CellPress

Feature Review

Sustained Activity Encoding Working Memories: Not Fully Distributed

Matthew L. Leavitt,[1] Diego Mendoza-Halliday,[2] and Julio C. Martinez-Trujillo[3,*]

Working memory (WM) is the ability to remember and manipulate information for short time intervals. Recent studies have proposed that sustained firing encoding the contents of WM is ubiquitous across cortical neurons. We review here the collective evidence supporting this claim. A variety of studies report that neurons in prefrontal, parietal, and inferotemporal association cortices show robust sustained activity encoding the location and features of memoranda during WM tasks. However, reports of WM-related sustained activity in early sensory areas are rare, and typically lack stimulus specificity. We propose that robust sustained activity that can support WM coding arises as a property of association cortices downstream from the early stages of sensory processing.

Introduction

The memory systems of the brain resemble time machines for thought: they transport sensory experiences from the past to the present, to guide our current decisions and actions. Memories have been classified into long-term, stored for time intervals of days, months, or years, and short-term, stored for shorter intervals of seconds or minutes. There is a consensus that these two types of memories involve different brain systems and have different underlying mechanisms.

WM is a particular type of short-term memory that involves the maintenance and manipulation, usually for a matter of seconds, of information either recently acquired from the environment or retrieved from long-term storage [1]. The main theoretical distinction between WM and other forms of short-term memory is that WM requires manipulation of the remembered information rather than only storage [2]. For many neurophysiologists, however, WM and short-term memory are used interchangeably, usually to refer to the maintenance aspect of both constructs [3]. In this review we use the term WM to refer to maintenance of information. This review aims to summarize findings from studies conducted over the past four decades and relate them to more recent reports, focusing primarily on electrophysiological studies in non-human primates.

Brain Signal Correlates of WM

The development of techniques to measure neural signals in awake, behaving subjects (animals and humans) has allowed researchers to relate the variations in such signals to specific behaviors to provide insight into the neural basis of cognition. Techniques to measure neural activity in behaving animals can be classified as measuring signals at the single-neuron level (single-unit activity, SUA), measuring signals integrated over tissue volumes (multiunit activity,

Trends

Neuronal correlates of WM include sustained spiking activity in individual and populations of neurons, and changes in BOLD signal intensity. The similarity of the information contained in these different signals depends on their strength and the spatial scale of the topography in the recorded area.

Robust, sustained single-unit spiking representing the contents of WM is present in association areas of the parietal, frontal, and temporal lobes.

Sustained single-unit activity representing the contents of WM is absent in early sensory cortices or consists of a transient increase in baseline activity encoding information about the location of spatial attention, but not the features of the memorandum.

We propose that the ability to generate robust sustained spiking activity capable of supporting WM coding emerges in association cortices, downstream from initial cortical processing.

[1]Department of Physiology, McGill University, Montreal, QC H3G 1Y6, Canada
[2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Robarts Research Institute, Brain and Mind Institute, Department of Psychiatry, and Department of Physiology and Pharmacology, University of Western Ontario, London, ON N6A 5B7, Canada

*Correspondence:
matthew.l.leavitt@gmail.com
(M.L. Leavitt) and
julio.martinez@robarts.ca
(J.C. Martinez-Trujillo).

MUA; and local field potentials, LFPs), and measuring secondary signals resulting from SUA, MUA, and LFPs such as the blood oxygenation level-dependent (BOLD) signal. To properly interpret results from studies of WM referenced in this review, it is crucial to consider what the signal is measuring (e.g., action potentials for SUA vs metabolic activity for BOLD), its temporal and spatial resolution, and the cortical topography and anatomy of the brain region in which it is measured. Box 1 offers a summary of the techniques we reference in this review.

Whether SUA, MUA, LFPs, and BOLD signals contain similar information depends on the topography and architecture within the recorded region. Take for example a single neuron with a receptive field located in the upper right of the visual field. If the neuron is embedded into a cluster of excitatory neurons with receptive fields positioned in the lower left of the visual field (i.e., 180° away), then recording SUA could identify and isolate the tuning properties of this unit. However, MUA recording would only show the tuning of the neuronal cluster, dominated by cells with opposite tuning (Figure 1A). Such a discrepancy would not exist if all neurons within the area covered by the electrode were similarly tuned.

This is not a hypothetical scenario: studies recording SUA and MUA using 4 mm × 4 mm multielectrode arrays in area 8a of the lateral prefrontal cortex (LPFC) have shown that neurons with receptive fields tuned for different visual field quadrants can be separated by distances of 0.4 mm [4,5]. Contrast this with area V4, in which neurons with receptive fields at similar

Box 1. Techniques Used To Explore Neural Correlates of WM

Single-Unit Activity (SUA)

The finite resistance of the brain tissue extracellular environment creates the mean extracellular field potential (EFP) that can be measured by an electrode relative to a reference. Placing a high-impedance electrode close to a neuron allows it to record changes in the EFP that reflect the action potentials or spikes fired by the cell. SUA recordings have been the gold standard of electrophysiological studies in behaving animals for at least two main reasons: (i) they have the highest spatial and temporal resolution (spatial resolution of about 0.14 mm and temporal resolution as high as 70 kHz depending on the recording system capabilities) [7], and (ii) they measure spikes, considered to be the fundamental unit of information transmission in the nervous system.

Multiunit Activity (MUA)

The aggregated spiking activity of multiple single neurons. It can be measured with relatively lower impedance electrodes than those used to measure SUA. MUA can be obtained by high-pass filtering recorded signals at 300 Hz. If neurons within the recorded area have homogeneous tuning properties, then SUA and MUA can be considered to be similarly tuned.

Local Field Potentials (LFPs)

Low-frequency fluctuations of the EFP in the vicinity of a recording electrode measured relative to a reference. LFP signals are usually filtered below 200 Hz. LFPs strongly depend on the geometry of the synaptic trees in the recording site. In the cortex, pyramidal cells oriented orthogonal to the cortical surface (dendrites on apical layers and axons in deep layers) are the main contributors to the signal. LFPs are thought to reflect a weighted average of synchronized dendrosomatic components of the synaptic signals of a neural population, as well as non-synaptic events such as voltage-dependent membrane and spike after potentials from within 0.5–3 mm of the electrode tip [7].

fMRI Activation

The BOLD signal originates from the realignment of hydrogen nuclei to a low-energy state in a magnetic field after a perturbation. The signal termed T2* is orthogonal to the field orientation, and it correlates with changes in regional blood flow and hemoglobin concentration that in turn correlate with the neural activity within a volume of tissue. fMRI resolution is determined by the relationship between the local vasculature, the functional organization of a given area, and the size of each voxel (i.e., a cubic volume within which the BOLD signal is integrated and measured). Spatial resolution can reach 1 mm^3 and temporal resolution is in the order of hundreds of milliseconds to seconds. BOLD responses primarily correlate with input strength into a given region and local processing of neuronal information (LFPs) [7].

