

# Revisiting the role of persistent neural activity during working memory

Kartik K. Sreenivasan<sup>1</sup>, Clayton E. Curtis<sup>2</sup>, and Mark D'Esposito<sup>3</sup>

<sup>1</sup> Division of Science and Mathematics, New York University Abu Dhabi, 19 Washington Square North, New York, NY 10011, USA

<sup>2</sup> Department of Psychology, and Center for Neural Science, New York University, 6 Washington Place, New York, NY 10003, USA

<sup>3</sup> Helen Wills Neuroscience Institute, and Department of Psychology, University of California, Berkeley, 132 Barker Hall, Berkeley, CA 94720, USA

**What are the neural mechanisms underlying working memory (WM)? One influential theory posits that neurons in the lateral prefrontal cortex (LPFC) store WM information via persistent activity. In this review, we critically evaluate recent findings that together indicate that this model of WM needs revision. We argue that sensory cortex, not the LPFC, maintains high-fidelity representations of WM content. By contrast, the LPFC simultaneously maintains representations of multiple goal-related variables that serve to bias stimulus-specific activity in sensory regions. This work highlights multiple neural mechanisms supporting WM, including temporally dynamic population coding in addition to persistent activity. These new insights focus the question on understanding how the mechanisms that underlie WM are related, interact, and are coordinated in the LPFC and sensory cortex.**

## Introduction

WM comprises the set of operations that support the active retention of behaviorally relevant information over brief intervals. Given the central role of WM in goal-directed behavior, establishing the neural basis of WM has been a priority of neuroscience research. Early WM studies observed that selective increases in neural activity during the presentation of a to-be-maintained sample item persisted throughout the blank 'delay' interval of a WM delay task, bridging the temporal gap between the sample and the subsequent contingent response [1,2]. This work inspired the theoretical framework that has predominated in the field: neurons or neuronal populations that are selectively tuned to the to-be-remembered information hold this information in an active state through persistent activation [3]. We refer to this model, which emphasizes stable persistent neural activity (see [Glossary](#)) in selective neurons as the fixed-selectivity model. Motivated by this model, functional MRI (fMRI) studies in humans and

electrophysiological studies in monkeys have consistently identified persistent neural activity in the LPFC, leading many to conclude that the LPFC stores representations of WM memoranda.

A decade ago, we provided a critique of the literature on persistent activity in the context of contemporary models of prefrontal cortical function [4]. We hypothesized that, in contrast to existing theories of WM, persistent LPFC activity signifies attention directed to internal representations maintained in sensory cortices. Viewed through the lens of the fixed-selectivity model, evidence for this proposal is limited. Studies of sensory and motor function, however, suggest that information is likely to be represented through the combined activity of neural populations with diverse tuning properties rather than individual highly-tuned neurons [5,6]. This notion offers a promising framework for understanding WM.

In recent years, analytic and methodological advances ([Box 1](#)) have expanded researchers' ability to capture the multivariate nature of population coding and the causal relationships between neural activity and behavior. The findings generated using these approaches underscore the need for a revision of existing views of WM. In light of these results, we revisit the issue of how information remains active during WM. The studies we discuss here focus on

## Glossary

**Delay tasks:** the experimental paradigm typically used to study the neural basis of working memory (WM). A trial in a delay task begins with a brief presentation of a sample item. The subject encodes this item into WM and maintains this item over a blank 'delay' period of a few to several seconds. At the end of the delay period, a probe stimulus appears and the subject initiates a behavioral response contingent on the WM representation of the sample item. A key feature of delay tasks is that they temporally segregate subcomponents of WM such as stimulus encoding, storage, and retrieval/response.

**Persistent neural activity:** above-baseline neural activity that remains stable and elevated during a trial of a delay task. Persistent neural activity begins during the sample presentation and persists throughout the delay period, returning to baseline at the end of the trial. According to the fixed-selectivity model of WM (see main text), persistent neural activity in neurons selective for WM memoranda is the mechanism by which WM information is actively maintained.

