortical areas into grids of nested hypercolumns (HCs) and MCs, sometimes referred to as macro columns, and "functional columns," respectively. The STM network is simulated with $n^{STM}_{HC} = 25$ HCs spread out on a grid with spatial extent of 17×17 mm. This spatially distributed network of columns has sizable conduction delays due to the distance between columns and can be interpreted as a spatially distributed subsampling of columns from the extent of dorsolateral PFC (e.g., BA 46 and 9/46, which also have a combined spatial extent of ~289 mm^2 in macaque).

Each of the nonoverlapping HCs has a diameter of ~640 μm, comparable to estimates of cortical column size (Mountcastle, 1997), contains 48 basket cells, and its pyramidal cell population has been divided into 12 MCs. This constitutes another subsampling from the ~100 MCs per HC when mapping the model to biological cortex. We simulate 20 pyramidal neurons per MC to represent approximately the layer 2 population of an MC, 5 cells for layer 3A, 5 cells for layer 3B, and another 30 pyramidal cells for layer 4, as macaque BA 46 and 9/46 have a well developed granular layer (Petrides and Pandya, 1999). The STM model thus contains ~18,000 simulated pyramidal cells in four layers (although layers 2, 3A, and 3B are often treated as one layer 2/3).

STM network connectivity

The most relevant connectivity parameters are found in Tables 1, 2, and 3. Pyramidal cells project laterally to basket cells within their own HC via AMPA-mediated excitatory projections with a connection probability of p_{p-B} (i.e., connections are randomly drawn without duplicates until the target fraction of all possible pre–post connections exist). In turn, they receive GABAergic feedback (FB) inhibition from basket cells (p_{B-p}) that connect via static inhibitory synapses rather than plastic BCPNN synapses. This strong loop implements a competitive soft WTA (winner-take-all) subnetwork within each HC (Douglas and Martin, 2004). Local basket cells fire in rapid bursts, and induce alpha/beta oscillations in the absence of attractor activity and gamma, when attractors are present and active.

Pyramidal cells in layer 2/3 form connections both within and across HCs at connection probability $p_{L23e\text{-}L23e}$. These projections are implemented with plastic synapses and contain both AMPA and NMDA components, as explained in the subsection Spike-based BCPNN learning rule. Connections across columns and areas may feature sizable conduction delays due to the implied spatial distance between them (Table 1).

Pyramidal cells in layer 4 project to pyramidal cells of layer 2/3, targeting 25% of cells within their respective MC only. Experimental characterization of excitatory connections from layer 4 to layer 2/3 pyramidal cells have confirmed similarly high fine-scale specificity in rodent cortex (Yoshimura and Callaway, 2005) and, in turn, full-

Table 3: Projections

Scope	Source	Target	Type	Symbol	Value
Cortical area	Pyramidal	Basket	Probability	$p_{P\text{-}B}$	0.7
	Pyramidal	Basket	Conductance (static)	$g_{P\text{-}B}$	+3.5 nS
	Basket	Pyramidal	Probability	$p_{B\text{-}P}$	0.7
	Basket	Pyramidal	Conductance (static)	$g_{B\text{-}P}$	−20 nS
	L23e	L23e	Probability	$p_{L23e\text{-}L23e}$	0.2
	L23e	L23e	AMPA gain (BCPNN)	w^{AMPA}_{gain}	3.93nS
	L23e	L23e	NMDA gain (BCPNN)	w^{NMDA}_{gain}	0.21nS
	L4e	L23e	Probability	$p_{L4e\text{-}L23e}$	0.25
	L4e	L23e	Conductance (static)	$g_{L4e\text{-}L23e}$	25 nS
Feed forward	LTM L3Ae	STM MC	Probability	$p^{FF}_{L3Ae-MC}$	0.0015
	LTM L3Ae	STM MC	Branching factor	$b^{FF}_{L3Ae-MC}$	0.25
	LTM L3Ae	STM L23e	Conductance (static)	$g^{FF}_{L3Ae-L23e}$	±7.2 nS
	LTM L3Ae	STM L4e	Conductance (static)	$g^{FF}_{L3Ae-L4e}$	±7.2 nS
Feedback	STM PYR	LTM PYR	Probability	p^{FB}_{P-P}	0.0066
	STM L3Be	LTM HC	Branching factor	$b^{FB}_{L3Be-HC}$	0.25
	STM L3Be	LTM L23e	AMPA gain (BCPNN)	w^{AMPA}_{FB}	7.07 nS
	STM L3Be	LTM L23e	NMDA gain (BCPNN)	w^{NMDA}_{FB}	0.4 nS

scale cortical simulation models without functional columns have found it necessary to specifically strengthen these connections to achieve defensible firing rates (Potjans and Diesmann, 2014).

In summary, the STM model thus features a total of 16.2 million plastic AMPA- and NMDA-mediated connections between its 18,000 simulated pyramidal cells, as well as 67,500 static connections from 9000 layer four pyramidals to layer 2/3 targets within their respective MC, and 1.2 million static connections to and from 1200 simulated basket cells.

LTM network

We simulate two structurally identical LTM networks, referred to as LTMa and LTMb. LTM networks may be interpreted as a spatially distributed subsampling of columns from areas of the parietotemporal cortex commonly associated with modal LTM stores. For example, inferior temporal cortex (ITC) is often referred to as the storehouse of visual LTM (Miyashita, 1993). Two such LTM areas are indicated in Figure 2.

We simulate $n^{LTM}_{HC} = 16$ HCs in each area and 9 MCs per HC (Tables 1, 2, 3, for further details). Both LTM networks are structurally very similar to the previously described STM, yet they do not feature plasticity among their own cells, beyond short-term dynamics in the form of synaptic depression. Unlike STM, LTM areas also do not feature an input layer 4, but are instead stimulated directly to cue the activation of previously learned long-term memories (see Stimulation protocol). Various previous models with identical architecture have demonstrated how attractors can be learned via plastic BCPNN synapses (Lansner et al., 2013; Tully et al., 2014, 2016; Fiebig and Lansner, 2017). We load each LTM network with nine orthogonal attractors [see Fig. 4*B*, 10 in the example (which features two sets of five memories each)]. Each memory pattern consists of 16 active MCs, distributed across the 16 HCs of the network. We load in BCPNN weights from a previously trained network (Table 2), but thereafter set $\kappa = 0$ to deactivate plasticity of recurrent connections in LTM stores.

In summary, the two LTM models thus feature a total of 7.46 million connections between 8640 pyramidal cells, as well as 435,456 static connections to and from 1152 basket cells.

Interarea connectivity

In this model, we focus on layers 2/3, as its high degree of recurrent connectivity (Thomson et al., 2002; Yoshimura and Callaway, 2005) supports attractor function. The high fine-scale specificity of dense stellate cell (Yoshimura et al., 2005) and double-bouquet cell inputs (DeFelipe et al., 2006; Chrysanthidis et al., 2019) enable strongly coding subpopulations in the superior layers of functional columns. This fits with the general observation that layers 2/3 are more input selective than the lower layers (Crochet and Petersen, 2009; Sakata and Harris, 2009) and thus of more immediate concern to our computational model.

The recent characterization of supragranular feedforward (FF) and FB projections (from large cells in layer 3B and 3A, respectively), between association cortices and at short and medium cortical distances (Markov et al., 2014), allows for the construction of a basic cortical hierarchy without explicit representation of infragranular layers (and its long-range FB projections from large cells in layer 5 and 6). This is not to say that nothing would be gained by explicitly modeling infragranular layers, but it would go beyond the scope of this model.

Accordingly, our model implements supragranular FF and FB pathways between cortical areas that are at a medium distance in the cortical hierarchy. The approximate

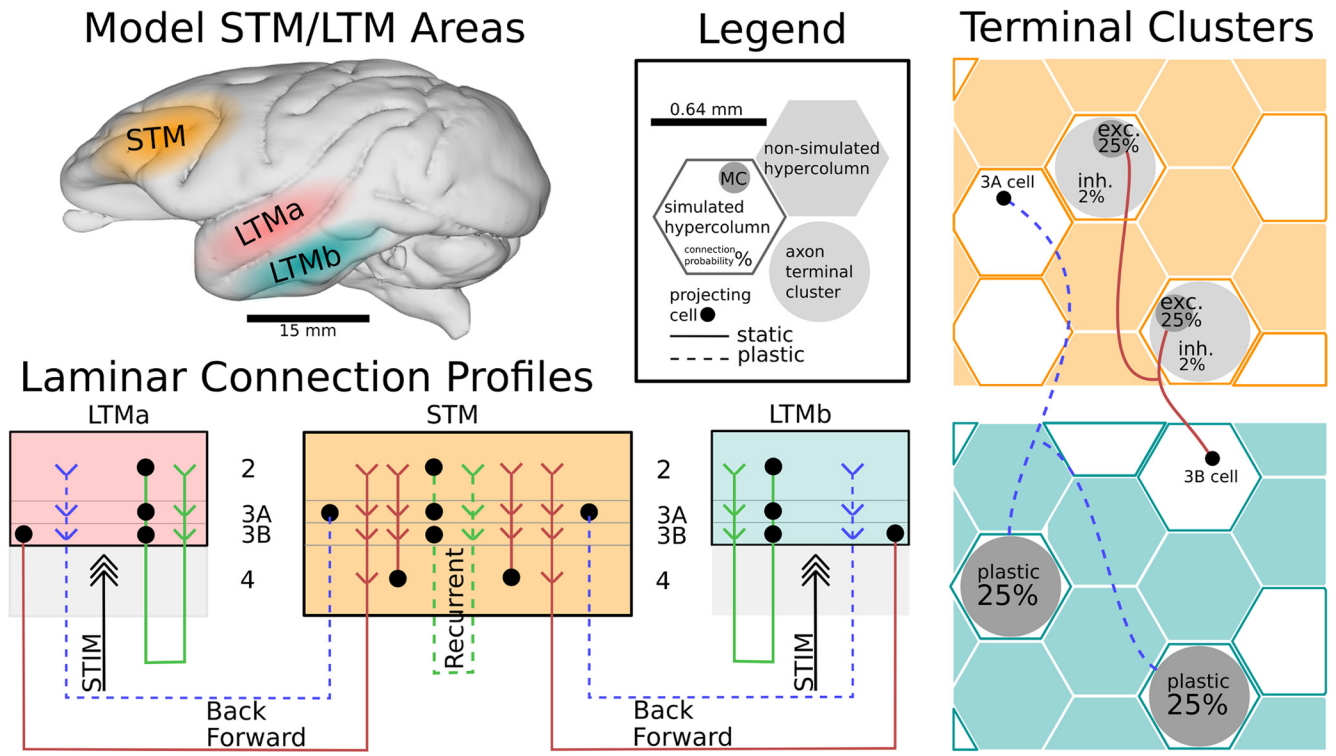

Figure 2. Schematic of modeled connectivity within and across representative STM and LTM areas in macaque. STM features 25 HCs, whereas LTMa and LTMb both contain 16 simulated HCs. Each network spans several hundred square millimeters, and the simulated columns constitute a spatially distributed subsample of biological cortex, defined by conduction delays. Pyramidal cells in the simulated supragranular layers form connections both within and across columns. STM features an input layer 4 that shapes the input response of cortical columns, whereas LTM is instead stimulated directly to cue the activation of previously learned long-term memories. Additional corticocortical connections (feedforward in brown, feedback in dashed blue) are sparse (<1% connection probability) and implemented with terminal clusters (rightmost panels) and specific laminar connection profiles (bottom left). The connection schematic illustrates laminar connections realizing a direct supragranular forward-projection, as well as a common supragranular backprojection. Layer 2/3 recurrent connections in STM (dashed green) and corticocortical backprojections (dashed blue) feature fast Hebbian plasticity. For an in-depth model description, including the columnar microcircuits, please refer to Materials and Methods and Figure 1.

cortical distance between ITC and dlPFC in macaque is ~40 mm and with an axonal conductance speed of 2 m/s, distributed conduction delays in our model (Eq. 10) average just >20 ms between these areas (Girard et al., 2001; Thorpe and Fabre-Thorpe, 2001; Caminiti et al., 2013).