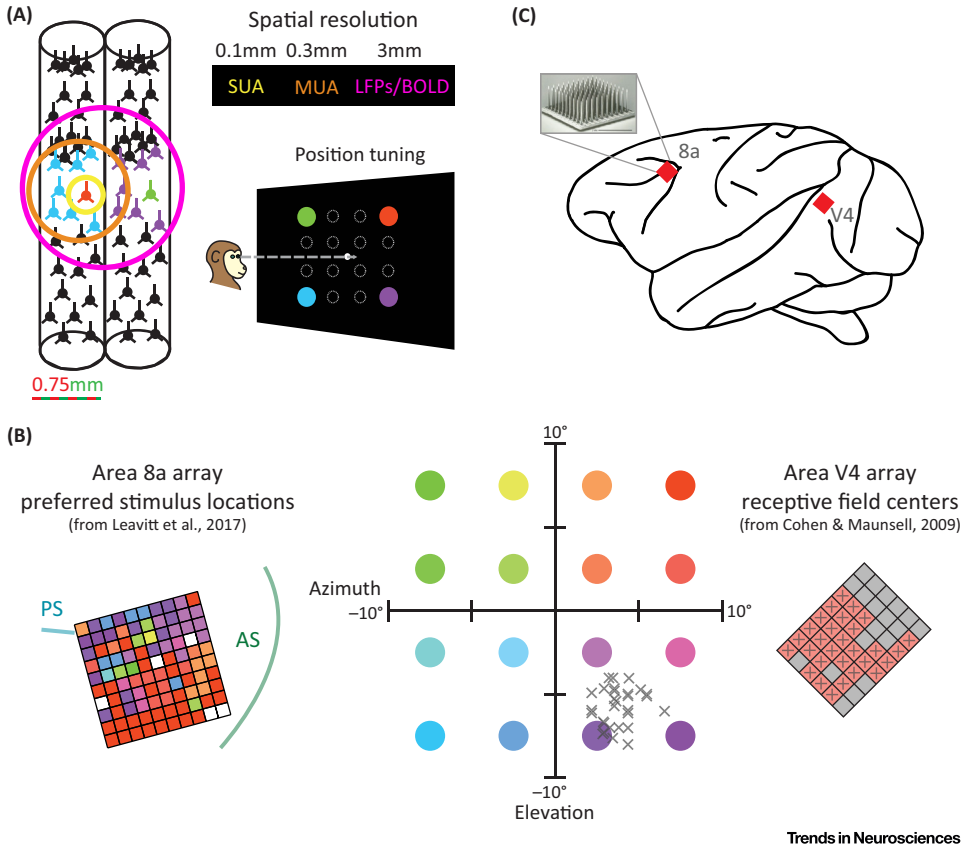

Figure 1. The Integration Volume of a Measured Signal and the Strength of Local Topography Can Limit the Amount of Information that Can Be Obtained about a Stimulus. (A) Hypothetical cortical columns containing neurons with heterogeneous tuning for the spatial location of a stimulus (i.e., the region is weakly topographic; stimulus locations are illustrated by colors are depicted in the right panel). The circles indicate the resolution or integration area of different signals: single-unit activity (yellow), multiunit activity (orange), and LFPs/BOLD (magenta). If a single neuron with a receptive field located in the upper right of the visual field is embedded into a cluster of excitatory neurons with receptive fields positioned in the lower left of the visual field, SUA could identify and isolate the tuning properties of this unit. However, MUA recording would only show the aggregate tuning of the cluster. (B) Position of microelectrode arrays implanted in two areas: V4 (right) and LFPC area 8a (left). (C) Receptive field topography of the signals from the microelectrodes in (B), demonstrating the amount of the visual field (center) represented in areas 8a (left) and V4 (right). The data collected from 8a [4] comprise mnemonic responses to one of 16 stimuli. The different locations are represented by different colors which are overlaid on to the array electrodes to denote the selectivity of the neuronal activity recorded on each electrode, shown on the left. The size of the stimuli is denoted by the grey circles in each quadrant of the visual field, shown in the center. The data collected from V4 (courtesy of Drs Cohen and Maunsell [6]) comprise receptive field centers (denoted by x in the center panel), as determined using fine-grained response mapping. The quadrant containing the receptive field center for each neuron is overlaid on the array electrode on which that neuron was recorded. (C) Multielectrode arrays implantation locations for data shown in (C). Abbreviations: AS, arcuate sulcus; BOLD, blood oxygen level-dependent; LFP, local field potential; LPFC, lateral prefrontal cortex; MUA, multiunit activity; PS, principal sulcus; SUA, single-unit activity.

locations are spread over areas as large as 3 mm [6] (Figure 1B,C). Thus, measurements that integrate signals over areas larger than 0.4 mm will not be able to characterize the tuning of single neurons and neuronal populations in area 8A, but will reliably do so in V4.

The same principles apply to the relationship between the BOLD signal and SUA/MUA. If the population of neurons driving the metabolic demands is homogenous over the volume measured by a voxel (e.g., 1–8 mm^3), measurements of BOLD, SUA, and MUA will yield similar tuning estimates. However, if single neurons or groups of neurons with different tuning are

clustered within a space smaller than the voxel resolution, BOLD measurements will not reflect the coding properties of single neurons (Figure 1A, blue circle). Moreover, because BOLD signal more strongly correlates with LFPs, which reflect the interactions of excitatory and inhibitory postsynaptic potentials (e.g., EPSPs, IPSPs, low-frequency components of spikes), the relationship between spiking activity (SUA/MUA) and BOLD can be non-linear [7,8]. Thus, researchers should be cautious when extrapolating BOLD measurements from neuronal tuning properties and account for neuronal selectivities within a voxel volume.

Sustained Spiking Activity as a Neural Correlate of WM

The first SUA correlates of WM were reported in 1971 in macaque monkeys as sustained increases in the firing rate of neurons in the LPFC (area 46 of Brodman) during a WM task [9], soon followed by a second study reporting a similar finding [10]. These findings supported reports from decades earlier in non-human primates showing impairments in WM after PFC lesions [11]. The task used in these studies became a canonical paradigm for studying WM: a memorandum (also termed a cue or sample) is presented to the subject for a short time (Figure 2A, left panel). After it disappears, the subject must remember its location or a specific feature(s) during a delay (or memory) period (middle panel). Finally, the subject must use this memorized information to guide a response, such as a saccade to the remembered location or object in an oculomotor delay response (ODR) task (right upper panel), or a match/non-match judgment relative to a test stimulus in a delayed-match-to-sample (DMTS) task (lower panels).

One limitation of the ODR task is that it confounds signals related to WM for the location of the sample and preparation for a motor response. Take for example the proposal by an early study that prefrontal area 46 neurons have 'memory fields' signaling the remembered location of a sample during an ODR task [12]. If a saccade is made to the location of the sample, neural activity during the delay period may reflect memory for the sample location or preparation of a

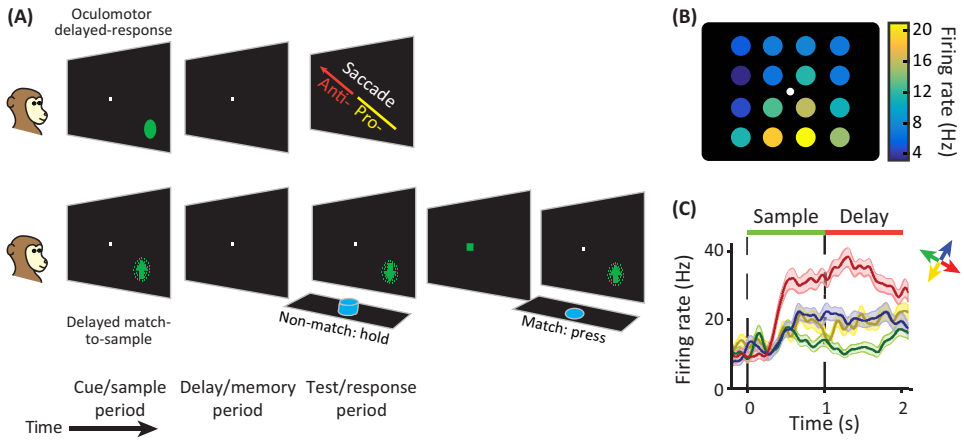

Figure 2. Example Working Memory Tasks and Single-Unit Responses. (A) Task designs for oculomotor delayed response (ODR, top) and delayed match-to-sample (DMTS, bottom) tasks. In the ODR task, a stimulus is presented at one of multiple possible locations in the visual field, removed, and then after a delay period the subject is cued to saccade to the remembered location. In the anti-saccade version of the task, the subject saccades to the location opposite the presented stimulus. In the DMTS task, the subject must remember a feature of the sample stimulus (e.g., motion direction). They are then shown a test stimulus, and must indicate whether the test stimulus matches the sample stimulus (e.g., press a lever). If the test stimulus is not a match, they may continue to be shown additional test stimuli until they are shown a match. (B) Mean activity of an example 8a neuron across the first 500 ms of the delay/memory period when remembering the location of a presented stimulus, for each of the 16 different possible locations (adapted from [4]). (C) Responses of an example area 9/46 neuron to four different motion directions of a random dot kinematogram in the receptive field of a neuron during the sample and delay periods (adapted from [88]).

saccade toward that location. This issue was addressed by experiment [13] which recorded the responses of single neurons in prefrontal area 46 during a task in which a saccade could be made to a location opposite to the remembered location (antisaccade), or to the same remembered location (pro-saccade) (right upper panel of Figure 2A). Many prefrontal neurons encoded the remembered location rather than the saccade location. These studies not only provide evidence in favor of neural correlates of WM processes in the primate PFC but also established the foundations for subsequent studies in non-human primates. The work of Goldman-Rakic, her collaborators, and trainees has formed the bedrock of our knowledge of the primate WM circuits [14], and led to an eminent hypothesis: 'sustained spiking activity in lateral prefrontal cortex neurons provides a mechanism for maintenance of information during working memory tasks'. The specific neural mechanisms by which sustained activity arises are still the subject of intensive research [15,16].

It is possible that sustained activity *per se* does not encode WM. Several studies have shown that populations of neurons with heterogeneous temporal dynamics can encode stable representations of stimuli during the delay period of WM tasks [17–19]. However, these studies were all conducted using populations of neurons recorded from LPFC, a region known to contain a significant proportion of neurons that exhibit sustained activity. Mnemonic maintenance does not necessitate that every neuron in a given region fires continuously for the entirety of a delay period. Furthermore, even if neurons do not exhibit tonically sustained firing throughout the entirety of a delay period, integrating their firing rates over this period will still demonstrate an increase in firing rate. Thus we apply this standard in our evaluation of the literature for this review.

Mapping Reports of Sustained Activity Across the Brain

SUA correlates of WM were isolated in early studies in areas of the parietal cortex [20] and temporal lobe [21]. These findings led to the current view that the sustained activity underlying WM is not exclusive to prefrontal neurons, but exists within a network of cortical association areas [3]. The role of each area in the origin of this phenomenon and in WM in general remains unclear. In recent decades several studies have confirmed these findings, and additionally reported sustained activity in a variety of brain areas including extrastriate cortex, superior colliculus, and the basal ganglia. To gain a bird's-eye perspective on the state of the research, we generated a database of SUA and MUA studies of WM (Table 1) comprising >90 studies involving tasks that incorporate delay periods between a stimulus and a response (i.e., a task design similar to Figure 2A). We then determined the cortical areas examined in each study and whether the authors positively or negatively reported sustained activity that encoded memory for locations, features, or objects (Figure 3). We then mapped the recorded areas into a comprehensive map of brain areas kindly provided by Dr Henry Kennedy (INSERM, France) that includes 91 different cortical areas [22] (Figure 4). Our database does not account for different studies using different stimuli and variations of the delay tasks, different animals and training protocols, or different criteria to consider sustained activity as a neural correlate of WM. Our intent is not to perform a meta-analysis (which would go beyond the scope of this review). We simply desire to provide a point of departure for consolidating facts, and identifying gaps and contradictions in our knowledge of WM networks in the primate brain, with the goal of guiding future studies. This database and interactive versions of Figures 3 and 4 will be made available online at http://martinezlab.robarts.ca. We aim to update this with current literature and invite researchers to submit new or overlooked studies to the database.