**TMS:** transcranial magnetic stimulation.

**Voxel:** the spatial unit for measuring changes in blood-oxygenation-level dependent (BOLD) signal with fMRI. A voxel is a 3D volumetric pixel, typically of the order of 3 mm<sup>3</sup>. BOLD signal within a voxel is an indirect measure of the summed activity of many tens of thousands of neurons. A single whole-brain fMRI image can comprise 60,000–100,000 voxels.

**Working memory:** the set of operations that support the ability to maintain information in an active state, to manipulate that information, and to use that information to guide behavior. WM is essential for several aspects of

Corresponding author: Sreenivasan, K.K. ([kartik.sreenivasan@nyu.edu](mailto:kartik.sreenivasan@nyu.edu)).

Keywords: working memory; prefrontal cortex; top down; decoding; forward encoding; MVP; fMRI.

1364-6613/\$ – see front matter

© 2014 Elsevier Ltd. All rights reserved. <http://dx.doi.org/10.1016/j.tics.2013.12.001>

### Box 1. Methodological advances

Here we briefly describe analytic and methodological advances that have furthered our understanding of the neural basis of WM. The reader is encouraged to seek out some of the excellent reviews on these approaches (referenced below) for more details.

#### *Decoding analysis*

Unlike standard univariate analyses, which independently examine data from individual neurons or voxels for differences across conditions, multivariate decoding methods consider data from several neurons or voxels at once to identify patterns of activity that encode task-related information [10–13]. This technique uses machine learning algorithms to decode, or categorize, unlabeled test data using labeled training data. Successful (above-chance) decoding signifies that the activity pattern entered into the algorithm differs between the categories of interest, implying that the underlying neural activity encoded information about these categories. The chief advantage of this approach is potentially increased sensitivity [78]. However, patterns of neural activity may reliably distinguish between conditions for various reasons, some of which are not anticipated by the experimental design [18,79]. Caution is therefore required when interpreting the nature of the information identified via decoding analysis [80].

#### *Forwarding encoding models*

Conversely, encoding models predict fMRI activity from task conditions [81–83]. These models rely on *a priori* assumptions about the features of task conditions that will result in changes in the hemodynamic response. In WM studies, forward encoding models of visual cortex have been constructed using knowledge about tuning for visual features [17]. Neural activity in hypothetical populations of neurons (channels) tuned to different values in feature space can be reconstructed from training data by estimating the degree to which each voxel's response contributes to a given channel. The critical advantage over decoding analyses is that this approach can predict fMRI responses to novel stimuli [14]. Encoding approaches are potentially more powerful for identifying information encoded in neural activity, but are constrained by the validity of the underlying assumptions of the model.

#### *TMS*

TMS uses magnetic fields to focally modulate cortical excitability [84]. In WM studies, TMS is used either offline to modulate cortical function for the duration of the experiment or online to modulate activity during specific epochs of a task. TMS effects on behavior or neural activity in distal regions can support strong causal inferences about the functional role of the regions targeted with TMS. Attenuation of TMS effects as a function of distance from the coil imposes restrictions on which brain regions can be targeted.

visual WM, but the general principles discussed herein apply to WM in other modalities.