In the forward path, layer 3B cells in LTM project toward STM (Fig. 2). We do not draw these connections one by one, but as branching axons targeting 25% of the pyramidal cells in a randomly chosen MC (the chance of any layer 3B cell to target any MC in STM is only 0.15%). The resulting split between targets in layer 2/3 and 4 is typical for FF connections at medium distances in the cortical hierarchy (Markov et al., 2014) and has important functional implications for the model (LTM-to-STM forward dynamics). We also branch off some inhibitory corticocortical connections as follows: for every excitatory connection within the selected targeted MC, an inhibitory connection is created from the same pyramidal layer 3B source cell onto a randomly selected cell outside the targeted MC, but inside the local HC. This way of drawing random forward-projections retains a degree of functional specificity due to its spatial clustering and yields patchy sparse forward-projections as observed in the cortex (Houzel et al., 1994; Voges et al., 2010), with a resulting interarea connection probability of only 0.0125% (648 axonal projections from L3B cells to STM layers 2/3 and 4 results in ~20,000 total connections after branching, as described above.

In the FB path, we draw sparse plastic connections from layer 3A cells in STM to layer 2/3 cells in LTM: branching axons target 25% of the pyramidal cells in a randomly chosen HC in LTM, simulating a degree of axonal branching found in the literature (Zufferey et al., 1999). Using this method, we obtain biologically plausible sparse and structured FB projections with an interarea connection probability of 0.66%, which, unlike the forward pathway, do not have any built-in MC specificity but may develop such through activity-dependent plasticity. More parameters on corticocortical projections can be found in Table 3. On average, each LTM pyramidal cell receives ~120 corticocortical connections from STM. Because ~5% of STM cells fire together during memory reactivation (see Results), this means that a mere 6 active synapses per target cell are sufficient for driving (and thus maintaining) LTM activity from STM (there are 96 active synapses from coactive pyramidal cells in LTM).

Notably LTMa and LTMb have no direct pathways connecting them in our model since we assume that the use of previously not associated stimuli in our simulated multimodal tasks and further, that plasticity of biological connections between them are likely too slow (LTP timescale) to make a difference in WM dynamics. This arrangement also guarantees that any binding of long-term memories across LTM areas must be the result of interaction via STM instead. Overall in our model, corticocortical connectivity is very sparse, <1% on a cell-to-cell basis.

Stimulation protocol

The term I_{ext} in Equation 1 subsumes specific and unspecific external inputs. To simulate unspecific input from nonsimulated columns, and other areas, pyramidal cells are continually stimulated with a zero mean noise background throughout the simulation. In each layer, two independent Poisson sources generate spikes at rate r_{bg}^{layer} and connect onto all pyramidal neurons in that layer, via nondepressing conductances $\pm g_{bg}$ (Table 2). Before each simulation, we distribute the initial values of all plastic weights by a process of learning from 1.5 s low, unstructured background activity (Table 2; r_{bg-low}^{L23}). To cue the activation of a specific memory pattern (i.e., attractor), we excite LTM pyramidal cells belonging to a memory patterns component MC with an additional excitatory Poisson spike train (rate, r_{cue}, length, t_{cue}; conductance, g_{cue}). As LTM patterns are strongly encoded in each LTM, a brief 50 ms stimulus is usually sufficient to activate any given memory.

Synthetic field potentials and spectral analysis

We estimate local field potentials (LFPs) by calculating a temporal derivative of the average low-pass filtered (cutoff frequency at 250 Hz) potential for all pyramidal cells in local populations at every time step, similarly to the approach adopted by Ursino and La Cara (2006). Although LFP is more directly linked to the synaptic activity (Logothetis, 2003), the averaged membrane potentials have been reported to be correlated with LFPs (Okun et al., 2010). In particular, low pass-filtered components of synaptic currents reflected in differentiated membrane potentials appear to carry the portion of the power spectral content of extracellular potentials that is relevant to our key findings (Lindén et al., 2010). As regards the phase response of estimated extracellular potentials, the delays of different frequency components are spatially dependent (Lindén et al., 2010). However, irrespective of the LFP synthesis, the phase-related phenomena reported in this study remain qualitatively unaffected since they hinge on relative rather than absolute phase values.

Most spectral analyses have been conducted on the synthesized field potentials with the exception of population firing rates, shown in Figure 3, *A* and *B*. Spectral information is extracted with a multitaper approach using a family of orthogonal tapers produced by Slepian functions (Slepian, 1978; Thomson, 1982), with frequency-dependent window lengths corresponding to five to eight oscillatory cycles and frequency smoothing corresponding to 0.3–0.4 of the central frequency, which was sampled with the resolution of 1 Hz (this configuration implies that two to three tapers are usually used). To obtain the spectral density, spectrotemporal content is averaged within a specific time interval.

The coherence for a pair of synthesized field potentials at the spatial resolution corresponding to a hypercolumn was calculated using the multitaper auto-spectral and cross-spectral estimates. The complex value of coherence (Carter, 1987) was evaluated first based on the spectral components averaged within 0.5 s windows. Next, its magnitude was extracted to produce the time-windowed estimate of the coherence amplitude. In addition, phase-locking statistics were estimated to examine synchrony without the interference of amplitude correlations (Lachaux et al., 1999; Palva et al., 2005). In particular, the phase-locking value (PLV) between two signals with instantaneous phases $\Phi_1(t)$ and $\Phi_2(t)$ was evaluated within a time window of size $N = 0.5$ s as follows:

$$\mathrm{PLV} = \frac{1}{N}\left|\sum_{i=1}^{N}\exp\left(j\left(\Phi_1(t_i) - \Phi_2(t_i)\right)\right)\right|.$$

The instantaneous phase of the signals was estimated from their analytic signal representation obtained using a Hilbert transform. Before the transform was applied, the signals were narrow band filtered with low-time domain spread, finite-impulse response filters (in the forward and reverse directions to avoid any phase distortions). The analysis was performed mainly for gamma-range oscillations. A continuous PLV estimate was obtained with a sliding window approach, and the average along with SE were calculated typically over 25 trials.

Spike train analysis and memory activity tracking

We track memory activity in time by analyzing the population firing rate of pattern-specific and network-wide spiking activity usually using an exponential moving average filter time constant of 20 ms. We do not use an otherwise common low-pass filter with symmetrical window, because we are particularly interested in characterizing activation onsets and onset delays. As activations are characterized by sizable gamma-like bursts, a simple threshold detector can extract candidate activation events and decode the activated memory. This is trivial in LTM due to the known nature of its patterns. In STM, we decode the stimulus specificity of each cell individually by finding the maximum correlation between input pattern and the untrained STM spiking response in the 320 ms following cue onset (which is the stimulation interval during the plasticity-modulated stimulation period; Fig. 3*D*) following the pattern cue to LTM. Thereafter, we can filter the population response of cells in STM with the same selectivity on that basis to obtain a more robust readout. We validate the specificity by means of cross-correlations, which reveal that the pattern-specific populations are rather orthogonal according to the covariance matrix (off-diagonal magnitude, <0.1). In all three networks, we measure the onset and offset of pattern activity by thresholding each individual activation at half of its population peak firing rate. In LTM, we further check

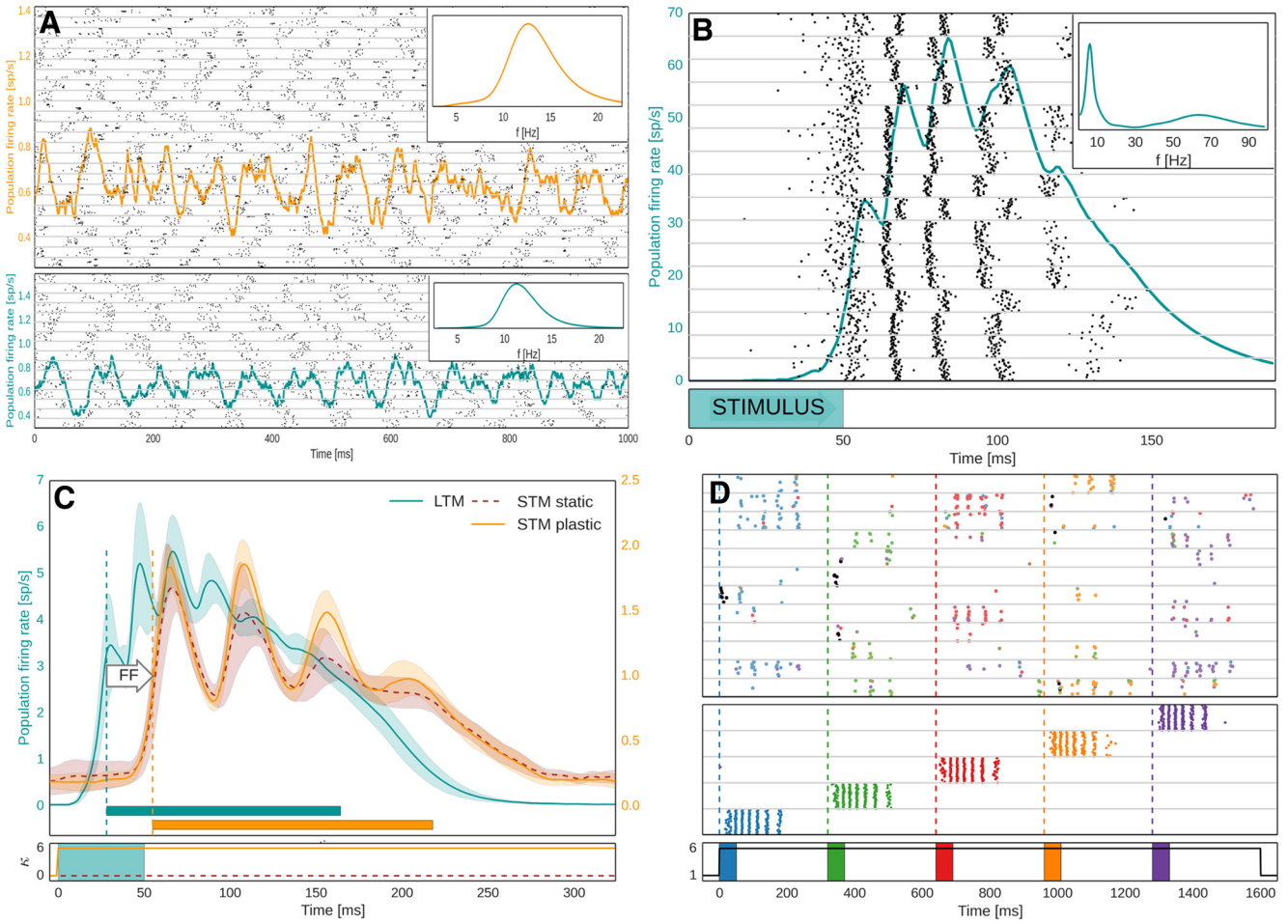

Figure 3. Basic network behavior in spike rasters and population firing rates. Activity in the untrained network under strong background input. ***A***, Subsampled spike raster of STM (top) and LTM (bottom) layer 2/3 activity. HCs are separated by gray horizontal lines. Global oscillations in the alpha range (10–13 Hz) characterize this activity state in both STM (top) and LTM (bottom) in the absence of attractors. Inset, Power spectral density of LFP of each network. ***B***, Cued LTM memory activation express as fast oscillation bursts of selective cells (50–80 Hz), organized into a theta-like envelope (4–8 Hz), see also power spectrum inset. The gamma band is broad due to the varying lengths of the underlying cycles (i.e., noticeably increasing over the short memory activation period). The underlying spike raster shows layer 2/3 activity of the activated MC in each HC, revealing spatial synchronization. The brief stimulus is a memory-specific cue. ***C***, LTM-to-STM forward dynamics as shown in population firing rates of STM and LTM activity following LTM activation induced by a 50 ms targeted stimulus at time 0. LTM-driven activations of STM are characterized by an FF delay. Shadows indicate the SD of 100 peristimulus activations in LTM (blue) and STM (orange) with and without plasticity enabled (dashed, dark orange). Horizontal bars indicate the activation half-width (Materials and Methods). Onset is denoted by vertical dashed lines. The stimulation of LTM and the activation of plasticity is denoted underneath. ***D***, Subsampled spike raster of STM (top) and LTM (middle) during forward activation of the untrained STM by five different LTM memory patterns, triggered via specific memory cues in LTM at times marked by the vertical dashed lines. Bottom spike raster shows LTM layer 2/3 activity of one selective MC per activated pattern (colors indicate different patterns). Top spike raster shows layer 2/3 activity of one HC in STM. STM spikes are colored according to each cells dominant pattern selectivity (based on the memory pattern correlation of individual STM cell spiking during initial pattern activation, see Materials and Methods, Spike train analysis and memory activity tracking). Bottom, The five stimuli to LTM (colored boxes) and modulation of STM plasticity (black line). Extended Data Figure 3-1 shows basic network behavior in spike rasters and population firing rates under low-input feature fluctuations in membrane voltages and low-rate, asynchronous spiking activity, while Extended Data Figure 3-2 shows network activity during plasticity-modulated stimulation with 20% spatial extent, illustrating the impact of conductance delays on cortical dynamics (see Model robustness).

pattern completion by analyzing component MC activation. Whenever targeted stimuli are used, we analyze peristimulus activation traces. When activation onsets are less predictable, such as during free STM-paced maintenance, we extract activation candidates via a threshold detector trained at the 50th percentile of the cumulative distribution of the population firing rate signal.