There is one caveat regarding the interpretation of electrophysiological studies that bears mentioning: it is not an uncommon practice for researchers to target neurons for recording that exhibit specific response properties, and this can introduce a sampling bias in subsequent analyses. We did not account for this when aggregating findings for our database. One

Table 1. Works Reporting WM-Related Activity in Different Brain Regions

Area	Positive finding	Negative finding
10		[23]
11	[89–92]	[93]
12	[89–92,94–97]	
13	[89–92]	[93]
14	[89–92]	[93]
1		[27,50,51]
24a	[91,98–102]	
24b	[91,98–102]	
24c	[91,98–102]	
24d	[91,98–102]	
2	[49]	[27,50]
3		[27,50,51]
45a	[56,94,95,97,103,104]	
45b	[56,94,95,97,103]	
46d	[9,10,12,13,56,88,94–97,105–118]	
46v	[9,10,12,13,56,88,94–97,105,109,110,112–118]	
5	[119]	
7a	[20,107,108,111,120–122]	[37]
8b	[106,123]	
8l	[5,12,88,105,109,110,112,114,115,122,124–128]	
8m	[5,12,88,105,108–112,114,115,122–128]	
8r	[5,12,88,105,122,124,127–129]	
9	[107]	
9/46d	[9,10,12,13,56,88,92,94–97,102,105–107, 113–117,122,123,125,126,128,130–134]	[41,57]
9/46v	[9,10,12,13,56,88,92,94–97,102,104–107, 113–117,122,123,125,126,128,130–134]	[41,57]
Core	[135]	[53–55]
Entorhinal	[136]	
F2	[24,27,123]	
F3	[137,138]	
F4	[139]	
F5	[58,139,140]	
F6	[137,138]	
F7	[24,27,59,133,141,142]	
LB	[55]	
LIP	[42,102,143–147]	
MST	[88]	[37]
MT		[37,39,41,42,88]
Perirhinal	[148]	
S2	[50,51]	
STPc	[135]	
STPl	[135]	

Table 1. (continued)

Area	Positive finding	Negative finding
STPr	[135]	
Subiculum	[149,150]	
Temporal pole	[148]	[151]
TEam a	[35,94,152–154]	
TEam p	[35,94,152–154]	
TEad	[21,35,94,152–155]	[156]
TEav	[35,94,148,152–154,157]	
TEpd	[21,155]	[55,156]
V1	[31,32]	[33–35]
V2	[34]	
V4	[34,36]	[33,37,38]
VIP	[134]	
Superior colliculus	[158]	
Mediodorsal nucleus of thalamus	[159]	
Ventral posterior lateral nucleus of thalamus		[52]
Caudate nucleus	[160]	
Hippocampus	[149,150]	

advantage of chronically implanted microelectrode array recordings is that preselection procedures are not typically performed, thus limiting the likelihood of sampling bias.

Sustained Spiking Activity in Association Areas

Figures 3 and 4 show that sustained activity has been consistently isolated by single-cell studies in the PFC (orbitofrontal areas 11–14, anterior cingulate areas 24a–d, lateral prefrontal areas 8, 9, 9/46, and 45) with the exception of the frontal pole (area 10), where a single study has reported a negative result [23]. In the parietal cortex, sustained activity has been isolated by multiple studies in areas LIP and 7A. Sustained activity has also been isolated in different visual areas of the inferotemporal cortex (TEad, TEav, TEama, TEamp, TEapd, temporal pole, subiculum, STPr, STPl, STPc, SII, and perirhinal). The number of studies reporting sustained activity in areas of the LPFC exceeds that of any other region.

Sustained activity has also been reported in areas in the frontal lobe (areas F2, F3, F4, F5, F6, and F7). These areas are considered to be premotor areas [24], but are also known to encode somatosensory signals [25–27]. Here the distinction between sustained activity underlying representations of locations, features, or objects and sustained activity underlying preparation for action becomes complicated. A main issue is whether representing an action or a movement goal should be considered as a fully separate process from representing a remembered location, feature, or object. Clarifying this goes beyond the scope of this review. However, the fluid nature of many sensorimotor transformations precludes a clear boundary between these different types of representations. For example, sustained activity representing motion direction (a non-spatial feature) has been found in area 8Ad of the PFC, a region that is also known to contain neurons that fire bursts of action potentials before the onset of a saccade [5]. The spatially-selective sustained activity of these neurons during the delay period of an ODR task could merge with a burst of action potentials before the saccade is executed. If one defines sustained activity as the ability of a circuit to maintain information in the absence of inputs, this

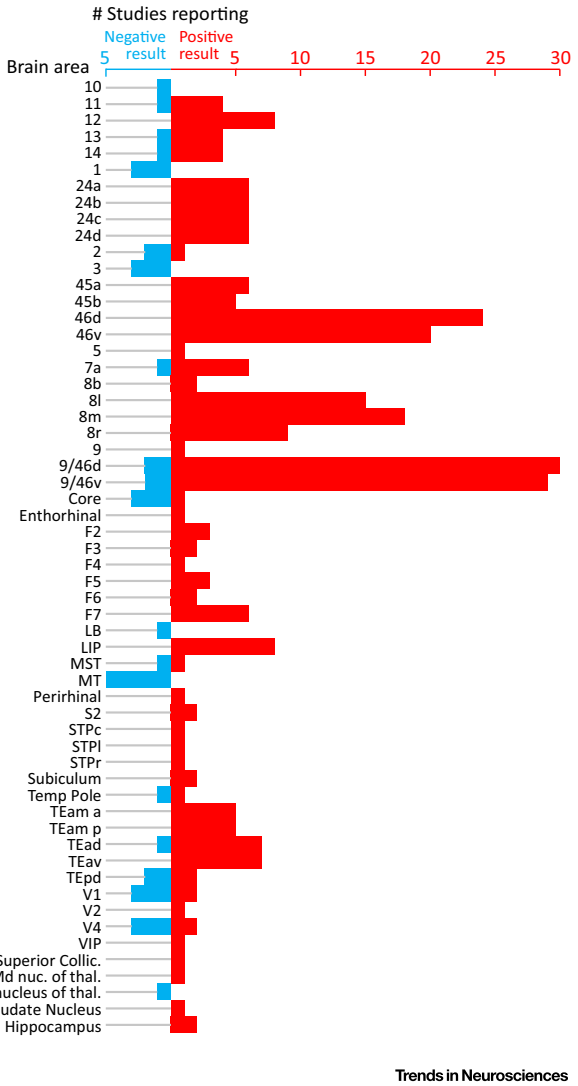

Figure 3. Positive and Negative Reports of Working Memory-Related Activity in Different Brain Regions. The number of positive reports (red) and negative reports (blue) of sustained activity across brain regions. The criterion defining a positive report was reported delay activity during a working memory task, and the criterion for a negative report was an absence of delay activity reported in the study. References are provided in Table 1. Abbreviations: Md nuc thal, mediodorsal nucleus of thalamus; Vpl nuc thal, ventral posterolateral nucleus of thalamus.

includes the presaccadic activity, which may be related to motor preparation. Thus, sustained activity can also be considered as a neural correlate of motor preparation or planning. One way to approach this issue is to consider sustained activity as a mechanism that allows mapping information from the past onto current events along the chain of sensorimotor transformations.

Sustained Activity in Early Sensory Cortices

One increasingly popular hypothesis is that sustained activity encoding the contents of WM is a property of all cortical neurons, including those in early sensory areas such as V1. This hypothesis can only be tested by recording the activity of single neurons in early sensory areas during WM tasks. However, negative results in such studies may be under-reported because of publication bias against negative results [28–30]. Nevertheless, we were able to find both positive and negative findings in studies using delayed response tasks in early sensory areas of multiple modalities.