### Evidence for persistent WM representations in visual cortex

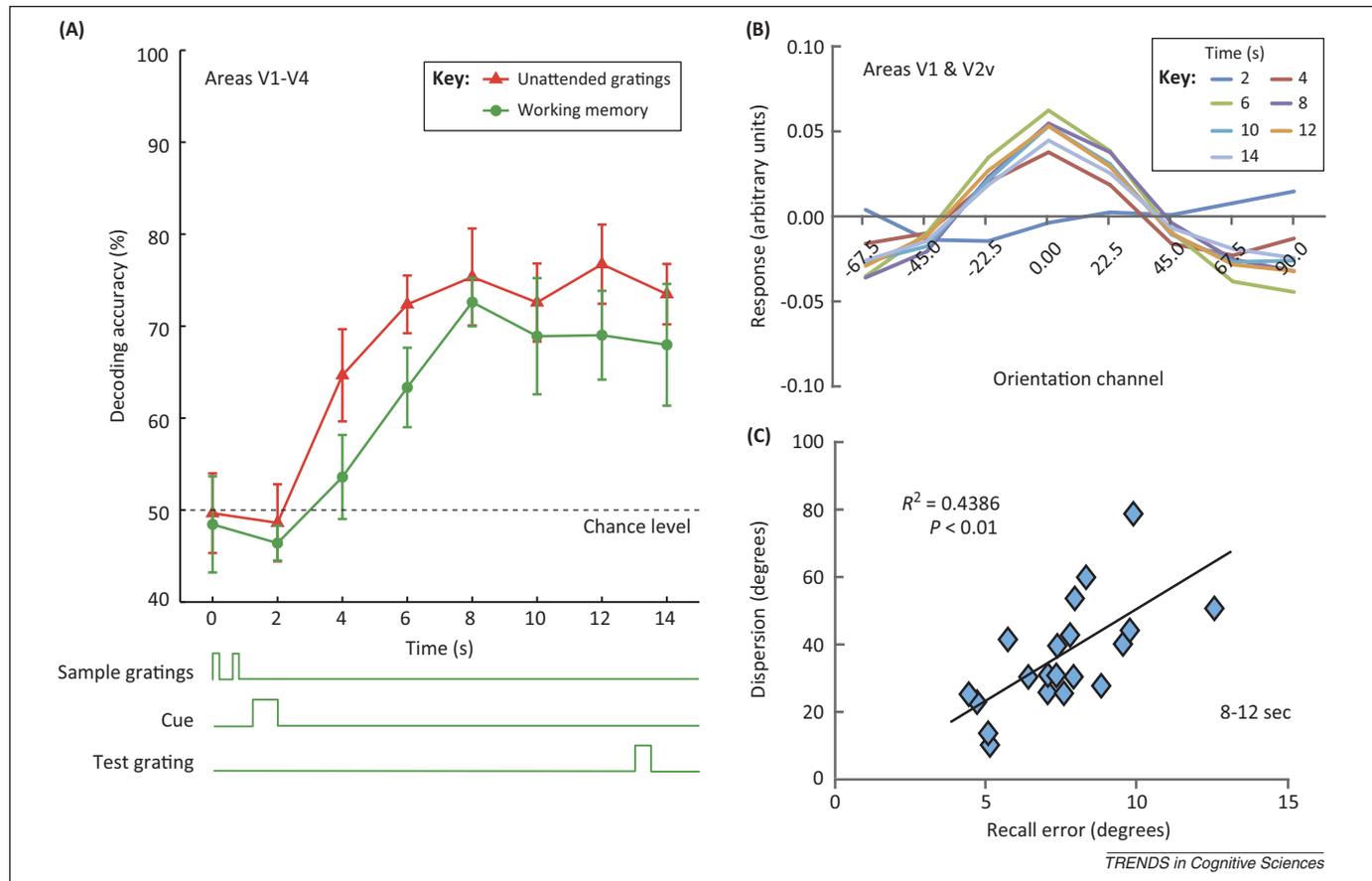
Neurons in visual cortex are selectively tuned to visual stimulus features and are consequently well suited for maintaining high-fidelity representations of visual information in the service of WM [7]. Yet, from the perspective of the fixed-selectivity model, evidence for sustained WM representations in visual cortex has been equivocal. Although sustained responses have been observed in temporal cortex [8], studies typically describe transient neural responses to sample stimuli without any subsequent sustained activation. Studies of early visual regions routinely note an absence of persistent activity [9].