Code accessibility

We used the NEST simulator (Gewaltig and Diesmann, 2007) version 2.2 for our simulations (RRID:SCR_002963), running on a Cray XC-40 Supercomputer of the PDC Centre for High Performance Computing. The custom-built spiking neural network implementation of the BCPNN learning rule for MPI (message passing interface) parallelized NEST is

freely available on github (https://github.com/Florian-Fiebig/BCPNN-for-NEST222-MPI) and is included in the Extended Data 1. Further, the model is also available on ModelDB (https://modeldb.yale.edu/257610).

Model robustness

Our model incorporates a plethora of biological constraints, such as estimates of the extent and distance of areas (e.g., STM patch size approximates macaque dlPFC and is 40 mm from either LTM patch), laminar cell distributions (n_{MC}^{pYR-L2}, $n_{MC}^{pYR-L3b}$,...), and hypercolumnar size. The model also abides by various electrophysiological constraints, such as plausible EPSP, IPSP sizes, estimates on laminar connection densities, laminar characterization of cortical FF/FB pathways with remote patchy connectivity, estimates on axonal conductance speeds, dendritic arbor sizes (branching factors), commonly accepted synaptic time constants for various receptor types, depression, adaptation, and builds on top of established models, such as the neuron model or the synaptic resource model. References to many of these constraints can be found throughout the Materials and Methods.

Because our model is quite complex and synthesizes many different components and processes, it is beyond the scope of this work to perform a detailed parameter sensitivity analysis. However, from our extensive simulations we conclude that it is robust and degrades gracefully. Almost all uncertain parameters can be varied ±30% without breaking WM function. The model is dramatically subsampled, and scaling up would be possible. This could be expected to further improve overall robustness. Highly related modular cortical network models have been studied extensively previously (Lundqvist et al., 2010, 2011; Tully et al., 2013, 2014; Fiebig and Lansner, 2017). For example, the model sensitivity to important short-term plasticity parameters affecting active maintenance mechanisms and intermittent gamma bursts (e.g., neural adaptation and synaptic depression time constants) were specifically explored in a single-network model (Fiebig and Lansner, 2017; see Fig. 8).

In the following, we briefly address new aspects of model sensitivity, previously unexplored, such as the parameterization of corticocortical connectivity, spatial scale (and associated conduction delays), as well as the transient modulation of Hebbian plasticity during rapid WM encoding.

In the FB pathway, a mere 0.6% connectivity is sufficient to support LTM activation in maintenance and recall. As rigorous testing (data not shown here) revealed, lower connectivity degrades WM capacity, unless we increase the total number of coactive STM cells by other means. FF connectivity can be even lower (0.015% in this model) because terminal clusters in STM are smaller and provide more information contrast (corticocortical connectivity). In both cases, our model uses very sparse connectivity, yet it could be increased or decreased if single synaptic currents were reduced/increased, respectively. Somewhat peculiarly, we also found that we needed to increase the corticocortical conductance of the backprojections (w_{FB}^{syn}) by the same factor of 1.8 (over the local conductance gain w_{gain}^{syn}) as another highly detailed multiarea model of macaque visual cortex (Schmidt et al., 2018) to achieve functional WM at the stated long-distance connection probabilities.

There are upper and lower limits on conduction delays in our model. When corticocortical conduction delays exceed 65 ms (corresponding to 130 mm in distance), STM FB can no longer activate the LTM network because bursts desynchronize before they arrive. STM and LTM could be adjacent, as we briefly mention at the end of the Results section, but there is a minimum spatial scale for each component network. The length of gamma bursts decreases if we reduce the spatial extent (and thus the connection delays between HCs) by 45%. At 20%, when the largest inter-HC delays fall to <5 ms (Extended Data Fig. 3–2), the spiking activity of activated memories collapses into a single brief burst, which degrades learning and effective information transmission both within and across networks. Networks may be much smaller, however, if this is compensated by slower axonal conductance velocities (<2 mm/ms). Furthermore, we verified that the relative temporal delay dither in Equation 10 can be varied considerably (0–30%) without noticeable effects on memory performance.

The Hebbian plasticity of the model can be modulated via the parameter κ (Eq. 7). While κ is normally 1 (κ_{normal}, a transient increase of $\kappa = \kappa_{encoding}$; Table 1), it enables rapid, one-shot encoding in STM (Fig. 3*D*). Halving or doubling $\kappa_{encoding}$ affects the overall working memory performance of the model only slightly, as measured by the number of items maintained during the delay period, or the overall rate of gamma bursts (Extended Data Fig. 4–4). It is, however, not possible to maintain normal WM operation without upregulating plasticity during encoding (leaving $\kappa_{encoding} = \kappa_{normal}$), unless additional compensatory changes are made to increase STM excitability, background excitation, or excitatory long-range connectivity. The strong correlation between working memory load and gamma-burst rate was previously discussed by Fiebig and Lansner (2017) in the context of evidence from multi-item WM recordings in macaque by Lundqvist et al. (2016).

Results

Our model implements WM function arising from the interactions of STM and LTM networks, which manifest as multi-modal memory binding phenomena. To this end, we simulate three cortical patches with significant biophysical detail: one STM and two LTM networks (LTMa, LTMb), representing PFC and parietotemporal areas, respectively (Fig. 2). The computational network model used here represents a detailed modular cortical microcircuit architecture in line with previous models (Lundqvist et al., 2006, 2011; Tully et al., 2016). Like those models, the new model can reproduce a wide range of mesoscopic and macroscopic biological manifestations of cortical memory function, including complex oscillatory dynamics and synchronization effects (Silverstein and Lansner, 2011; Lundqvist et al., 2011, 2013). The current model is built directly on a recent STM model of human word-list learning (Fiebig and Lansner, 2017). We subdivided the

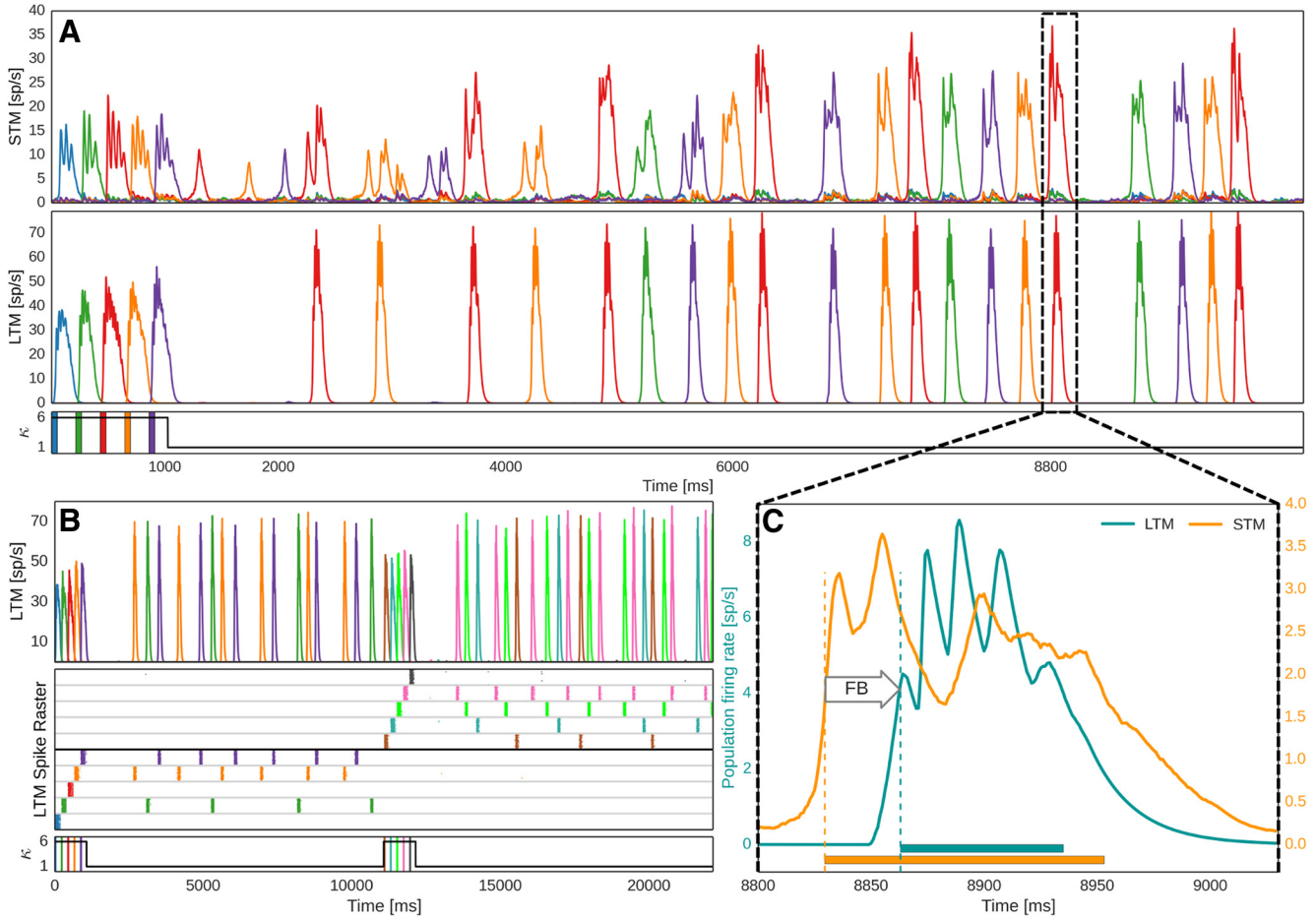

Figure 4. Encoding and feedback-driven reactivation of LTM. ***A***, Firing rates of pattern-specific subpopulations in STM and LTM during encoding and subsequent maintenance of five memories. Just as in the plasticity-modulated stimulation phase shown in Figure 2*D*, five LTM memories are cued via targeted 50 ms stimuli (shown underneath). Plasticity of STM and its backprojections is again elevated sixfold during the initial memory activation. Thereafter, a strong noise drive to STM causes spontaneous activations and plasticity induced consolidation of pattern-specific subpopulations in STM (lower plasticity, κ = 1). Backprojections from STM cell assemblies help reactivate associated LTM memories. ***B***, Updating of WM. Rapid encoding and subsequent maintenance of a second group of memories following an earlier set. The LTM spike raster shows layer 2/3 activity of one LTM HC (MCs separated by gray horizontal lines), and the population firing rate of pattern-specific subpopulations across the whole LTM network is seen above. Underneath, we denote stimuli to LTM and the modulation of plasticity, κ, in STM and its backprojections. ***C***, STM-to-LTM loop dynamics during a spontaneous reactivation event. STM-triggered activations of LTM memories are characterized by a feedback delay and a second peak in STM after LTM activations. Horizontal bars at the bottom indicate activation half-width (Materials and Methods). Onset is denoted by vertical dashed lines. Extended Data Figure 4-1 shows a more detailed spike raster of WM encoding and maintenance. Extended Data Figures 4-2 and 4-3 show spike rates and a subsampled spike rasters during WM updating and maintenance. Extended Data Figure 4-4 shows the sensitivity of delay activity to the plasticity modulation κ_v during encoding.

associative cortical layer 2/3 network of that model into layers 2, 3A, and 3B. Importantly, we also extended this model with an input layer 4 and corticocortical connectivity to LTM stores in temporal cortical regions. This large, multiarea network model synthesizes many different anatomic and electrophysiological cortical data and produces complex output dynamics. Here, we specifically focus on the dynamics of memory specific subpopulations in the interaction of STM and LTM networks.