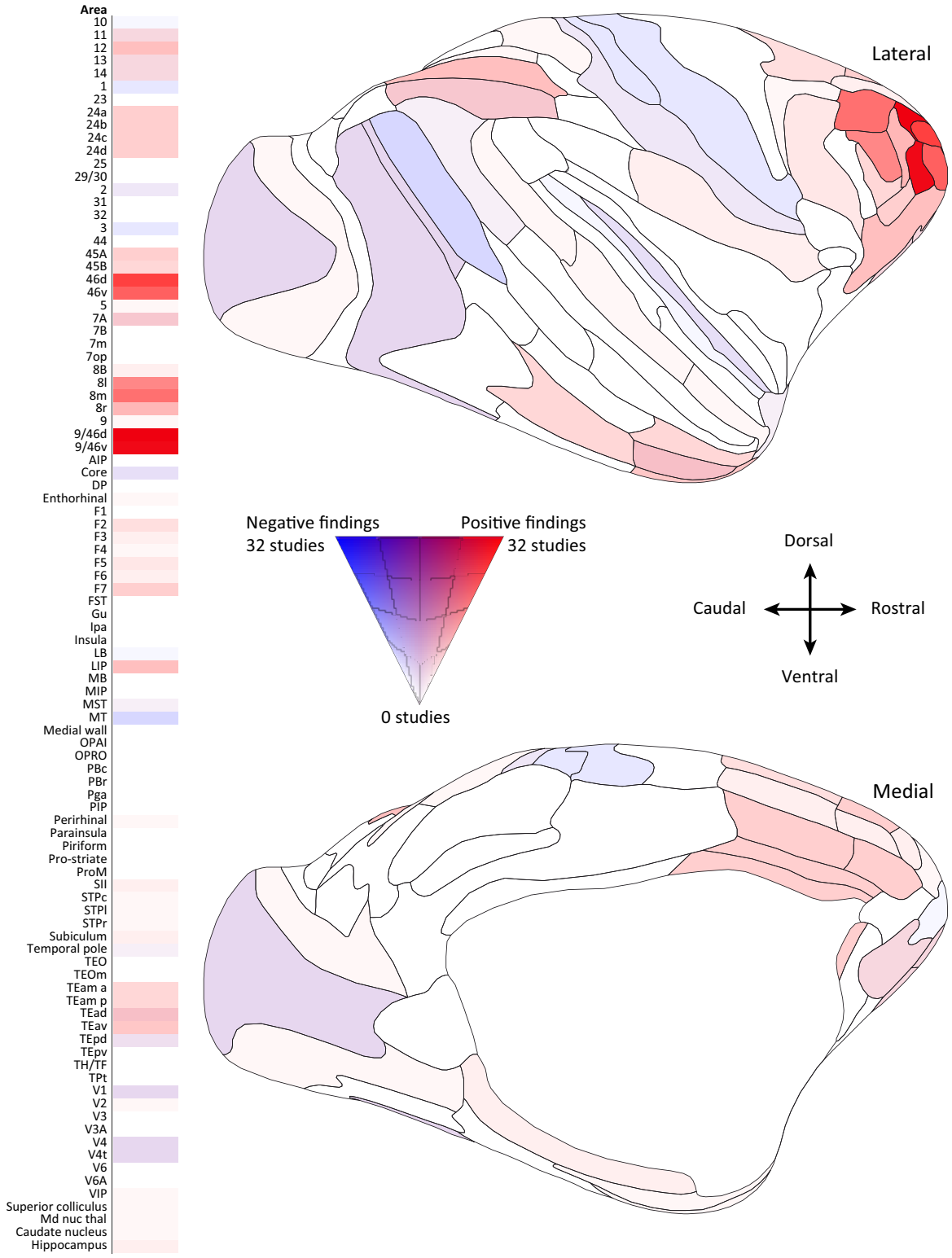

Figure 4. Brain Map of Evidence for Working Memory (WM)-Related Sustained Activity. The data from Figure 3 are visualized on an inflated macaque brain, kindly provided by Dr Henry Kennedy (Lyon, France) [22]. Each region is color-coded according to the number of studies positively or negatively reporting WM-related

(Figure legend continued on the bottom of the next page.)

Primary Visual Cortex (Area V1)

We found two studies reporting sustained MUA in area V1 encoding the remembered location of a curve segment. The first study used a figure–ground segregation paradigm in which the receptive fields of V1 neurons were tonically stimulated, and which could be responsible for the reported elevation in background firing during the memory period [31]. A second and more recent study using a similar paradigm controlled for this potential confound, reporting that sustained activity is present even in the absence of a stimulus in the receptive field during the first 500 ms of the delay period (Figure 2e,f of [32]). However, at least three other studies have reported an absence of sustained activity in area V1 when using a match-to-sample task for orientation [33,34] or color/shape [35]. Luck and coworkers reported that the 'baseline shift' (an increase in response during the delay period when the sample stimulus was inside vs outside the receptive field) was absent in V1, although it was present in areas V2 and V4 of the same animals. They interpreted baseline shift as an effect of attending to the spatial location of the sample – which happens to be inside the receptive of the recorded neurons – rather than as sustained activity related to WM. However, when they examined whether V2 and V4 neurons were selective for the features of the sample or memoranda (e.g., a specific orientation), or showed any selective enhancement relative to neurons poorly selective for the same feature, they found no effect. Considering that the task required matching the features of the sample to the test, one would conclude that neurons in areas V1, V2, and V4 do not carry information about the features held in WM during the delay period.

We identified several differences that can reconcile the seemingly contradictory findings from studies in V1. First, they used different tasks: curve-tracing in studies reporting positive results, and match-to-sample tasks in studies reporting negative results. Second, the curve-tracing studies reporting positive results used MUA recordings, while the studies reporting negative results used SUA recordings. Finally, the activity reported by the curve-tracing studies is considerably attenuated after cue offset, and decays to baseline within 500 ms. This is very different from the activity found in the PFC, parietal, and inferotemporal cortices (Figure 2B for examples), which can be as high as the activity evoked by the sample presentation, tuned for non-spatial features, sustained over the entire delay period, and only present during the delay period. It is possible that the small baseline shift (identified as sustained activity by the authors) reported in both studies results from using the same curve-tracing task and/or MUA recordings. In sum, the majority of V1 studies we examined failed to isolate sustained activity that encoded the contents of WM during delay periods of WM tasks.

Visual Area V4

In V4, an area in the ventral visual processing pathway targeted by V1 projections, we identified two studies reporting delay activity [34,36] and at least three studies reporting an absence of delay activity [33,37,38]. In the studies reporting positive results, the average magnitude of the delay activity was 1.8 Hz [36] and in the second study 4.3 Hz [34]. Although these small numbers can account for a modulation of 20–30% of the baseline firing rate, the relative number of spikes is very low in relationship to the activity evoked by a stimulus. This resembles the results of curve-tracing WM studies in area V1 that report a small modulation of baseline activity. One issue that is unclear is how much information these small modulations can provide about the identity or the target in comparison with the information provided by the sustained activity of downstream neurons in the inferotemporal cortex or the PFC. Unfortunately, the aforementioned studies do not provide a measure of the amount of information represented in such spiking activity [e.g., decoding analyses, receiver operating characteristic (ROC) analysis,

sustained activity. Note the tendency toward more negative reports in early sensory cortices, and more positive reports in association cortices. Also note that the color map is scaled by the largest total number of studies in any single area (32 studies, area 9/46d); there is no region with more than 32 negative or positive reports. Abbreviations, Md nuc thal, mediodorsal nucleus of thalamus.

mutual information, etc.]. This is necessary to properly evaluate the contribution of activity modulations in a given area to information coding during WM tasks.

Middle Temporal Visual Area (MT)

In area MT, part of the dorsal visual processing pathway and a downstream projection target of V1 neurons, we found five different studies that failed to report sustained activity in single neurons during delayed response memory tasks [37,39–42]. One study in area MT reported a decrease in response relative to baseline after the offset of a sample moving in the preferred motion direction of the neurons, and a rebound of responses coupled to the offset of a sample moving in the antipreferred direction [41]. Despite the fact that monkeys were capable of memorizing motion directions through the entire 1500 ms delay period, this effect was limited to the initial 300–400 ms of the delay period following stimulus offset, in contrast to sustained activity that lasts for the entire delay period in prefrontal neurons [43]. Hence, this effect is more likely to represent response adaptation [44] than WM-related activity. Another study recorded the activity of neurons in areas MT, MST (medial superior temporal area), 7a, and V4 of macaque monkeys during a match-to-sample task similar to that illustrated in Figure 1A (lower panel). They found that during the delay period the activity of neurons in the three areas was significantly higher than the activity before sample presentation. However, the effect was small and did not contain information about the identity of the sample (direction of motion) [37].

More recently, a study recorded the responses of MT, MST, and LPFC (areas 8d, 8v, 46/9d, and 46/9v) neurons in the same animals using a similar same match-to-sample task for motion direction and found no sustained activity during the delay period in MT but robust sustained activity in MST and LPFC. This study compared how neurons represent the sample direction during the sample (sensory stimulation) and delay periods of the task and found that (i) neurons in MT represent motion direction during the sample but not throughout the delay period; (ii) MST represent motion direction throughout both sample and delay periods, but representations were more robust during the sample than the delay; and (iii) in the LPFC (areas 8 and 9/46) motion direction representations during the sample and delay periods were of similar strength (Figure 5). To contextualize these results, consider that area MST is a polymodal association area receiving projections from MT. MST neurons have complex response properties such as receptive fields covering both visual hemifields and optic flow selectivity. MST lies at a similar stage in the dorsal visual pathway as areas of the anterior inferior temporal lobe lie in the ventral pathway [45]. Area MT, by contrast, is better compared to area V4: both MT and V4 neurons have receptive fields covering the contralateral hemifield and show relatively simpler selectivities [46]. If one considers the pattern of absence of robust delay activity in MT and V4, but presence in MST [40] and inferotemporal areas [35], it is reasonable to conclude that robust sustained activity encoding the contents of WM may emerge at stages in which neuronal response properties reach some threshold of complexity. For example, receptive fields in MST are much larger than in MT and extend to the ipsilateral visual hemifield. Attentional modulation of responses is also larger in area MST than in MT [47]. Finally, MST seems to integrate signals from the vestibular and visual modalities [48], thus it is considered to be a multisensory area. One can find a similar trend from area V4 to areas of the inferior temporal cortex.