Contemporary multivariate encoding and decoding statistical analyses (Box 1), however, consistently demonstrate that visual cortex does retain sensory WM

representations. Decoding analysis applied to fMRI or electrophysiological data can identify activity distributed across neurons or neural populations that encodes task-relevant information [10–13]. By contrast, forward encoding models take advantage of assumptions about neural population tuning to reconstruct the response of hypothetical channels from fMRI voxels that represent the weighted sum of subpopulations of neurons tuned to these channels [14]. Both approaches can test whether feature or item information is encoded in the multivariate patterns of activity during WM, regardless of whether this activity exhibits sustained stimulus-selective responses during sample presentation that persist across the blank delay interval of the WM task. Studies incorporating these methods find that patterns of delay period activity in early visual cortex contain information about simple visual features held in WM [15–19] (Figure 1A,B). Similarly, delay patterns in occipital and temporal regions specialized for object representation encode actively maintained visual objects [20–24], consistent with studies that inferred a role for temporal cortex in WM storage on the basis of persistent neural activity in these regions [8].

Moreover, this work establishes four key properties of population coding of WM information in visual cortex. First, decoding and forward encoding analyses have extracted information specific to the contents of WM from visual cortex activity across multiple timepoints during the delay period [15,17,24,25], indicating that visual cortical WM representations persist throughout the period separating the visual stimulus and the contingent behavioral response. Second, given the limited capacity of WM [26], neural coding of sensory representations should prioritize task-relevant over task-irrelevant information [27]. Selectivity for task-relevant information was illustrated in a study where decoding based on the multivoxel pattern of delay period activity in early visual cortex was successful only for the task-relevant feature (orientation or color) of the memoranda [16]. Similarly, other work has shown that multivoxel patterns of delay period activity encode only items cued in the sample display as task relevant [15,24]. Third, and in contrast to the fixed-selectivity model, information about items maintained in WM can be encoded by neural populations that are not highly selective for the maintained stimuli. A recent fMRI study examined the degree to which decoding information about items maintained in WM was dependent on voxels that were highly selective for the WM items. The key finding was that removing highly selective voxels from the analysis did not substantially reduce the ability to decode information about the WM items [24]. These results are in line with studies demonstrating that perceptual [28] and motor [5] information is distributed across neural populations with diverse tuning preferences.

Fourth, and perhaps most importantly, fMRI measures of sensory representations in visual cortex are tied to the precision of WM representations. For example, one study found decreases in the ability to decode maintained directions of motion from multivoxel delay period activity as the number of to-be-maintained motion directions increased, possibly due to interference between spatially overlapping representations [29]. Reductions in the ability to decode the contents of WM predicted decrements in the precision



**Figure 1.** Evidence for sustained working memory (WM) representations in the early visual cortex. Subjects maintained an orientation of a grating over a delay period. **(A)** A decoding analysis demonstrated successful decoding of the cued orientation throughout the delay period. Decoding accuracy for maintained orientation (green circles) was comparable to decoding accuracy for the orientation of visually presented gratings (red triangles). Adapted from [15]. **(B)** A forward encoding analysis found tuning for maintained orientation, indicating that information about the maintained orientation was present during the delay. As in (A), this information was preserved throughout the delay period (different color channel response functions represent tuning at different points during the trial; error bars have been omitted for clarity). **(C)** The relationship between the degree of tuning of the channel response functions (estimated as dispersion of the best-fit Gaussian), which was taken as a proxy for the precision of the WM representation, and behavioral accuracy in reconstructing the maintained orientation. Greater tuning (i.e., less dispersion in the channel response functions) predicted increased accuracy across subjects, establishing a relationship between the quality of WM information stored in visual cortex and memory precision. Adapted from [17].

of memorized motion direction in individual subjects. Forward encoding estimates of WM representations also correspond to memory precision. Ester and colleagues [17] estimated population tuning curves for maintained orientations from visual cortical activity and used the width of these tuning curves as an inverse proxy for tuning for the maintained orientation. Tuning precision was correlated with subjects' behavioral precision in reconstructing the remembered orientation (Figure 1C). Although compelling, these studies do not indicate whether the observed modulation of memory precision is a direct or an indirect consequence of neural activity in sensory cortex. Fortunately, transcranial magnetic stimulation (TMS) (Box 1) experiments can provide causal evidence for the role of sensory activity in maintaining high-fidelity WM representations. TMS applied to early visual cortex during short-term retention of visual stimuli results in a reduction in behavioral measures of WM precision [30–32]. Together, this work supports the notion that visual cortex stores precise representations of visual WM contents.