We introduce the operation of the WM model in several steps. First, we take a brief look at background activity and active memory states in isolated cortical networks of this kind to familiarize the reader with some of its dynamical properties. Second, we describe the effect of memory activation on STM with and without plasticity. Third, we add the plastic backprojections from STM to LTM and monitor the encoding and maintenance of several memories in the resulting STM–LTM loop. We track the evolution of acquired cell assemblies with shared pattern selectivity in STM and show their important role in WM maintenance (called delay activity). We then demonstrate that the emerging WM network system is capable of flexibly updating the set of maintained memories. Finally, we simulate multimodal association and analyze its dynamical correlates. We explore temporal characteristics of network activations, the accompanying oscillatory behavior of the synthesized field potentials, cross-cortical delays as well as gamma-band coupling (coherence and phase

synchronization) between LTM networks during WM encoding, maintenance, and cue-driven associative recall of multimodal memories (LTMa–LTMb pairs of associated memories).

Background activity and activated memory

At sufficiently high background input levels, the empty network transitions from asynchronous spiking activity into a state characterized by global oscillations of the population firing rates in the alpha/beta range (Fig. 3*A*). This is largely an effect of fast feedback inhibition from local basket cells (Fig. 1), high connection density within MCs, and low latency local spike transmission (Lundqvist et al., 2010). If the network has been trained with structured input so as to encode memory (i.e., attractor states), background noise, or a specific cue (Materials and Methods) can trigger memory item reactivations accompanied by fast broadband oscillations modulated by an underlying slow oscillation in the lower theta range (~4–8 Hz; Lundqvist et al., 2011; Herman et al., 2013; Fig. 3*B*). The spiking activity of memory activations (called attractors) is short lived due to neural adaptation and synaptic depression. When unspecific background excitation is very strong, this can result in a random walk across stored memories (Lundqvist et al., 2011; Fiebig and Lansner, 2017).

LTM-to-STM forward dynamics

We now consider cued activation of several memories embedded in LTM. Each HC in LTM features selectively coding MCs for given memory patterns that activate synchronously in theta-like cycles each containing several fast oscillation bursts (Fig. 3*B*). Five different LTM memory patterns are triggered by brief cues, accompanied by an upregulation of STM plasticity (Fig. 3*D*, bottom). To indicate the spatiotemporal structure of evoked activations in STM, we also show a simultaneous subsampled STM spike raster (Fig. 3*D*, top). STM activations are sparse (~5%), but despite this, nearby cells (in the same MC) often fire together. The distributed, patchy character of the STM response to memory activations (Fig. 3*D*, top) is shaped by branching forward-projections from LTM layer 3B cells, which tend to activate cells that are close by. STM input layer four receives half of these corticocortical connections and features very high fine-scale specificity in its projections to layer 2/3 pyramidal neurons, which furthers the recruitment of local clusters with shared selectivity. STM cells initially fire less than those in LTM because the latter received a brief, but strong, activation cue and have strong recurrent connections if they code for the same embedded memory pattern. STM spikes in Figure 3*D* are colored according to the dominant memory pattern selectivity of the cells (Materials and Methods, Spike train analysis and memory activity tracking), which reveals that STM activations are mostly nonoverlapping as well. Unlike the organization of LTM with strictly nonoverlapping memory patterns, MC activity in STM is not exclusive to any given input pattern. Nevertheless, nearby STM cells often develop similar pattern selectivity. On the other hand, different stimulus patterns typically develop quite nonoverlapping STM representations. This is due to the randomness in LTM–STM connectivity, competition via basket cell feedback inhibition, and short-term dynamics, such as neural adaptation and synaptic depression. STM neurons that have recently been activated by a strong, bursting input from LTM are refractory and thus less prone to spike again for some time thereafter (τ_{rec} and τ_{I_w}; Table 1), further reducing the likelihood of creating overlapping STM representations for different patterns.

Figure 3*C* shows peristimulus population firing rates of both STM and LTM networks (the mean across 100 trials with five triggered memories each). There is a bottom-up response delay between stimulus onset at $t = 0$ and LTM activation, as well as a substantial forward delay. Oscillatory activity in STM is lower than in LTM mostly because the untrained STM lacks strong recurrent connections. It is thus less excitable, and therefore does not trigger its basket cells (the main drivers of fast oscillations in our model) as quickly as in LTM. Fast oscillations in STM and the amplitude of their theta-like envelope build up within a few seconds as new cell assemblies become stronger [Fig. 4*A* (see also Fig. 8)]. As seen in Figure 3*B*, bursts of coactivated MCs in LTM can become asynchronous during activation. Dispersed forward axonal conduction delays further decorrelate this gamma-like input to STM. Activating strong plasticity in STM ($\kappa = \kappa_{\mathrm{p}}$; Materials and Methods; Table 1) has a noticeable effect on the amplitude of stimulus-locked oscillatory STM activity after as little as 100 ms (Fig. 3*C*, STM).

Multi-item working memory

In Figure 3*D*, we have shown pattern-specific subpopulations in STM emerging from FF input. Modulated STM plasticity allows for the quick formation of rather weak STM cell assemblies from one-shot learning. When we include plastic STM backprojections, these assemblies can serve as an index for specific LTM memories and provide top-down control signals for memory maintenance and retrieval. STM backprojections with fast Hebbian plasticity can index multiple activated memories in the closed STM–LTM loop. In Figure 4*A*, we show network activity following targeted activation of five LTM memories. Under an increased unspecific noise drive ($r^{L23}_{bg-high}$; Table 2), STM cell assemblies formed during the brief plasticity-modulated stimulus phase (Fig. 3*D*) may activate spontaneously. These brief bursts of activity are initially weak and different from the theta-like cycles of repeated fast bursting seen in LTM attractor activity.

STM recurrent connections remain plastic ($\kappa = 1$) throughout the simulation, so each reactivation event further strengthens memory-specific cell assemblies in STM. As a result, there is a noticeable ramp-up in the strength of STM pattern-specific activity over the course of the delay period (Fig. 4*A*, increasing burst length and amplitude). STM backprojections are also plastic and thus acquire memory specificity from STM–LTM coactivations, especially during the initial stimulation phase. Given enough STM cell assembly firing, their sparse but potentiated backprojections can trigger associated memories in LTM. Weakly active assemblies may fail to do so. In the example of Figure 4*A*, we can see a few early STM

reactivations that are not accompanied (or quickly followed) by a corresponding LTM pattern activation (of the same color) in the first 2 s after the plasticity-modulated stimulation phase. When LTM is triggered, there is a noticeable FB delay (Fig. 4*C*), which we will address together with aforementioned FF delays in the analysis of recall dynamics during a multi-item, multimodal recall task.

Cortical FF and FB pathways between LTM and STM form a loop, so each LTM activation will again feed into STM, typically causing a second peak of activation in STM 40 ms after the first (Fig. 4*C*). The forward delay from LTM to STM, which we have seen earlier in the stimulus-driven input phase (Fig. 3*C*), is still evident here in this delayed secondary increase of the STM activation following LTM onset. The reverberating cross-cortical activation extends/sustains the memory activation and thus helps to stabilize item-specific STM cell assemblies and their specificity. This effect may be called auto-consolidation, and it is an emergent feature of the plastic STM–LTM loop in our model. It occurs on a timescale governed by the unmodulated plasticity time constant ($\kappa = \kappa_{normal}$, $\tau_p = 5$ s; Table 1). After a few seconds, the network has effectively stabilized and typically maintains a small set of three to four activated long-term memories. The closed STM–LTM loop thus constitutes a functional multi-item WM.

A crucial feature of any WM system is its flexibility, and Figure 4*B* highlights an example of rapid updating. The maintained set of activated memories can be weakened by stimulating yet another set of input memories. Generally speaking, earlier items are reliably displaced from active maintenance in our model if activation of the new items is accompanied by the same transient elevation of plasticity (k_p/k_{normal}; Table 1) used during the original encoding of the first five memories.

In line with the earlier results by Fiebig and Lansner (2017), cued activation can usually still retrieve previously maintained items. The rate of decay for memories outside the maintained set depends critically on the amount of noise in the system, which erodes the learned associations between STM and LTM neurons as well as STM cell assemblies. We note that such activity-dependent memory decay is substantially different from time-dependent decay, as shown by Mi et al. (2017).

Multimodal, multi-item working memory

Next, we explore the ability of the closed STM–LTM loop system to flexibly bind coactive pairs of long-term memories from different modalities (LTMa and LTMb, respectively). As both LTM activations trigger cells in STM via FF projections, a unique joint STM cell assembly with shared pattern selectivity is created. Forward activations include excitation and inhibition, and combine nonlinearly with each other (Materials and Methods) and with prior STM content.

Figure 5 illustrates how this new index then supports WM operations, including delay maintenance through STM-paced coactivation events and stimulus-driven associative memory pair completion. The three columns of Figure 5 illustrate the following three fundamental modes of the closed STM–LTM loop: stimulus-driven encoding, WM maintenance, and associative recall. The top three rows show sampled activity of a single trial, whereas the bottom row shows multitrial averages.

During stimulus-driven fast binding, we coactivate memories from both LTMs by brief 50 ms cues that trigger activation of the corresponding memory patterns. The average of peristimulus activations reveals 45 ± 7.3 ms LTM attractor activation delay, followed by 43 ± 7.8 ms FF delay (about half of which is explained by axonal conduction delays due to the spatial distance between LTM and STM) from the onset of the LTM activations to the onset of the input-specific STM response (Fig. 5, top left, bottom left).

During WM maintenance, a 10 s delay period, paired LTM memories reactivate together. The onset of these paired activations is a lot more variable than during cued activation with an FB delay mean of 41.5 ± 15.3 ms, mostly because the driving STM activations are of variable size and strength. Over the course of the maintenance period, the oscillatory dynamics of the LTMs changes. In particular, LFP spectral power as well as coherence between LTMs in the broad gamma band (30–80 Hz) increases ($p < 0.001$ for each of two permutation tests comparing average spectral power/coherence in the gamma band between two intervals during the delay period: 4–8 s and 8–12 s; $n = 25$ trials). To study the fast oscillatory dynamics of the LFP interactions between LTMs during the WM maintenance, mediated by STM, we follow up the coherence analysis and examine the gamma phase synchronization effect using PLV with 0.5 s sliding window (see Materials and Methods). It appears that the gamma phase coupling also increases during the second part of the WM maintenance period ($p < 0.001$ in the analogous permutation test, as described above; Fig. 6).

Following the maintenance period, we test the ability of the memory system for bimodal associative recall. To this end, we cue LTMa, again using a targeted 50 ms cue for each memory, and track the systems response across the STM–LTM loop. We compute multitrial averages of peristimulus activations during recall testing (Fig. 5, bottom right). Following cued activation of LTMa, STM responds with the related joint cell assembly activation as the input is strongly correlated to the learned inputs, as a result of the simultaneous activation with LTMb earlier on. Similar to the mnemonic function of an index, the completed STM pattern then triggers the associated memory in LTMb through backprojections. STM activation now extends far beyond the transient activity of LTMa because STM recurrent connectivity and the STM–LTMb recurrence re-excites it. The temporal overlap between associated LTMa and LTMb memory activations peaks at ~125 ms after the initial stimulus to LTMa.