Somatosensory Cortex

At least three studies have recorded the responses of neurons in area S1 (areas 1, 2, and 3) of macaque monkeys during delayed response tasks. Only one study reported sustained activity during the delay period [49], while the others reported no delay activity in S1 neurons [27,50,51]. One of the studies recorded the responses of neurons in S1 and dorsal premotor cortex (DPC, areas F2 and F7) in the same animals and during the same task [27]. They found no sustained activity during the delay period of a vibrotactile frequency discrimination task (VTFD) in SI, but robust sustained activity in DPC. Moreover, the activity of neurons in S1 poorly

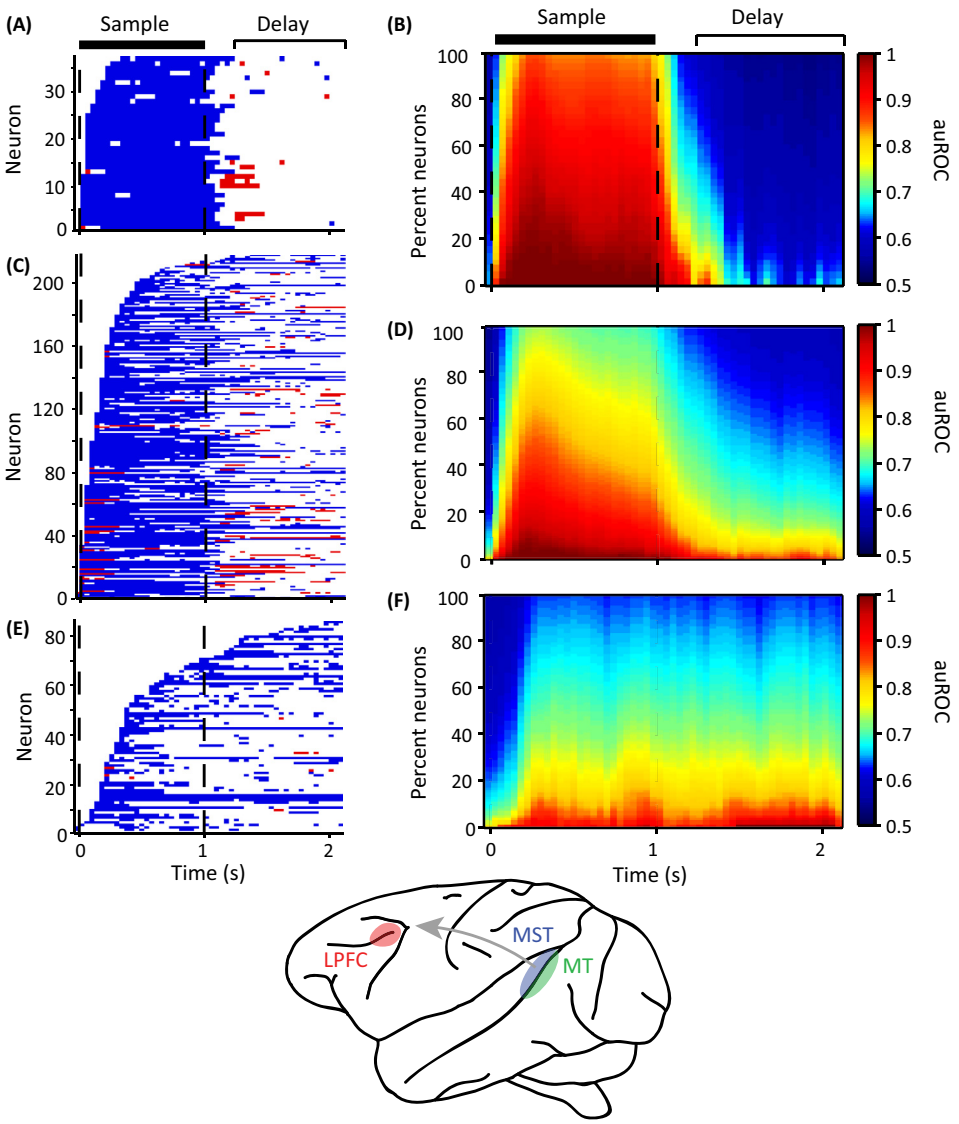

Figure 5. Working Memory (WM)-Related Activity in MT, MST, and LPFC. ROC analysis of the selectivity in neurons recorded from three different areas (A, top, MT; C, middle, MST; E, bottom, LPFC areas 8a and 9/46). The abscissa indicates the time period, sample, and delay of the DMTS task illustrated in Figure 2A and the ordinate the recorded neuron. Each row represents a neuron. Time bins in blue illustrate when the ROC yields significance for discriminating the preferred and the antipreferred direction (defined during the sample period) with the preferred direction evoking the larger response. Bins in red illustrate significance when the preferred direction gives a weaker response than the antipreferred. On the right, average auROC across MT (B), MST (D), and LPFC (F) neurons over time as a function of the percentage of averaged neurons (organized from maximum to minimum auROC in each time bin). At the bottom, a cartoon of the macaque brain illustrating the flow of information from MT to MST to LPFC. Adapted from [88]. Abbreviations: auROC, area under the ROC curve; DMTS, delayed match to sample; LPFC, lateral prefrontal cortex; MST, medial superior temporal area; MT, middle temporal visual area; ROC, receiver operating characteristic.

correlated with the behavior of the animals (choice probability), while the activity of neurons in DPC did correlate with behavior. Furthermore, one study showed that the thalamus does not show delay activity in the somatosensory modality, and that delay activity representing the contents of WM emerges downstream from S1 [52]. This pattern of results resembles the one found in visual areas: absence of delay activity in early sensory cortex, and increasingly robust

activity that increasingly correlates with the behavior of the animals as one moves further downstream [40]. This may be a general property of WM-related activity across sensory domains. Comprehensive reviews and analyses examining delay activity across different areas during the somatosensory-to-motor transformation involved in a vibrotactile discrimination task can be found in [25,26].

Auditory Cortex

There are comparatively fewer studies of auditory WM than visual WM. We found one study that recorded the responses of neurons in AI during a match-to-sample task for auditory stimuli. It was reported that most neurons show transient enhancement or suppression during the initial part of the delay period, but no robust sustained activity [53,54]. WM-related delay activity in the auditory belt (part of primary auditory cortex) and core is also minimal [55]. This is in contrast to a recent study reporting delay activity for sound stimuli in the LPFC areas 45 and 9/46 [56]. Another study in LPFC reported that only 10% of the neurons show sustained activity during the delay period [57]. Interestingly, a series of studies using an acoustic flutter discrimination task found sustained delay activity in ventral premotor [58], SII [50], and SMA [59]. While there are fewer studies in auditory than in visual cortex, the basic pattern matches that found in other WM modalities such as visual and somatosensory, supporting the hypothesis that sustained activity emerges as a property of association cortices.

BOLD Signal and Delay Activity in Early Sensory Areas

One of the first reports of a BOLD signal correlate of WM was made by Courtney and colleagues in 1997 using a match-to-sample task for faces. They demonstrated increased BOLD activation in occipitotemporal areas during the presentation of sample faces, and in prefrontal areas during the period in which subjects maintained a memory representation of the face [60]. In 1998, Belger and colleagues demonstrated increases in BOLD signal in the same areas during a delayed-response task, but also found activation in several other areas such as inferotemporal and cingulate regions. These findings have been corroborated by many laboratories ([2,61,62] for reviews). The results of these studies revealed a homology between sustained spiking activity isolated by SUA studies in monkeys and sustained BOLD activation in humans [63].

More recently, studies using multivoxel pattern classification analysis (MVPA) have shown it is possible to decode the contents of WM from the activity in early sensory cortices, even if the mean signal intensity in these areas during the memory delay period remains near baseline levels [64–66]. However, applying the same technique to BOLD signals in PFC, where WM-related activity has consistently found SUA, MUA, and LFPs, the classification performance of the algorithms remains close to chance [67]. This has been presented as evidence that early sensory cortex, not association cortex, encodes the contents of WM. A recent paper by Ester *et al.* is an exception to this trend [68]. Using an inverted encoding model, which can be more sensitive than traditional MVPA, they were able to robustly reconstruct remembered orientations from human PFC activity. Nevertheless, this is a complex issue that needs to be considered in greater detail to account for the relationships between BOLD signal, SUA, MUA, and LFP measurements.

Each axis in the multidimensional space of the MVPA is represented by a voxel. A voxel integrates BOLD responses over millimeters of cortical volume (Box 1 and Figure 1). Sensory cortex has a topographical organization characterized by cortical columns for relevant features (e.g., space, orientation or motion direction) [69], and thus voxels in sensory cortex likely contain neurons with homogeneous or similar tuning. In association areas of the prefrontal and parietal cortex, however, columnar organization may not follow the same principles. For example, neurons with dissimilar tuning can co-occur in small volumes, within a few hundred

micrometers of each other. This results in poorly tuned voxels with lower signal-to-noise ratio (e.g., Figure 1C). Studies of attention using single units and multiunit activity recorded with small microelectrode arrays (4 × 4 mm) in PFC can classify the allocation of attention or the contents of WM with high accuracy [4,5]. However, when the same studies use LFPs recorded in the same animals, region, and task trials, the classification performance deteriorates considerably, particularly in the low-frequency bands where most of the LFP energy is concentrated [70]. A recent study using SUA and MUA from a 4 × 4 mm array positioned in area 8A was able to decode which of 16 locations was being held in WM during an ODR task [4]. Moreover, the number of neurons required for reaching the highest possible classification accuracy was between 25 and 30 units. It is difficult to compare the classification performance reported by fMRI studies and by the SUA and MUA studies. However, the studies of SUA and MUA have proved that the spiking activity of a small number of simultaneously active LPFC neurons concentrated within a small cortical area can carry enough information to decode the contents of WM. By contrast, MVPA studies using BOLD signal in a homologous area of the human brain produce poor or negative results [2].