### The role of the IPFC in WM

The most pervasive observation in the WM literature is that IPFC activity persists throughout WM maintenance.

This finding has been interpreted as evidence that IPFC delay activity encodes sensory features of WM items [3]. However, in addition to displaying coarse selectivity for WM items and features [33], IPFC activity exhibits selectivity for a broad range of task variables during the delay period of WM tasks. For example, IPFC neurons show differential preferences for task rules [34], contingent motor responses [35], and stimulus–response mappings [36]. Studies examining population coding of IPFC delay activity have similarly found information about stimuli [37], rules [25], and object categories [38] throughout the delay.

### Not-so-selective selectivity

How is the IPFC simultaneously selective for multiple task variables? An increasing number of studies find that, rather than utilizing distinct populations to encode each task variable, activity in the IPFC encodes multiple task variables within a single population of neurons [37,39,40]. For example, Machens and colleagues demonstrated that individual IPFC neurons responded to combinations of two task variables (maintained stimulus identity and elapsed time), but that information about each task variable could be independently extracted from the population code [41]. This finding suggests that IPFC representations can be

high dimensional, because they simultaneously encode multiple pieces of information that can be interpreted by neural populations in hierarchically lower regions according to their functional relevance.

Recent work by Rigotti and colleagues further characterized high-dimensional representations of multiple task variables in the IPFC [42,43]. They demonstrated that activity in a population of IPFC neurons simultaneously encoded information about all task variables (task, object 1, and object 2) during the delay period of a complex object-sequence WM task. To examine the degree to which these high-dimensional representations arose from neurons' selectivity for individual task variables (e.g., task A vs task B), the authors artificially abolished classical neuronal selectivity. This was accomplished by adding noise to each neuron's response to equate the average response across single task variables (e.g., such that the average response to tasks A and B were equivalent), while preserving differences in firing as a result of different combinations of task variables. Strikingly, and counter to the predictions of the fixed-selectivity model, population selectivity for task variables persisted even when selectivity for individual task variables was abolished. Most critically, this work highlights two conditions that give rise to high-dimensional IPFC representations. The first condition is nonlinear mixed selectivity, or nonlinear neural responses to combinations of task variables. Artificially abolishing nonlinear mixed selectivity reduced the dimensionality of the representations, limiting the information that was available for readout by other regions, and error trials were associated with reduced nonlinear mixed selectivity. The second condition is randomly connected networks of neurons, which can support complex and diverse input-output mappings [44]. The perspective presented in this intriguing study has the potential to demystify the seemingly limitless ability of the IPFC to represent task information; however, an important future direction will be to specify the constraints (such as anatomical input from sensorimotor regions [45]) that limit the dimensionality of information encoded by IPFC activity.

#### *Interpreting stimulus selectivity*

How do we interpret evidence that IPFC activity is selective for maintained stimuli? Does this imply that the IPFC maintains sensory properties of the WM stimulus [33], analogous to sensory cortices? Duplicating the sustained sensory information in sensory cortices seems unnecessary and metabolically costly [46]. Furthermore, from the perspective of the fixed-selectivity model, the maintenance of sensory representations in the IPFC would seem to require that IPFC neurons exhibit a degree of tuning for sensory features comparable with that of neurons in sensory cortex. This does not appear to be the case. For example, neurons in the dorsal IPFC exhibit preferences for task-relevant directions of motion, but are markedly less motion selective than neurons in the motion-sensitive middle temporal (MT) visual area [47].