Network power spectra and the nonassociative control case

Figure 7 (top) shows multitrial peristimulus/periactivation activity traces for a control task, where learned and maintained LTMa items are not associated with concurrent LTMb activations. LTMa items are still encoded in STM, maintained over the delay, and recalled by specific

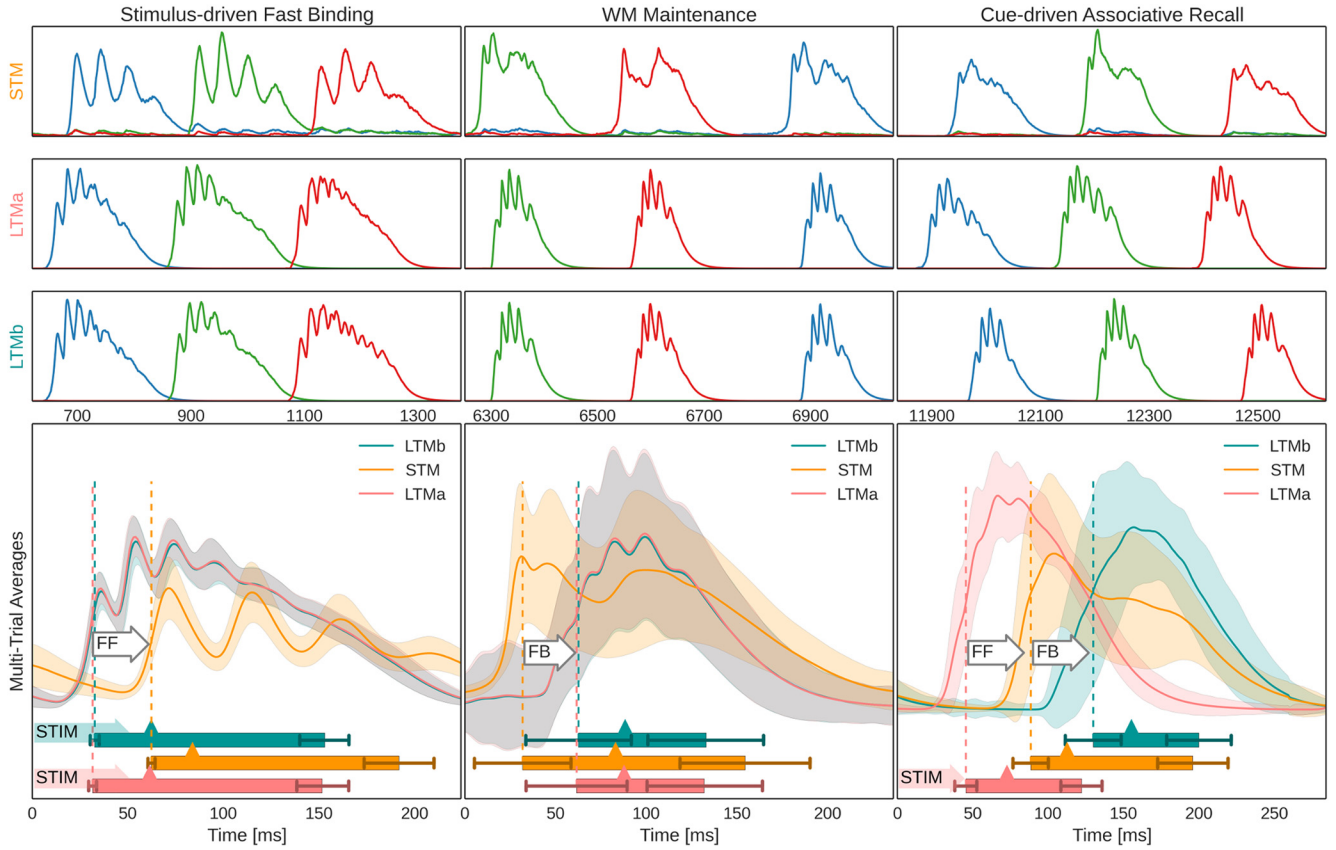

Figure 5. Population firing rates of networks and memory-specific subpopulations during three different modes of network activity. Top half, Exemplary activation of three memories (blue, green, and red, respectively) in STM (first row), LTMa (second row), and LTMb (third row) during the following three different modes of network activity: the initial association of pairs of LTM memory activations in STM (left column), WM maintenance through spontaneous STM-paced activations of bound LTM memory pairs (middle column), and cue-driven associative recall of previously paired stimuli (right column). Bottom half, Multitrial peristimulus activity traces from the three cortical patches across 100 trials (495 traces, as each trial features five activated and maintained LTM memory pairs and very few failures of paired activation). Shaded areas indicate an SD from the underlying traces. Vertical dashed lines denote mean onset of activity for each network, as determined by activation half-width (Materials and Methods), also denoted by a box underneath the traces. Error bars indicate an SD from activation onset and offset. Mean peak activation is denoted by a triangle on the box, and shaded arrows to the left of the box denote targeted pattern stimulation of a network at time 0. As there are no external cues during WM maintenance (i.e., the delay period), we use detected STM activation onset to align firing rate traces of 5168 STM-paced LTM reactivations across trials and reactivation events for averaging. White arrows annotate FF and FB delay, as defined by respective network onsets. Extended Data Figure 5-1 further illustrates the subsampled spiking activity in the three networks, during the multimodal LTM binding task.

cues, but LTMb now remains silent throughout the maintenance period (Fig. 7, top left) and, as expected, does not show any evidence of memory activation following LTMa-specific cues during recall testing (Fig. 7, top right, Fig. 8, LFP signal). The logarithmic power spectra (Fig. 7, bottom) show a noticeable difference between the normal associative and the nonassociative control trials. The latter displays a significant drop in LTMb power across the board, particularly during the maintenance period. This can be explained by the overall lower number of memory reactivations in STM during the nonassociative control task (2.58 $\pm$ 0.28 vs 1.62 $\pm$ 0.47 reactivations/s).

Top-down and bottom-up delays

We collected distributions of FF and FB delays during associative recall (Fig. 9). To facilitate a more immediate comparison with biological timing data, we also computed the bottom-up and top-down response latencies of the model in analogy to Tomita et al. (1999). Their study explicitly tested widely held beliefs about the executive control of PFC over ITC in memory retrieval. To this end, they identified and recorded neurons in ITC of monkeys trained to memorize several visual stimulus–stimulus associations. They used a posterior-split brain paradigm to cleanly disassociate the timing of the bottom-up (contralateral stimuli) and top-down (ipsilateral stimuli) responses in 43 neurons that were significantly stimulus selective in both conditions. They observed that the latency of the top-down response (178 ms) was longer than that of the bottom-up response (73 ms).

Our simulation is analogous to this experimental setup with respect to some key features, such as the spatial extent of memory areas (STM/dlPFC, ~289 mm^2) and

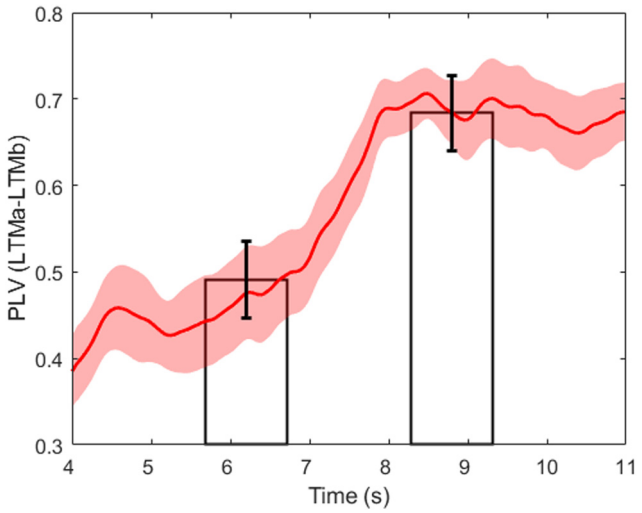

Figure 6. Gamma band PLV between LTMa and LTMb during WM maintenance. PLV is estimated using sliding window of size 0.5 s (the period between 4 and 12 s is shown). Two bars demonstrate the average gamma-band PLV over the first (4–8 s) and the second part (8–12 s) of the WM maintenance period. Shaded area and error bars correspond to the SEM calculated over $n = 25$ trials.

interarea distances (40 mm cortical distance between PFC and ITC). These measures heavily influence the resulting connection delays and the time needed for information integration. In analogy to the posterior-split brain experiment, the LTMa and LTMb in our model are unconnected. However, we now have to consider them as ipsilateral and contralateral visual areas in ITC. The display of a cue in one hemifield in the experiment then corresponds to the LTMa-sided stimulation of an associated memory pair in the model. This arrangement forces any LTM interaction through STM (representing PFC) and allows us to treat the cued LTMa memory activation as a bottom-up response, whereas the much later activation of the associated LTMb representation is related to the top-down response in the experimental study. Figure 9 shows the distribution of these latencies in our simulations, where we also marked the mean latencies measured by Tomita et al. (1999). The mean of our bottom-up delay (72.9 ms) matches the experimental data (73 ms), whereas the mean of the broader top-down latency distribution (155.2 ms) is a bit lower than that in the monkey study (178 ms). Only 31 % (48 ms) of the top down delay (155.2 ms) was explained by the spatial distance between networks, as verified by a simulated model with 0 mm distance between networks.

Discussion

In this work, we have proposed and studied a novel theory for WM that rests on the dynamic interactions between STM and LTM stores enabled by fast synaptic plasticity. In particular, it hypothesizes that activity in parietotemporal LTM stores targeting PFC via fixed or slowly plastic and patchy synaptic connections triggers an activity pattern in PFC, which then gets rapidly encoded by means of fast Hebbian plasticity to form a cell assembly. Equally plastic backprojections from PFC to the LTM stores are enhanced as well, thereby associating the formed PFC "index" specifically with the active LTM cell assemblies. This rapidly but temporarily enhanced connectivity produces a functional WM superassembly (a distributed constellation of cell assemblies) capable of encoding and maintaining multiple individual LTM items (i.e., bringing these LTM representations "online") and forming novel associations within and between several connected LTM areas and modalities. The PFC cell assemblies themselves do not encode much information but act as indices of LTM stores, which contain additional information that is also more permanent. The underlying highly plastic connectivity and thereby the WM itself is flexibly remodeled and updated as new incoming activity gradually overwrites previous WM content. How quickly working memory is established after the initial encoding period, critically depends on plasticity modulation, network size, and overall activity. Our model does not address other important aspect of WM (e.g., the task relevance filtering and attentional gating involving upstream subcortical structures like basal ganglia and amygdala; O'Reilly and Frank, 2006; McNab and Klingberg, 2008).

We have studied the functional and dynamical implications of this theory by implementing and evaluating a special case of a biologically plausible large-scale spiking neural network model representing PFC reciprocally connected with two LTM areas (e.g., visual and auditory) in temporal cortex. We have shown how a number of single LTM items can be encoded and maintained online, and how pairs of simultaneously activated items can become jointly indexed and associated. Activating one pair member now also activates the other one indirectly via PFC with a short latency. We have further demonstrated that this kind of WM can readily be updated, such that as new items are encoded, old items are fading away, whereby the active WM content is replaced. Notably, unlike in our model, in a biological brain many long-range connections exist between LTM areas, and they will significantly influence the sequence of recalled items.

Recall dynamics in the presented model are in most respects identical to a previous cortical associative memory model (Lansner, 2009) and also to that of single-item persistent activity WM models (Camperi and Wang, 1998). Any activated memory item, whether randomly or specifically triggered, is subject to known and previously well characterized associative memory dynamics, such as pattern completion, rivalry, bursty reactivation dynamics, and oscillations in different frequency bands (Lundqvist et al., 2010, 2013; Silverstein and Lansner, 2011; Herman et al., 2013). Moreover, sequential learning and recall could readily be incorporated (Tully et al., 2016). This could, for example, support encoding of sequences of items in WM rather than a set of unrelated items, resulting in reactivation dynamics reminiscent of, for instance, the phonological loop (Baddeley et al., 1998; Burgess and Hitch, 2006).