Interestingly, at least one functional imaging study has pointed out that MVPA in area V1 does not reflect the underlying columnar organization of the area [71]. Considering the results of studies reporting changes in LFPs in visual sensory cortex without measurable changes in spiking activity, it is very likely that the results of MVPA studies reflect increases in LFP activity rather than sustained spiking activity. Increases in LFP power can be explained by neurons in association areas with strong sustained activity feeding back into sensory cortices [40]. Such a mechanism would also explain what has been considered by some researchers as the most compelling evidence for the role of sensory areas in WM – that the precision of population tuning curves in areas V1 and V2 estimated from the delay period BOLD signal using MVPA predicts the fidelity with which a subject can reconstruct the remembered sample features at the end of the delay period [72]. This result is expected if representations held by tuned neurons in associative areas are the source of the top-down bias that modulates LFPs and BOLD signal in sensory areas [40]. Interestingly, a study using transcranial magnetic stimulation (TMS) has shown that stimulation of cortex including area MT changes the perception of coherent motion in a manner dependent on the direction of motion the subject holds in WM [73]. Although this has been used to support the claim that sensory areas encode WM [62], it could also be considered to support the position that top-down signals to sensory cortices influence visual perception. Studies using TMS of the frontal eye fields and the parietal cortex (superior parietal lobe and intraparietal sulcus) in humans reporting performance changes during WM tasks [74] support the causal role of these areas in WM.

A relevant question to ask is whether increases in LFP power in sensory cortices indicate involvement of these areas in the maintenance of WM or a mechanism different from sustained activity that can provide a substrate for WM functions. A recent study suggested that a form of temporary synaptic storage in sensory areas, which does not necessarily result in spikes but is visible in BOLD signal, could be the mechanism underlying WM maintenance [75]. However, physiological evidence directly measuring such synaptic storage mechanisms is lacking. Moreover, it is not clear how such a mechanism could account for performance in tasks different from match to sample – for example in tasks in which memoranda should trigger a future action in the absence of inputs into the synapses storing the memories (i.e., in a DMTS task when the location of the sample and the test are in different hemifields). This is not trivial, because synaptic potentials cannot travel from sensory areas to prefrontal and premotor areas where populations of neurons representing motor actions need to be activated by mnemonic representations. The only signal that can travel long distances through myelinated axons (e.g., from area MT to the LPFC, and from there to motor areas) in the nervous system is the action

potential. Thus, sustained trains of action potentials by single neurons in association areas are, so far, the most parsimonious mechanism for WM maintenance in the brain.

A Circuit Mechanism Underlying Sustained Activity

The most accepted theory for the origins of sustained activity underlying WM is that it originates within cortical circuits that support recurrent network dynamics [15,76,77]. Wang *et al.* have proposed that a microcircuit composed of pyramidal neurons and three different types of interneurons, differentiated by containing different calcium-binding proteins, can generate sustained activity in the PFC. The different types of interneurons are perisoma-targeting cells containing parvalbumin (PV), peridendritic-targeting cells containing calbindin (CB) and also possibly somatostatin, and interneuron-targeting cells containing calretinin (CR) and also vasoactive intestinal peptide. Within such a circuit, activity is sustained after the number of recurrent connections exceeds a given threshold [78]. Considering this modeling work, and our proposal that sustained activity is a property of association cortices, one straightforward prediction is the existence of microcircuit differences between early sensory cortices and association cortices that should be linked to the ability of a network to generate sustained activity (e.g., excitatory recurrent connectivity). Indeed, such differences have been reported. For example, layer III pyramidal neurons in the PFC have 16-fold more spines than in area V1. This is not simply because of larger cell bodies in pyramidal neurons but reflects an increase in spine density (i.e., the number of spines per unit of dendritic length is fourfold higher in PFC than in V1) [79]. Thus, pyramidal neurons in the PFC are more heavily interconnected than in area V1. Moreover, anatomical studies in the PFC of macaques have revealed stripe-like patterns and long lateral connections originating in layer IV that have the potential to connect cortical columns over millimeters [80]. Electrophsyiological results also allude to differentiated circuit properties even between the PFC and parietal cortex: the PFC contains significantly more neurons with inverted tuning [43], and cross-correlation strength (i.e., effective connectivity) is stronger but more spatially restricted in parietal cortex than in PFC [81]. Interestingly, it has been shown that decay time constants of neurons in prefrontal and association areas are longer than those of sensory areas, alluding to increased recurrent strength being necessary for sustained activity [77]. Furthermore, elevated activity is observed after the removal of stimuli during passive viewing even in monkeys naïve to any training on WM tasks, further supporting the notion that the capability for sustained activity is an intrinsic property of association cortex ([82] for review).

Another important difference between association areas such as the PFC and early sensory areas such as V1 is found in the proportions of different types of interneurons. These cells are thought to play an important role in recurrent dynamics in neural circuits, and therefore in the origin of sustained activity [83]. In area V1, PV-containing cells are the most prevalent type of interneuron. By contrast, in the PRC, CR cells are the most prevalent type [84]. Interestingly, a study has reported that in human and monkey brain slices of PFC and temporal cortex, CR neurons generate sustained spiking activity after stimulation by a single action potential, and that spiking is terminated by bursts of action potentials [85]. CR interneurons have been shown to selectively drive the activity of CB interneurons that project to dendritic trees of pyramidal cells in the rat hippocampus [86]. Further research of this cortical circuitry may provide insight into the cellular mechanisms of WM.

Finally, the PFC, but not area V1, is a target of afferents from dopaminergic, noradrenergic, serotoninergic, and cholinergic systems. Studies in the monkey PFC have shown that sustained activity is sensitive to manipulations of dopaminergic systems [87]. Although it is not clear how these differences would impact on sustained activity, there is plenty of evidence that cortical architecture varies substantially from primary sensory areas to association areas. Future studies will be necessary to elucidate the precise properties that allow particular cortical circuits to exhibit sustained activity (see Outstanding Questions).

Outstanding Questions

Studies of sustained activity during WM tasks are still lacking in several macaque brain areas. There is a need for studies recording from different areas using the same animals and tasks to clarify this issue.

What are the neural circuit mechanisms of sustained activity in associative areas? Although there are computational models of sustained activity underlying WM that make specific predictions, there is a lack of studies testing these predictions. The most fruitful approach should combine anatomical and physiological methods. Calcium-imaging and associated techniques that are able to label neuronal types and register their activity during WM tasks may open new venues to clarify this issue.

Although the evidence in favor of sustained activity as the mechanism underlying WM coding is strong, other proposed mechanisms such as synaptic storage and transient oscillatory dynamics need to be further investigated.

Acknowledgments

We would like to thank the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council (NSERC), and the Western Research Chair in Autism for funding, Drs. Christos Constantinidis and John Maunsell for their helpful comments on an initial version of this manuscript, Drs. Xiao-Jing Wang and John D. Murray for their thoughtful discussion on this topic, and Roberto Gulli for technical insights.

References

1. Baddeley, A. (2010) Working memory. *Curr. Biol.* 20, R136–R140
2. Postle, B.R. (2015) The cognitive neuroscience of visual short-term memory. *Curr. Opin. Behav. Sci.* 1, 40–46
3. Goldman-Rakic, P.S. (1995) Cellular basis of working memory. *Neuron* 14, 477–485
4. Leavitt, M.L. *et al.* (2017) Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles. *PNAS* 114, E2494–E2503
5. Tremblay, S. *et al.* (2014) Attentional filtering of visual information by neuronal ensembles in the primate lateral prefrontal cortex. *Neuron* 85, 202–215
6. Cohen, M.R. and Maunsell, J.H.R. (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* 12, 1594–1600
7. Logothetis, N.K. and Wandell, B.A. (2004) Interpreting the BOLD Signal. *Annu. Rev. Physiol.* 66, 735–769
8. Lippert, M.T. *et al.* (2010) Coupling of neural activity and fMRI-BOLD in the motion area MT. *Magn. Reson. Imaging* 28, 1087–1094
9. Fuster, J.M. and Alexander, G.E. (1971) Neuron activity related to short-term memory. *Science* 173, 652–654
10. Kubota, K. and Niki, H. (1971) Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J. Neurophysiol.* 34, 337–347
11. Jacobsen, C.F. and Nissen, H.W. (1937) Studies of cerebral function in primates. IV. The effects of frontal lobe lesions on the delayed alternation habit in monkeys. *J. Comp. Psychol.* 23, 101
12. Funahashi, S. *et al.* (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349
13. Funahashi, S. *et al.* (1993) Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature* 365, 753–756
14. Arnsten, A.F.T. (2013) The neurobiology of thought: the groundbreaking discoveries of Patricia Goldman-Rakic 1937–2003. *Cereb. Cortex* 23, 2269–2281
15. Wang, X.-J. (2001) Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* 24, 455–463
16. Curtis, C.E. and D'Esposito, M. (2003) Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423
17. Murray, J.D. *et al.* (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *PNAS* 114, 394–399
18. Barak, O. *et al.* (2010) Neuronal population coding of parametric working memory. *J. Neurosci.* 30, 9424–9430
19. Brody, C.D. *et al.* (2003) Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb. Cortex* 13, 1196–1207
20. Andersen, R.A. *et al.* (1987) Neurons of area 7 activated by both visual stimuli and oculomotor behavior. *Exp. Brain Res.* 67, 316–322
21. Fuster, J.M. and Jervey, J.P. (1981) Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* 212, 952–955
22. Markov, N.T. *et al.* (2014) A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb. Cortex* 24, 17–36
23. Tsujimoto, S. *et al.* (2010) Evaluating self-generated decisions in frontal pole cortex of monkeys. *Nat. Neurosci.* 13, 120–126
24. Cisek, P. and Kalaska, J.F. (2005) Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron* 45, 801–814
25. Romo, R. and Salinas, E. (2003) Cognitive neuroscience: flutter discrimination: neural codes, perception, memory and decision making. *Nat. Rev. Neurosci.* 4, 203–218
26. Hernandez, A. *et al.* (2010) Decoding a perceptual decision process across cortex. *Neuron* 66, 300–314
27. Rossi-Pool, R. *et al.* (2016) Emergence of an abstract categorical code enabling the discrimination of temporally structured tactile stimuli. *PNAS* 113, E7966–E7975
28. Dickersin, K. *et al.* (1987) Publication bias and clinical trials. *Control. Clin. Trials* 8, 343–353
29. Easterbrook, P.J. *et al.* (1991) Publication bias in clinical research. *Lancet* 337, 867–872
30. Button, K.S. *et al.* (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376
31. Supèr, H. *et al.* (2001) A neural correlate of working memory in the monkey primary visual cortex. *Science* 293, 120–124
32. van Kerkoerle, T. *et al.* (2017) Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat. Commun.* 8, 13804
33. McAdams, C.J. and Maunsell, J.H. (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* 19, 431–441
34. Luck, S.J. *et al.* (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77, 24–42
35. Fuster, J.M. (1990) Inferotemporal units in selective visual attention and short-term memory. *J. Neurophysiol.* 64, 681–697
36. Hayden, B.Y. and Gallant, J.L. (2013) Working memory and decision processes in visual area v4. *Front. Neurosci.* 7, 18
37. Ferrera, V.P. *et al.* (1994) Responses of neurons in the parietal and temporal visual pathways during a motion task. *J. Neurosci.* 14, 6171–6186
38. Chelazzi, L. *et al.* (2001) Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb. Cortex* 11, 761–772
39. Bisley, J.W. *et al.* (2004) Activity of neurons in cortical area MT during a memory for motion task. *J. Neurophysiol.* 91, 286–300
40. Mendoza-Halliday, D. *et al.* (2014) Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat. Neurosci.* 17, 1255–1262
41. Zaksas, D. and Pasternak, T. (2006) Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J. Neurosci.* 26, 11726–11742
42. Freedman, D.J. and Assad, J.A. (2006) Experience-dependent representation of visual categories in parietal cortex. *Nature* 443, 85–88
43. Zhou, X. *et al.* (2012) Neurons with inverted tuning during the delay periods of working memory tasks in the dorsal prefrontal and posterior parietal cortex. *J. Neurophysiol.* 108, 31–38
44. Kohn, A. and Movshon, J.A. (2003) Neuronal adaptation to visual motion in area MT of the macaque. *Neuron* 39, 681–691
45. Britten, K.H. (2008) Mechanisms of self-motion perception. *Annu. Rev. Neurosci.* 31, 389–410
46. Born, R.T. and Bradley, D.C. (2005) Structure and function of visual area MT. *Annu. Rev. Neurosci.* 28, 157–189
47. Treue, S. and Maunsell, J.H. (1999) Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J. Neurosci.* 19, 7591–7602