The results from population decoding studies also support a dissociation between stimulus-specific IPFC representations and sensory representations. In one study,

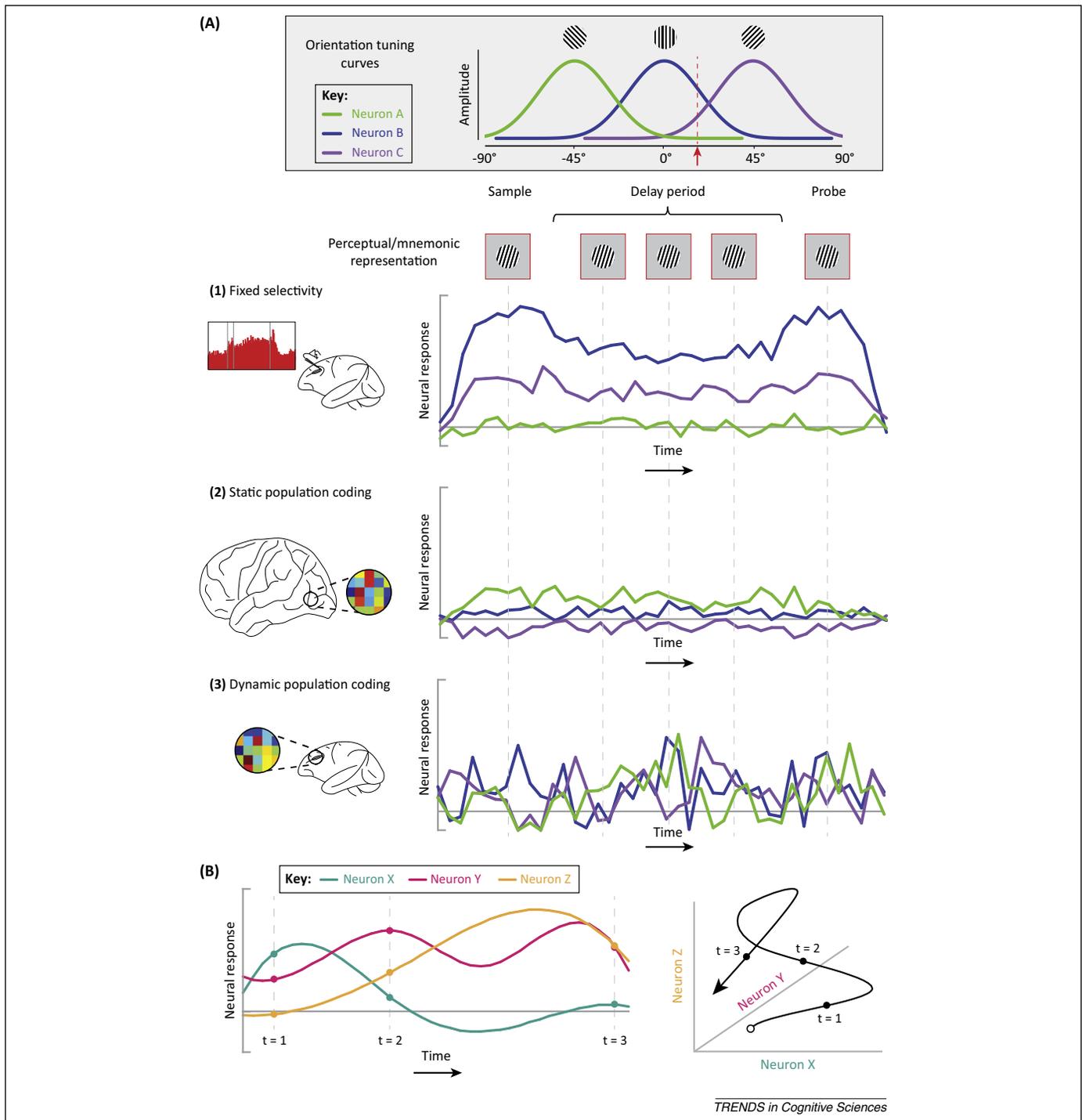
researchers were able to decode the identity of the memoranda from delay period activation patterns in visual cortex – but not the IPFC – when subjects were required to maintain a visual representation [48]. The pivotal finding was that stimulus-specific activation patterns emerged in the IPFC when subjects retained the category of the item, rather than the item itself. This finding accords well with prior work showing that the IPFC preferentially encodes and maintains arbitrary and abstract representations of object category over representations of visual similarity [38,49–51]. Further support for the distinction between stimulus-selective IPFC representations and sensory representations comes from a second fMRI study [24]. This study demonstrated that, in contrast to visual cortex, where category-selective patterns of delay period activity were more similar for visually similar categories, category-selective IPFC activity patterns were not sensitive to visual similarity, suggesting that the representations encoded by these patterns were categorical but not sensory in nature.