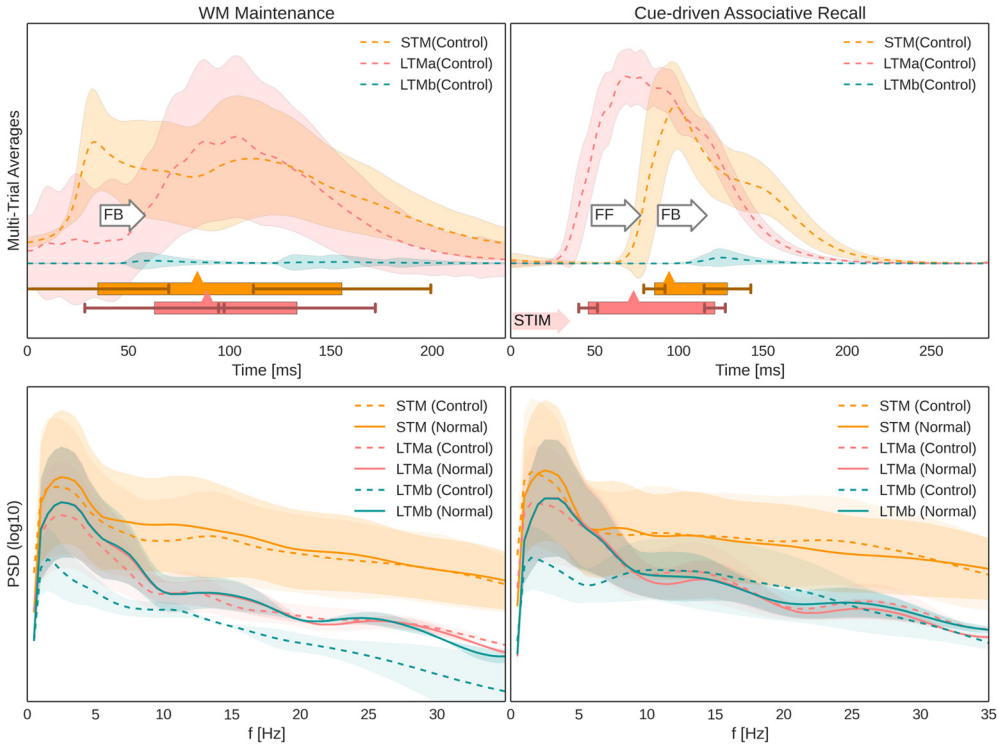

Figure 7. Nonassociative control case and power spectral analysis. Top half, Multitrial peristimulus activity traces from the three cortical patches across 25 trials following WM-encoded LTMa activations as before, but without associated LTMb memory activations. Shaded areas indicate a SD from the underlying traces. Activation half-widths (Materials and Methods) denoted by a box underneath the traces. Error bars indicate an SD from activation onset and offset. Mean peak activation is denoted by a triangle on the box, and shaded arrows to the left of the box denote targeted pattern stimulation of LTMa at time 0. As there are no external cues during WM maintenance (called the delay period), we use detected STM activation onset to align firing rate traces of 406 STM-paced LTMa reactivations across trials and reactivation events for averaging. There is no evidence of associated LTMb activations in the control case (only small increases in spike rate variability). White arrows annotate FF and FB delay, as defined by respective network onsets. Bottom half, Power spectral density of synthesized LFPs estimated over the maintenance (left) and recall (right) periods for STM and both LTMs in two cases: with (solid lines) and without (dashed line; control case) associated LTMb memory activations. Please note the log scale. Shaded areas correspond to the SD of the mean PSD over 25 trials. The decrease in theta- and gamma-band power observed during the maintenance (left) and recall (right) periods in the LTMb in the control case is due to lack of memory pattern reactivations in LTMb as they are not associated with LTMa via STM.

Cortical indexing theory for WM

Our model draws inspiration from the hippocampal indexing theory (Teyler and DiScenna, 1986), originally proposed to account for the role of hippocampus in storing episodic memories (Teyler and Rudy, 2007). While there are key similarities, there are also a number of important differences. Similar to the hippocampus, PFC is well connected with association cortices to directly or indirectly influence activity across most of cortex (Pandya and Yeterian, 1991; Pandya and Barnes, 2019). Unlike the hippocampal indexing theory, which posits that such influence is not seen until the eventual recall, we propose and demonstrate in simulation that the creation of the PFC index has immediate effects on neocortical activity patterns, manifested as WM delay activity across widely distributed cortical areas (Tomita et al., 1999; Sreenivasan and D'Esposito, 2019).

In line with Teyler and Discenna (1986), we propose that the rapid encoding necessitated by indexing is enabled by transient dopaminergic modulation of plasticity among recurrently connected neurons and their FB projections onto cortex. As suggested for hippocampus, PFC does not store the sensory content of WM itself, but rather an index to task-relevant information carried by areas lower in the cortical hierarchy. As PFC integrates information across cortex (Miller and Cohen, 2001), the index becomes part of a temporary WM superassembly. Hippocampal indexing is largely seen as a process preceding cortical long-term consolidation, whereby associations between indexed areas eventually become independent of hippocampus. Our model makes no such claim for the role of PFC. On the contrary, WM function rests on the fluidly changing selectivity of PFC neurons, which would be hampered by strong LTP and slow processes of consolidation. Yet an intriguing possibility suggested by our quantitative model is that a hippocampal indexing network with a longer plasticity time constant operating analogously to our PFC model could support such a consolidation process by reinstating activity in cortical LTM areas.

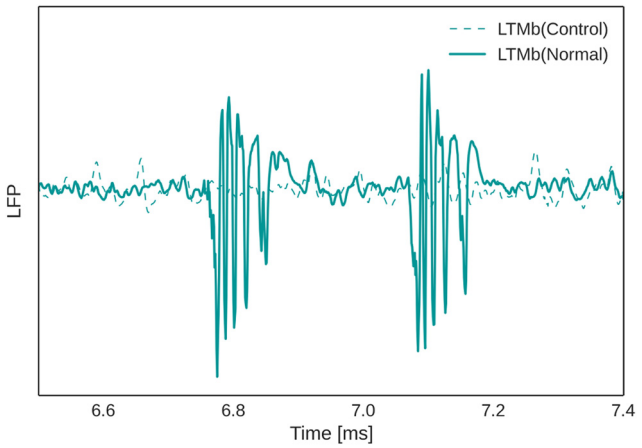

Figure 8. Exemplary recording of the LFP signal in LTMb following two cued activations of LTMa after learning and maintenance of associative LTMa–LTMb memory pairs (normal) or nonassociative LTMa memories without concurrent LTMb activation (control). While the LFP signal shows clear activation of associated LTMb items, LTMa-specific cues do not elicit memory activations in LTMb in the control case.

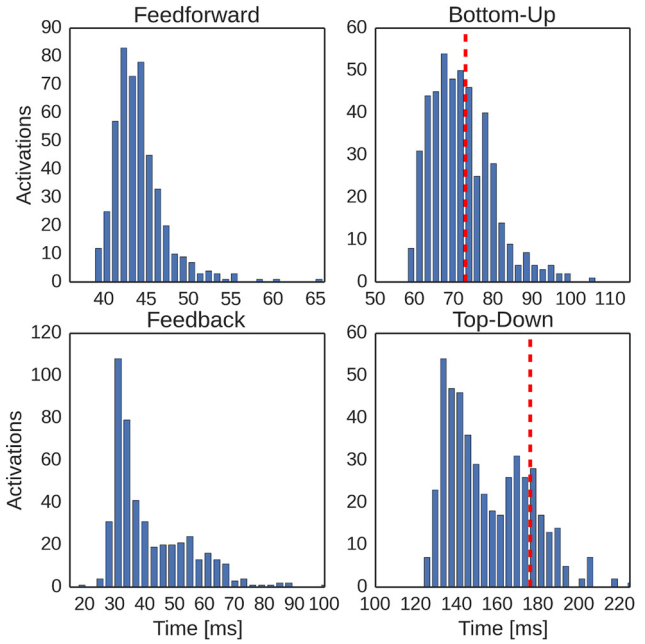

Figure 9. Comparison of key activation delays during associative recall in model and experiment following a cue to LTMa. Top left, Feedforward delay distribution in the model, as defined by the temporal delay between LTMa onset and STM onset (Fig. 4, bottom right). Top right, Bottom-up delay distribution in the model, as defined by the temporal delay between stimulation onset and LTMa peak activation. The red line denotes the mean bottom-up delay, as measured by Tomita et al. (1999). Bottom left, Feedback delay distribution in the model, as defined by the temporal delay between STM onset and LTMb onset (Fig. 4, bottom-right, measured by half-width). Bottom right, Top-down delay distribution in the model, as defined by the temporal delay between stimulation onset and LTMb peak activation. The red line denotes the mean bottom-up delay, as measured by Tomita et al. (1999). Model delays were averaged over 100 trials with five paired stimuli each.

The case for Hebbian plasticity

The underlying mechanism of our model is fast Hebbian plasticity, not only in the intrinsic PFC connectivity, but also in the projections from PFC to LTM stores. The former has some experimental support (Volianskis and Jensen, 2003; Erickson et al., 2010; Park et al., 2014; Volianskis et al., 2015; Pradier et al., 2018), whereas the latter remains a prediction of the model. D1R activation by DA is strongly implicated in reward learning and synaptic plasticity regulation in the basal ganglia (Wickens, 2009). In analogy, we propose that D1R activation is critically involved in the synaptic plasticity intrinsic to PFC and its projections to LTM stores, which would also explain the comparatively dense DA innervation of PFC and the prominent WM effects of PFC DA level manipulation (Goto et al., 2010; Arnsten and Jin, 2014). In the model presented here, the parameter κ represents the level of DA–D1R activation, which in turn regulates synaptic plasticity. We typically increase κ temporarily (Table 1) in conjunction with stimulation of LTM and WM encoding, in a form of attentional gating. Excessive modulation limits WM capacity to one to two items, while less modulation diminishes the strength of cell assemblies beyond what is necessary for reactivation and LTM maintenance.

When the synaptic plasticity WM hypothesis was first presented and evaluated, it was based on synaptic facilitation (Mongillo et al., 2008; Lundqvist et al., 2011). However, such non-Hebbian plasticity is only capable of less specific forms of memory. Activating a cell assembly comprising a subset of neurons in an untrained STM network featuring such plasticity would merely facilitate all outgoing synapses from active neurons. Likewise, an enhanced elevated resting potential resulting from intrinsic plasticity would make the targeted neurons more excitable. In neither case would there be any coordination of activity specifically within the stimulated cell assembly. Thus, if superimposed on an existing LTM, such forms of plasticity may well contribute to WM, but they are by themselves not capable of supporting encoding of novel memory items or the multimodal association of already existing ones. In contrast, previous work by Fiebig and Lansner (2017) showed that fast Hebbian plasticity similar to short-term potentiation (Erickson et al., 2010) allows effective one-shot encoding of novel STM items. In the extended model proposed here, PFC can additionally bind and bring online existing but previously unassociated LTM items across multiple modalities by means of the same kind of plasticity in backprojections from PFC to parietotemporal LTM stores.

On a side note, this implementation of fast Hebbian plasticity reproduces a remarkable aspect of short-term potentiation or labile LTP: it decays in an activity-dependent manner rather than with time (Volianskis and Jensen, 2003; Volianskis et al., 2015; Pradier et al., 2018). Although we used the BCPNN learning rule to reproduce these effects, we expect that other Hebbian learning rules

allowing for neuromodulated fast synaptic plasticity could give comparable results.

Experimental support and testable predictions

Our model has been built from available relevant microscopic data on neural and synaptic components as well as the modular structure and connectivity of selected cortical areas in macaque monkey. Its sparse corticocortical long-range connectivity is compatible with neuroanatomical data and can add specific predictions of the nature of this connectivity. The network so designed generates a well organized macroscopic dynamic working memory function, which can be interpreted in terms of manifest behavior and validated against cognitive experiments and data. Our model provides a powerful tool to investigate and examine the link between microscopic and macroscopic level processes and data. It suggests novel mechanistic hypotheses and inspiration for planning and performing experiments that can develop further the model, or potentially falsify it.

Unfortunately, the detailed neural processes and dynamics of our new model are not easily accessible experimentally as they are intrinsically expressed at multiple scales (e.g., mesoscopic field potentials and population spiking at macroscopic spatial scales). In consequence, it is difficult to find direct and quantitative results to validate the model. To our knowledge, no other WM model of comparable detail has been reported. On the one hand, some recent models that explain WM activity through the long-range interactions of STM and LTM systems (Bouchacourt and Buschman, 2019) lack defensible constraints on the density of the long-range projections involved. On the other hand, a more complete cortical model by Schmidt et al. (2018), accounting for available corticocortical connectivity data from layer-specific retrograde tracing experiments (Markov et al., 2014), was not concerned with any concrete aspects of cognitive function, such as working memory. We specifically tested extremely sparse and plastic corticocortical connectivity and demonstrated its effectiveness in indexing working memory.