48. Gu, Y. *et al.* (2006) Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J. Neurosci.* 26, 73–85
49. Zhou, Y.D. and Fuster, J.M. (1996) Mnemonic neuronal activity in somatosensory cortex. *PNAS* 93, 10533–10537
50. Lemus, L. *et al.* (2010) Do sensory cortices process more than one sensory modality during perceptual judgments? *Neuron* 67, 335–348
51. Salinas, E. *et al.* (2000) Periodicity and riring rate as candidate neural codes for the frequency of vibrotactile stimuli. *J. Neurosci.* 20, 5503–5515
52. Camarillo, L. *et al.* (2012) Coding perceptual discrimination in the somatosensory thalamus. *PNAS* 109, 21093–21098
53. Lemus, L. *et al.* (2009) Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. *PNAS* 106, 9471–9476
54. Bigelow, J. *et al.* (2014) Neural correlates of short-term memory in primate auditory cortex. *Front. Neurosci.* 8, 250
55. Scott, B.H. *et al.* (2014) Neural correlates of auditory short-term memory in rostral superior temporal cortex. *Curr. Biol.* 24, 2767–2775
56. Hwang, J. and Romanski, L.M. (2015) Prefrontal neuronal responses during audiovisual mnemonic processing. *J. Neurosci.* 35, 960–971
57. Plakke, B. *et al.* (2013) Neural correlates of auditory recognition memory in primate lateral prefrontal cortex. *Neuroscience* 244, 62–76
58. Lemus, L. *et al.* (2009) Neural encoding of auditory discrimination in ventral premotor cortex. *PNAS* 106, 14640–14645
59. Vergara, J. *et al.* (2016) A neural parametric code for storing information of more than one sensory modality in working memory. *Neuron* 89, 54–62
60. Courtney, S.M. *et al.* (1998) An area specialized for spatial working memory in human frontal cortex. *Science* 279, 1347–1351
61. Sreenivasan, K.K. *et al.* (2014) Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* 18, 82–89
62. D'Esposito, M. and Postle, B.R. (2015) The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142
63. Constantinidis, C. and Wang, X.-J. (2004) A neural circuit basis for spatial working memory. *Neuroscientist* 10, 553–565
64. Harrison, S.A. and Tong, F. (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635
65. Serences, J.T. *et al.* (2009) Stimulus-specific delay activity in human primary visual cortex. *Psychol. Sci.* 20, 207–214
66. Emrich, S.M. *et al.* (2013) Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *J. Neurosci.* 33, 6516–6523
67. Riggall, A.C. and Postle, B.R. (2012) The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *J. Neurosci.* 32, 12990–12998
68. Ester, E.F. *et al.* (2015) Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron* 87, 893–905
69. Hubel, D.H. *et al.* (1977) Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique. *Nature* 269, 328–330
70. Tremblay, S. *et al.* (2015) Single-trial decoding of visual attention from local field potentials in the primate lateral prefrontal cortex is frequency-dependent. *J. Neurosci.* 35, 9038–9049
71. Freeman, J. *et al.* (2011) Orientation decoding depends on maps, not columns. *J. Neurosci.* 31, 4792–4804
72. Ester, E.F. *et al.* (2013) A neural measure of precision in visual working memory. *J. Cogn. Neurosci.* 25, 754–761
73. Silvanto, J. and Cattaneo, Z. (2010) Transcranial magnetic stimulation reveals the content of visual short-term memory in the visual cortex. *Neuroimage* 50, 1683–1689
74. Hamidi, M. *et al.* (2008) Evaluating frontal and parietal contributions to spatial working memory with repetitive transcranial magnetic stimulation. *Brain Res.* 1230, 202–210
75. Rose, N.S. *et al.* (2016) Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 354, 1136–1139
76. Wimmer, K. *et al.* (2014) Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* 17, 431–439
77. Murray, J.D. *et al.* (2014) A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* 17, 1661–1663
78. Wang, X.-J. (2009) A microcircuit model of prefrontal functions: Ying and Yang of reverberatory neurodynamics in cognition. In *The Frontal Lobes – Development, Function, and Pathology* (Risberg, J. and Grafman, J., eds), pp. 92–127, Cambridge University Press
79. Elston, G.N. (2007) Specialization of the neocortical pyramidal cell during primate evolution. In *Evolution of Nervous Systems* (1 st Edn) (Kaas, J.H., ed.), pp. 191–242, Elsevier
80. Goldman-Rakic, P.S. and Schwartz, M.L. (1982) Interdigitation of contralateral and ipsilateral columnar projections to frontal association cortex in primates. *Science* 216, 755–757
81. Katsuki, F. *et al.* (2014) Differences in intrinsic functional organization between dorsolateral prefrontal and posterior parietal cortex. *Cereb. Cortex* 24, 2334–2349
82. Constantinidis, C. and Klingberg, T. (2016) The neuroscience of working memory capacity and training. *Nat. Rev. Neurosci.* 17, 438–449
83. Wang, X.J. *et al.* (2004) Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory. *PNAS* 101, 1368–1373
84. Wang, X.J. (2013) The prefrontal cortex as a quintessential 'cognitive-type' neural circuit: working memory and decision making. In *Principles of Frontal Lobe Function* (2nd edn) (Stuss, D.T. and Knight, R.T., eds), pp. 226–248, Oxford University Press
85. Wang, B. *et al.* (2015) A subtype of Inhibitory interneuron with intrinsic persistent activity in human and monkey neocortex. *Cell Rep.* 10, 1450–1458
86. Gulyás, A.I. *et al.* (1996) Interneurons containing calretinin are specialized to control other interneurons in the rat hippocampus. *J. Neurosci.* 16, 3397–3411
87. Williams, G.V. and Goldman-Rakic, P.S. (1995) Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature* 376, 572–575
88. Mendoza-Halliday, D. *et al.* (2014) Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat. Neurosci.* 17, 1255–1262
89. Rosenkilde, C.E. *et al.* (1981) Single cell activity in ventral prefrontal cortex of behaving monkeys. *Brain Res.* 209, 375–394
90. Hikosaka, K. and Watanabe, M. (2000) Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. *Cereb. Cortex* 10, 263–271
91. Kennerley, S.W. and Wallis, J.D. (2009) Encoding of reward and space during a working memory task in the orbitofrontal cortex and anterior cingulate sulcus. *J. Neurophysiol.* 102, 3352–3364
92. Cai, X. and Padoa-Schioppa, C. (2014) Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. *Neuron* 81, 1140–1151
93. Tremblay, L. and Schultz, W. (1999) Relative reward preference in primate orbitofrontal cortex. *Nature* 398, 704–708
94. Freedman, D.J. *et al.* (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.* 23, 5235–5246
95. Romanski, L.M. and Goldman-Rakic, P.S. (2002) An auditory domain in primate prefrontal cortex. *Nat. Neurosci.* 5, 15–16
96. Wallis, J.D. *et al.* (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411, 953–956
97. Freedman, D.J. *et al.* (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316
98. Akkal, D. *et al.* (2002) Comparison of neuronal activity in the rostral supplementary and cingulate motor areas during a task