In analyzing our resulting bottom-up and top-down delays, we drew an analogy to a split-brain experiment (Tomita et al., 1999) because of its clean experimental design (even controlling for subcortical pathways) and found similar temporal dynamics in our highly subsampled cortical model. The timing of interarea signals also constitutes a testable prediction for multimodal memory experiments. Furthermore, reviews of intracranial as well as electroencephalography (EEG) recordings conclude that theta band oscillations play an important role in long-range communication during successful memory retrieval (Sauseng et al., 2004; Johnson and Knight, 2015). With respect to theta band oscillations in our model, we have shown that STM leads the LTM networks during maintenance, engages bidirectionally during recall (due to the STM–LTM loop), and lags during stimulus-driven encoding and LTM activation, reflecting experimental observations (Anderson et al., 2010). These effects are explained by our model architecture, which imposes delays due to the spatial extent of networks and their distances from each other. Fast oscillations in the broad gamma band, often nested in the theta cycle, are strongly linked to local processing and activated memory items in our model, also matching experimental findings (Canolty and Knight, 2010; Johnson and Knight, 2015). Local frequency coupling is abundant with significant phase–amplitude coupling (Fig. 3*B*), and was well characterized in related models (Herman et al., 2013).

The most critical requirement and thus prediction of our theory and model is the presence of fast Hebbian plasticity in the PFC backprojections to parietotemporal memory areas. Without such plasticity, our model cannot explain the necessary STM–LTM binding. This plasticity is likely to be subject to neuromodulatory control, presumably with DA and D1R activation involvement. Since short-term potentiation decays with activity, a high noise level could be an issue since it could shorten WM duration (see The case for Hebbian plasticity). The evaluation of this requirement is hampered by lack of experimental evidence and characterization of the synaptic plasticity in corticocortical projections.

One of the neurodynamical manifestations of the fast associative plasticity in the PFC backprojections is a functional coupling between LTM stores. Importantly, this long-range coupling in our model is mediated by the PFC network alone, as manifested during the delay period free of any external cues, and is reflected in the synchronization of fast gamma oscillations. Although the predominant view has been that gamma is restricted to short distances, there is growing evidence for cortical long-distance gamma phase synchrony between task-relevant areas as a correlate of cognitive processes (Tallon-Baudry et al., 1998; Doesburg et al., 2008) including WM (Palva et al., 2010). In this regard, our model generates even a more specific prediction about the notable temporal enhancement of gamma phase coupling over the delay period, which could be tested with macroscopic human brain recordings (e.g., EEG or MEG), provided that a WM task involves a sufficiently long delay period.

Finally, our model suggests the occurrence of a double peak of frontal network activation in executive control of multimodal LTM association (Figs. 4*C*, 5, STM population activity during WM maintenance). The first one originates from the top-down control signal itself, and the second one is a result of corticocortical reentry and a successful activation of one or more associated items in LTM. As such, the second peak should also be correlated with successful memory maintenance or associative recall.

Possible role for fast Hebbian plasticity in variable binding

The “binding problem” is a classical and extensively studied problem in perceptual and cognitive neuroscience (Zimmer et al., 2012). Binding occurs in different forms and at different levels, from lower perceptual to higher cognitive processes (Reynolds and Desimone, 1999; Zimmer et al., 2006; Zmigrod et al., 2014).

Variable binding is a cognitive kind of neural binding in the form of a temporary variable–value association of items previously not connected by earlier experience and

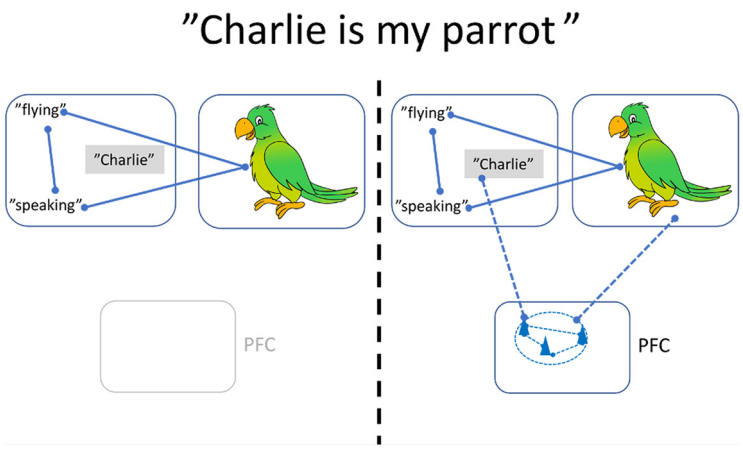

Figure 10. Variable value binding via index in PFC. Initially, the multimodal representation of “parrot” exists in LTM comprising symbolic and subsymbolic components side by side with “Charlie” as a representation of a proper name. It is hypothesized here that when someone states that “Charlie is my parrot,” the name “Charlie” is temporarily and reciprocally bound to the parrot representation via PFC, mediated by fast Hebbian plasticity. Pattern completion effect now allows “Charlie” to trigger the entire assembly and, analogously, makes “flying” or the sight of a given parrot trigger “Charlie.” If important enough or repeated a couple of times this association could consolidate in LTM.

learning (Cer and O’Reily, 2012; Garnelo and Shanahan, 2019). A simple special case is the association of a mathematical variable and its value “The value of x is 2” (i.e., $x = 2$) or of an object and its proper name as in “Charlie is my parrot” (Fig. 10). This and other more advanced forms of neural binding are assumed to underlie complex functions in human cognition including logical reasoning and planning (Pinkas et al., 2012), but have been a challenge to explain by neural network models of the brain (van der Velde and de Kamps, 2015; Legenstein et al., 2016). Work is in progress to uncover how such variable binding mechanisms can be used in neuro-inspired models of more advanced human logical reasoning (Pinkas et al., 2013).

Based on our WM model, we propose that a PFC/STM index mediated by fast Hebbian plasticity provides a neural mechanism that could explain such variable binding. The joint index formed in PFC during presentation of a name–value pair serves to temporarily bind the corresponding representations. The value could be multimodal and include symbolic as well as subsymbolic components. Turning to Figure 5 above, imagine that one of the LTMa patterns represents the image of a parrot and one pattern in LTMb represents the proper name “Charlie.” If someone says “Charlie is my parrot,” these previously not associated items are rapidly bound via a joint PFC index. While this short-term connectivity remains, the name “Charlie” will trigger the internal object representation of a parrot, and seeing a parrot will trigger the name “Charlie” with the dynamics shown in the right-most panels of Figure 5 Flexible updating of the PFC index (Fig. 4) will avoid confusion even if in the next moment my neighbor shouts “Charlie” to call his dog, also named Charlie. If important enough or repeated a number of times, this association could further consolidate in LTM.

Conclusions

We have formulated an indexing theory for cortical WM and tested it by means of computer simulations, which demonstrated the versatile WM properties of a large-scale spiking neural network model implementing key aspects of the theory. Our model provides a new mechanistic understanding of WM with distributed cortical correlates and variable binding phenomena, which connects microscopic neural processes with macroscopic observations and cognitive functions in a way that only computational models can do. While we designed and constrained this model based on macaque data, the theory itself is quite general, and we expect our findings to apply also to mammals, including humans, commensurate with changes in key model parameters (e.g., cortical distances, axonal conductance speeds). Many aspects of WM function remain to be tested and incorporated (e.g., its close interactions with basal ganglia; O’Reilly and Frank, 2006).

WM dysfunction has an outsized impact on mental health, intelligence, and quality of life. Progress in mechanistic understanding of its function and dysfunction is therefore very important for society. We hope that our theoretical and computational work provides inspiration for experimentalists to scrutinize the theory and model, especially with respect to neuromodulated fast Hebbian synaptic plasticity and large-scale network architecture and dynamics. Only in this way can we get closer to a more solid understanding and theory of WM, and position future computational research appropriately even in the clinical and pharmaceutical realm.

References

Anderson KL, Rajagovindan R, Ghacibeh GA, Meador KJ, Ding M (2010) Theta oscillations mediate interaction between prefrontal cortex and medial temporal lobe in human memory. Cereb Cortex 20:1604–1612.

Arnsten AFT, Jin LE (2014) Molecular influences on working memory circuits in dorsolateral prefrontal cortex. Prog Mol Biol Transl Sci 122:211–231.

Baddeley A, Gathercole S, Papagno C (1998) The Phonological Loop as a Language Learning Device. Psychol Rev 105:158–173.

Bouchacourt F, Buschman TJ (2019) A flexible model of working memory. Neuron 103:147–160.e8.
Brette R, Gerstner W (2005) Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. J Neurophysiol 94:3637–3642.
Burgess N, Hitch GJ (2006) A revised model of short-term memory and long-term learning of verbal sequences. J Mem Lang 55:627–652.
Caminiti R, Carducci F, Piervincenzi C, Battaglia-Mayer A, Confalone G, Visco-Comandini F, Pantano P, Innocenti GM (2013) Diameter, length, speed, and conduction delay of callosal axons in macaque monkeys and humans: comparing data from histology and magnetic resonance imaging diffusion tractography. J Neurosci 33:14501–14511.
Camperi M, Wang XJ (1998) A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. J Comput Neurosci 5:383–405.
Canolty RT, Knight RT (2010) The functional role of cross-frequency coupling. Trends Cogn Sci 14:506–515.
Carter GC (1987) Coherence and time delay estimation. Proc IEEE 75:236–255.
Cer DM, O'Reily RC (2012) Neural mechanisms of binding in the hippocampus and neocortex: insights from computational models. In: Handbook of binding and memory: perspectives from cognitive neuroscience (Zimmer HD, Mecklinger A, Lindenberger U, eds), New York: Oxford UP.
Chrysanthidis N, Fiebig F, Lansner A (2019) Introducing double bouquet cells into a modular cortical associative memory model. J Comput Neurosci 47:223–230.
Compte A, Brunel N, Goldman-Rakic PS, Wang XJ (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb Cortex 10:910–923.
Crochet S, Petersen CCH (2009) Cortical dynamics by layers. Neuron 64:298–300.
DeFelipe J, Ballesteros-Yáñez I, Inda MC, Muñoz A (2006) Double-bouquet cells in the monkey and human cerebral cortex with special reference to areas 17 and 18. Prog Brain Res 154:15–32.
D'Esposito M, Postle BR (2015) The cognitive neuroscience of working memory. Annu Rev Psychol 66:115–142.
Doesburg SM, Roggeveen AB, Kitajo K, Ward LM (2008) Large-scale gamma-band phase synchronization and selective attention. Cereb Cortex 18:386–396.
Douglas RJ, Martin K (2004) Neuronal circuits of the neocortex. Annu Rev Neurosci 27:419–451.
Erickson MA, Maramara LA, Lisman J (2010) A single brief burst induces glur1-dependent associative short-term potentiation: a potential mechanism for short-term memory. J Cogn Neurosci 22:2530–2540.
Eriksson J, Vogel EK, Lansner A, Bergström F, Nyberg L (2015) Neurocognitive architecture of working memory. Neuron 88:33–46.
Fiebig F, Lansner A (2014) Memory consolidation from seconds to weeks: a three-stage neural network model with autonomous reinstatement dynamics. Front Comput Neurosci 8:1–17.
Fiebig F, Lansner A (2017) A spiking working memory model based on Hebbian short-term potentiation. J Neurosci 37:83–96.
Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331–349.
Fuster JM (2009) Cortex and memory: emergence of a new paradigm. J Cogn Neurosci 21:2047–2072.
Garnelo M, Shanahan M (2019) Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. Curr Opin Behav Sci 29:17–23.
Gewaltig M-O, Diesmann M (2007) NEST (NEural Simulation Tool). Scholarpedia 2:1430.
Girard P, Hupé JM, Bullier J (2001) Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. J Neurophysiol 85:1328–1331.
Goldman-Rakic PS (1995) Cellular basis of working memory review. Neuron 14:477–485.
Goto Y, Yang CR, Otani S (2010) Functional and dysfunctional synaptic plasticity in prefrontal cortex: roles in psychiatric disorders. Biol Psychiatry 67:199–207.
Herman PA, Lundqvist M, Lansner A (2013) Nested theta to gamma oscillations and precise spatiotemporal firing during memory retrieval in a simulated attractor network. Brain Res 1536:68–87.
Hoffman DA, Magee JC, Colbert CM, Johnston D (1997) K+ channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons. Nature 387:869–875.
Houzel JC, Milleret C, Innocenti G (1994) Morphology of callosal axons interconnecting areas 17 and 18 of the cat. Eur J Neurosci 6:898–917.
Johnson EL, Knight RT (2015) Intracranial recordings and human memory. Curr Opin Neurobiol 31:18–25.
Kapfer C, Glickfeld LL, Atallah BV, Scanziani M (2007) Supralinear increase of recurrent inhibition during sparse activity in the somatosensory cortex. Nat Neurosci 10:743–753.
Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 8:194–208.
Lansner A (2009) Associative memory models: from the cell-assembly theory to biophysically detailed cortex simulations. Trends Neurosci 32:178–186.
Lansner A, Marklund P, Sikström S, Nilsson L (2013) Reactivation in working memory: an attractor network model of free recall. PLoS One 8:e73776.
Legenstein R, Papadimitriou CH, Vempala S, Maass W (2016) Variable binding through assemblies in spiking neural networks. Paper presented at 30th Annual Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, December.
Lindén H, Pettersen KH, Einevoll GT (2010) Intrinsic dendritic filtering gives low-pass power spectra of local field potentials. J Comput Neurosci 29:423–444.
Logothetis NK (2003) The underpinnings of the BOLD functional magnetic resonance imaging signal. J Neurosci 23:3963–3971.
Lundqvist M, Rehn M, Djurfeldt M, Lansner A (2006) Attractor dynamics in a modular network model of neocortex. Network 17:253–276.
Lundqvist M, Compte A, Lansner A (2010) Bistable, irregular firing and population oscillations in a modular attractor memory network. PLoS Comput Biol 6:e1000803.
Lundqvist M, Herman P, Lansner A (2011) Theta and gamma power increases and alpha/beta power decreases with memory load in an attractor network model. J Cogn Neurosci 23:3008–3020.
Lundqvist M, Herman P, Lansner A (2013) Effect of prestimulus alpha power, phase, and synchronization on stimulus detection rates in a biophysical attractor network model. J Neurosci 33:11817–11824.
Lundqvist M, Rose J, Herman P, Brincat SLL, Buschman TJJ, Miller EKK (2016) Gamma and beta bursts underlie working memory. Neuron 90:152–164.
Markov NT, Vezoli J, Chameau P, Falchier A, Quilodran R, Huissoud C, Lamy C, Misery P, Giroud P, Ullman S, Barone P, Dehay C, Knoblauch K, Kennedy H (2014) Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. J Comp Neur 522:225–259.
McNab F, Klingberg T (2008) Prefrontal cortex and basal ganglia control access to working memory. Nat Neurosci 11:103–107.
Mi Y, Katkov M, Tsodyks M (2017) Synaptic correlates of working memory capacity. Neuron 93:323–330.
Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.
Miyashita Y (1993) Inferior temporal cortex: where visual perception meets memory. Annu Rev Neurosci 16:245–263.
Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. Science 319:1543–1546.
Mountcastle VB (1997) The columnar organization of the neocortex. Brain 120:701–722.

O'Reilly RC, Frank MJ (2006) Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. Neural Comput 18:283–328.
Okun M, Naim A, Lampl I (2010) The subthreshold relation between cortical local field potential and neuronal firing unveiled by intracellular recordings in awake rats. J Neurosci 30:4440–4448.
Palva JM, Palva S, Kaila K (2005) Phase synchrony among neuronal oscillations in the human cortex. J Neurosci 25:3962–3972.
Palva JM, Monto S, Kulashekhar S, Palva S (2010) Neuronal synchrony reveals working memory networks and predicts individual memory capacity. Proc Natl Acad Sci U S A 107:7580–7585.
Pandya DN, Yeterian EH (1991) Prefrontal cortex in relation to other cortical areas in rhesus monkey: architecture and connections. Prog Brain Res 85:63–94.
Pandya DN, Barnes CL (2019) Architecture and connections of the frontal lobe. In: The frontal lobes revisited, pp 41–72. Hove, UK: Psychology Press.
Parisien C, Anderson CH, Eliasmith C (2008) Solving the problem of negative synaptic weights in cortical models. Neural Comput 20:1473–1494.
Park P, Volianskis A, Sanderson TM, Bortolotto ZA, Jane DE, Zhuo M, Kaang B-K, Collingridge GL (2014) NMDA receptor-dependent long-term potentiation comprises a family of temporally overlapping forms of synaptic plasticity that are induced by different patterns of stimulation. Philos Trans R Soc Lond B Biol Sci 369:20130131.
Petersson ME, Yoshida M, Fransén EA (2011) Low-frequency summation of synaptically activated transient receptor potential channel-mediated depolarizations. Eur J Neurosci 34:578–593.
Petrides M, Pandya DN (1999) Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. Eur J Neurosci 11:1011–1036.
Pinkas G, Lima P, Cohen S (2012) A dynamic binding mechanism for retrieving and unifying complex predicate-logic knowledge: In: Artificial neural networks and machine learning–ICANN 2012: 22nd International Conference on Artificial Neural Networks, Lausanne, Switzerland, September 11–14, 2012, Proceedings, Part I (Villa A, Wlodzislaw D, Erdi P, Masulli F, Palm G, eds), pp 482–490. Berlin Heidelberg: Springer-Verlag.
Pinkas G, Lima P, Cohen S (2013) Representing, binding, retrieving and unifying relational knowledge using pools of neural binders. In: Biologically inspired cognitive architectures (Samsonovich AV, Mason G, Editor-in-Chief), pp 87–95. Amsterdam: Elsevier.
Potjans TC, Diesmann M (2014) The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. Cereb Cortex 24:785–806.
Pradier B, Lanning K, Taljan KT, Feuille CJ, Nagy MA, Kauer JA (2018) Persistent but labile synaptic plasticity at excitatory synapses. J Neurosci 38:5750–5758.
Ren M, Yoshimura Y, Takada N, Horibe S, Komatsu Y (2007) Specialized inhibitory synaptic actions between nearby neocortical pyramidal neurons. Science 316:758–761.
Reynolds JH, Desimone R (1999) The role of neural mechanisms of attention in solving the binding problem. Neuron 24:19–29, 111–125.
Sakata S, Harris KD (2009) Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. Neuron 64:404–418.
Sandberg A, Lansner A, Petersson KM, Ekeberg O (2002) A Bayesian attractor network with incremental learning. Network 13:179–194.
Sauseng P, Klimesch W, Doppelmayr M, Hanslmayr S, Schabus M, Gruber WR (2004) Theta coupling in the human electroencephalogram during a working memory task. Neurosci Lett 354:123–126.
Schmidt M, Bakker R, Hilgetag CC, Diesmann M, van Albada SJ (2018) Multi-scale account of the network structure of macaque visual cortex. Brain Struct Funct 223:1409–1435.
Silberberg G, Markram H (2007) Disynaptic inhibition between neocortical pyramidal cells mediated by Martinotti cells. Neuron 53:735–746.
Silverstein DN, Lansner A (2011) Is attentional blink a byproduct of neocortical attractors? Front Comput Neurosci 5:13.
Slepian D (1978) Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: the discrete case. Bell Syst Tech J 57:1371–1430.
Slifstein M, van de Giessen E, Van Snellenberg J, Thompson JL, Narendran R, Gil R, Hackett E, Girgis R, Ojeil N, Moore H, D'Souza D, Malison RT, Huang Y, Lim K, Nabulsi N, Carson RE, Lieberman JA, Abi-Dargham A (2015) Deficits in prefrontal cortical and extrastriatal dopamine release in schizophrenia a positron emission tomographic functional magnetic resonance imaging study. JAMA Psychiatry 72:316–324.
Squire LR (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev 99:195–231.
Sreenivasan KK, D'Esposito M (2019) The what, where and how of delay activity. Nat Rev Neurosci 20:466–481.
Tallon-Baudry C, Bertrand O, Peronnet F, Pernier J (1998) Induced gamma-band activity during the delay of a visual short-term memory task in humans. J Neurosci 18:4244–4254.
Teyler TJ, DiScenna P (1986) The hippocampal memory indexing theory. Behav Neurosci 100:147–154.
Teyler TJ, Rudy JW (2007) The hippocampal indexing theory and episodic memory: updating the index. Hippocampus 17:1158–1169.
Thomson DJ (1982) Spectrum Estimation and Harmonic Analysis. Proc IEEE 70:1055–1096.
Thomson AM, West DC, Wang Y, Bannister AP (2002) Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2-5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling in vitro. Cereb Cortex 12:936–953.
Thorpe SJ, Fabre-Thorpe M (2001) Seeking categories in the brain. Science 291:260–263.
Tomita H, Ohbayashi M, Nakahara K, Hasegawa I, Miyashita Y (1999) Top-down signal from prefrontal cortex in executive control of memory retrieval. Nature 401:699–703.
Tsodyks MV, Markram H (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proc Natl Acad Sci U S A 94:719–723.
Tucker TR, Katz LC (2003) Recruitment of local inhibitory networks by horizontal connections in layer 2/3 of ferret visual cortex. J Neurophysiol 89:501–512.
Tully PJ, Lindén H, Hennig MH, Lansner A (2013) Probabilistic computation underlying sequence learning in a spiking attractor memory network. BMC Neurosci 14:P236.
Tully PJ, Hennig MH, Lansner A (2014) Synaptic and nonsynaptic plasticity approximating probabilistic inference. Front Synaptic Neurosci 6:8.
Tully PJ, Lindén H, Hennig MH, Lansner A (2016) Spike-based Bayesian-Hebbian learning of temporal sequences. PLoS Comput Biol 12:e1004954.
Ursino M, La Cara GE (2006) Travelling waves and EEG patterns during epileptic seizure: analysis with an integrate-and-fire neural network. J Theor Biol 242:171–187.
van der Velde F, de Kamps M (2015) The necessity of connection structures in neural models of variable binding. Cogn Neurodyn 9:359–370.
Voges N, Guijarro C, Aertsen A, Rotter S (2010) Models of cortical networks with long-range patchy projections. J Comput Neurosci 28:137–154.
Volianskis A, Jensen MS (2003) Transient and sustained types of long-term potentiation in the CA1 area of the rat hippocampus. J Physiol 550:459–492.
Volianskis A, France G, Jensen MS, Bortolotto Z. a, Jane DE, Collingridge GL (2015) Long-term potentiation and the role of N-methyl-D-aspartate receptors. Brain Res 1621:5–16.
Wahlgren N, Lansner A (2001) Biological evaluation of a Hebbian-Bayesian learning rule. Neurocomputing 38–40:433–438.
Wickens JR (2009) Synaptic plasticity in the basal ganglia. Behav Brain Res 199:119–128.

Yoshimura Y, Callaway EM (2005) Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. Nat Neurosci 8:1552–1559.
Yoshimura Y, Dantzker JLM, Callaway EM (2005) Excitatory cortical neurons form fine-scale functional networks. Nature 433:868–873.
Zimmer HD, Mecklinger A, Lindenberger U (2006) Levels of binding: types, mechanisms, and functions of binding in remembering. In: Handbook of binding and memory: perspectives from cognitive neuroscience, pp 3–22. Oxford: Oxford UP.
Zimmer HD, Mecklinger A, Lindenberger U (2012) Handbook of binding and memory: perspectives from cognitive neuroscience. Oxford: Oxford UP.
Zmigrod S, Colzato LS, Hommel B (2014) Evidence for a role of the right dorsolateral prefrontal cortex in controlling stimulus-response integration: a transcranial direct current stimulation (tDCS) study. Brain Stimul 7:516–520.
Zufferey PD, Jin F, Nakamura H, Tettoni L, Innocenti GM (1999) The role of pattern vision in the development of cortico-cortical connections. Eur J Neurosci 11:2669–2688.