with cognitive and motor demands. *Eur. J. Neurosci.* 15, 887–904

99. Isomura, Y. *et al.* (2003) Neural coding of 'attention for action' and 'response selection' in primate anterior cingulate cortex. *J. Neurosci.* 23, 8002–8012
100. Procyk, E. and Joseph, J.P. (2001) Characterization of serial order encoding in the monkey anterior cingulate sulcus. *Eur. J. Neurosci.* 14, 1041–1046
101. Niki, H. and Watanabe, M. (1976) Prefrontal unit activity and delayed response: relation to cue location versus direction of response. *Brain Res.* 105, 79–88
102. Bernacchia, A. *et al.* (2011) A reservoir of time constants for memory traces in cortical neurons. *Nat. Neurosci.* 14, 366–372
103. Romo, R. *et al.* (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399, 470–473
104. Miller, E.K. *et al.* (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* 16, 5154–5167
105. Markowitz, D.A. *et al.* (2015) Multiple component networks support working memory in prefrontal cortex. *PNAS* 112, 11084–11089
106. Wang, L. *et al.* (2015) Differential roles of delay-period neural activity in the monkey dorsolateral prefrontal cortex in visual-haptic crossmodal working memory. *PNAS* 112, E214–9
107. Crowe, D.A. *et al.* (2013) Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nat. Neurosci.* 16, 1484–1491
108. Katsuki, F. and Constantinidis, C. (2013) Time course of functional connectivity in primate dorsolateral prefrontal and posterior parietal cortex during working memory. *PLoS One* 8, e81601
109. Qi, X.-L. and Constantinidis, C. (2012) Correlated discharges in the primate prefrontal cortex before and after working memory training. *Eur. J. Neurosci.* 36, 3538–3548
110. Zhou, X. *et al.* (2011) Cholinergic modulation of working memory activity in primate prefrontal cortex. *J. Neurophysiol.* 106, 2180–2188
111. Qi, X.-L. *et al.* (2010) Comparison of neural activity related to working memory in primate dorsolateral prefrontal and posterior parietal cortex. *Front. Syst. Neurosci.* 4, 12
112. Meyer, T. *et al.* (2007) Persistent discharges in the prefrontal cortex of monkeys naive to working memory tasks. *Cereb. Cortex* 17, i70–i76
113. Lebedev, M.A. *et al.* (2004) Representation of attended versus remembered locations in prefrontal cortex. *PLoS Biol.* 2, e365
114. Constantinidis, C. *et al.* (2001) Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. *J. Neurosci.* 21, 3646–3655
115. Constantinidis, C. *et al.* (2001) The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nat. Neurosci.* 4, 311–316
116. Batuev, A.S. *et al.* (1985) Comparative characteristics of unit activity in the prefrontal and parietal areas during delayed performance in monkeys. *Behav. Brain Res.* 16, 57–70
117. Kojima, S. and Goldman-Rakic, P.S. (1982) Delay-related activity of prefrontal neurons in rhesus monkeys performing delayed response. *Brain Res.* 248, 43–49
118. Bodner, M. *et al.* (1996) Auditory memory cells in dorsolateral prefrontal cortex. *Neuroreport* 7, 1905
119. Koch, K.W. and Fuster, J.M. (1989) Unit activity in monkey parietal cortex related to haptic perception and temporary memory. *Exp. Brain Res.* 76, 292–306
120. Constantinidis, C. and Steinmetz, M.A. (1996) Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *J. Neurophysiol.* 76, 1352–1355
121. Crowe, D.A. *et al.* (2010) Rapid sequences of population activity patterns dynamically encode task-critical spatial information in parietal cortex. *J. Neurosci.* 30, 11640–11653
122. Chafee, M.V. and Goldman-Rakic, P.S. (2000) Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *J. Neurophysiol.* 83, 1550–1566
123. Fuster, J.M. *et al.* (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 405, 347–351
124. Zhou, X. *et al.* (2016) Neural correlates of working memory development in adolescent primates. *Nat. Commun.* 7, 13423
125. Rainer, G. *et al.* (1998) Memory fields of neurons in the primate prefrontal cortex. *PNAS* 95, 15008–15013
126. Rainer, G. *et al.* (1998) Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* 393, 577–579
127. Chafee, M.V. and Goldman-Rakic, P.S. (1998) Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.* 79, 2919–2940
128. Carlson, S. *et al.* (1997) Dissociation of mnemonic coding and other functional neuronal processing in the monkey prefrontal cortex. *J. Neurophysiol.* 77, 761–774
129. Armstrong, K.M. *et al.* (2009) Selection and maintenance of spatial information by frontal eye field neurons. *J. Neurosci.* 29, 15621–15629
130. Lara, A.H. and Wallis, J.D. (2014) Executive control processes underlying multi-item working memory. *Nat. Neurosci.* 17, 876–883
131. Takeda, K. and Funahashi, S. (2002) Prefrontal task-related activity representing visual cue location or saccade direction in spatial working memory tasks. *J. Neurophysiol.* 87, 567–588
132. Meyer, T. *et al.* (2011) Stimulus selectivity in dorsal and ventral prefrontal cortex after training in working memory tasks. *J. Neurosci.* 31, 6266–6276
133. di Pellegrino, G. and Wise, S.P. (1993) Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate. *J. Neurosci.* 13, 1227–1243
134. Jacob, S.N. and Nieder, A. (2014) Complementary roles for primate frontal and parietal cortex in guarding working memory from distractor stimuli. *Neuron* 83, 226–237
135. Gottlieb, Y. *et al.* (1989) Single unit activity in the auditory cortex of a monkey performing a short term memory task. *Exp. Brain Res.* 74, 139–148
136. Suzuki, W.A. *et al.* (1997) Object and place memory in the macaque entorhinal cortex. *J. Neurophysiol.* 78, 1062–1081
137. Hernandez, A. *et al.* (2002) Temporal evolution of a decision-making process in medial premotor cortex. *Neuron* 33, 959–972
138. Lemus, L. *et al.* (2007) Neural correlates of a postponed decision report. *PNAS* 104, 17174–17179
139. Graziano, M.S. *et al.* (1997) Coding the locations of objects in the dark. *Science* 277, 239–241
140. Romo, R. *et al.* (2004) Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron* 41, 165–173
141. Russo, G.S. and Bruce, C.J. (1996) Neurons in the supplementary eye field of rhesus monkeys code visual targets and saccadic eye movements in an oculocentric coordinate system. *J. Neurophysiol.* 76, 825–848
142. Olson, C.R. and Tremblay, L. (2000) Macaque supplementary eye field neurons encode object-centered locations relative to both continuous and discontinuous objects. *J. Neurophysiol.* 83, 2392–2411
143. Pesaran, B. *et al.* (2002) Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* 5, 805–811
144. Powell, K.D. and Goldberg, M.E. (2000) Response of neurons in the lateral intraparietal area to a distractor flashed during the delay period of a memory-guided saccade. *J. Neurophysiol.* 84, 301–310
145. Colby, C.L. *et al.* (1996) Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *J. Neurophysiol.* 76, 2841–2852
146. Gnadt, J.W. and Andersen, R.A. (1988) Memory related motor planning activity in posterior parietal cortex of macaque. *Exp. Brain Res.* 70, 216–220
147. Seo, H. *et al.* (2009) Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. *J. Neurosci.* 29, 7278–7289

148. Miyashita, Y. and Chang, H.S. (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* 331, 68–70

149. Cahusac, P.M. *et al.* (1989) Responses of hippocampal formation neurons in the monkey related to delayed spatial response and object–place memory tasks. *Behav. Brain Res.* 33, 229–240

150. Watanabe, T. and Niki, H. (1985) Hippocampal unit activity and delayed response in the monkey. *Brain Res.* 325, 241–254

151. Ng, C.-W. *et al.* (2014) Neural correlates of auditory recognition memory in the primate dorsal temporal pole. *J. Neurophysiol.* 111, 455–469

152. Chelazzi, L. *et al.* (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiol.* 80, 2918–2940

153. Miller, E.K. and Desimone, R. (1994) Parallel neuronal mechanisms for short-term memory. *Science* 263, 520–522

154. Miller, E.K. *et al.* (1991) A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* 254, 1377–1379

155. Baylis, G.C. and Rolls, E.T. (1987) Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. *Exp. Brain Res.* 65, 614–622

156. Mikami, A. and Kubota, K. (1980) Inferotemporal neuron activities and color discrimination with delay. *Brain Res.* 182, 65–78

157. Woloszyn, L. and Sheinberg, D.L. (2009) Neural dynamics in inferior temporal cortex during a visual working memory task. *J. Neurosci.* 29, 5494–5507

158. Takaura, K. *et al.* (2011) Neural substrate of spatial memory in the superior colliculus after damage to the primary visual cortex. *J. Neurosci.* 31, 4233–4241

159. Watanabe, Y. and Funahashi, S. (2004) Neuronal activity throughout the primate mediodorsal nucleus of the thalamus during oculomotor delayed-responses. I. Cue-, delay-, and response-period activity. *J. Neurophysiol.* 92, 1738–1755

160. Hikosaka, O. *et al.* (1989) Functional properties of monkey caudate neurons. I. Activities related to saccadic eye movements. *J. Neurophysiol.* 61, 780–798