#### *Source of top-down input*

Although the above findings are compatible with the well-described role of the IPFC in storing and integrating information, an alternate perspective is that, rather than representing the storage of information, the primary role of the IPFC during WM is to influence representations in other regions [52]. Specifically, patterns of activity in the IPFC may serve as top-down signals used to bias the competition of neural representation in hierarchically lower areas, such as sensorimotor [4] and parietal regions [53]. Indeed, the IPFC sits at the apex of the motor hierarchy [54] and possesses diverse anatomical connections capable of directly influencing a variety of regions [45]. We propose that activity in the IPFC is a likely source of top-down input to visual systems during visual WM.

A causal link between prefrontal activity and the properties of visual cortical neurons has been established through electrical microstimulation [55], pharmacological manipulations [56], and intervention with TMS [57]. Until recently, however, there was little direct evidence that IPFC input modulated visual activity during WM. Studies combining TMS and measures of neural activity such as fMRI and electroencephalography (EEG) provide such a link. TMS administered to focal regions of the IPFC causes a significant reduction in the selectivity of fMRI responses in visual cortex, suggesting that IPFC inputs enhance selectivity in visual cortex during WM [58,59]. Comparable results were found in patients with stroke-induced lesions to the IPFC: the selectivity of responses was reduced in ipsilesional relative to contralesional visual cortex [58].

Although highly informative, these experiments did not establish whether IPFC input phasically influenced visual cortical selectivity during a specific stage of the WM task or whether IPFC input operated at multiple stages during WM to preserve visual selectivity throughout WM maintenance. Two recent papers described TMS effects on selectivity during stimulus encoding and maintenance. One group found that TMS applied over the caudal IPFC caused reduced selectivity for the sample stimuli, as indicated by enhanced EEG responses to task-irrelevant sample stimuli as well as marginally suppressed EEG



**Figure 2.** Multiple neural mechanisms of working memory (WM). **(A)** A simplified schematic comparing and contrasting the fixed-selectivity model with population coding models involving static and dynamic temporal codes. Orientation tuning curves for three hypothetical neurons, A, B, and C, are shown in the top inset; neurons A–C are tuned to  $-45^\circ$ ,  $0^\circ$ , and  $45^\circ$ , respectively. Neural responses to the to-be-maintained orientation ( $17^\circ$ ) are indicated by the red arrow and broken line in the inset. The perceptual and mnemonic representations are depicted below the inset. The visually presented sample orientation (oriented at  $17^\circ$ ) elicits a perceptual representation and is followed by a delay period. The mnemonic representation of the sample orientation persists across the delay period and is followed by a probe stimulus that elicits a perceptual representation. Below this are schematics for three different potential neural models of WM. Note that the timecourses shown here are for illustrative purposes only and do not depict actual data. *Top row:* The fixed-selectivity model, primarily derived from single-unit recordings in monkey lateral prefrontal cortex (IPFC), predicts that neuron B, which is selective for the maintained orientation (refer to the tuning curves in the inset), will exhibit persistent, sustained activity. Neuron C, which is less selective for the maintained orientation, exhibits persistent activity to a lesser degree, while neuron A's activity remains at baseline. *Middle row:* Evidence for static population coding comes primarily from fMRI decoding and forward encoding studies of visual cortex. Here the pattern of activity across neurons can encode stimulus orientation in the absence of highly selective neural responses. This pattern is sustained throughout maintenance. *Bottom row:* Dynamic population coding has been demonstrated largely in monkey IPFC. Despite time-varying activity in all three neurons, the representation of orientation remains stable. The relevant orientation is encoded by a different combination of neural responses at each point in time. Note that each of these models is potentially compatible with the notion of mixed selectivity, where activity within a single neuron or neuronal population can be simultaneously selective for multiple goal-related variables. Portions of this figure are adapted from [2]. **(B)** Schematic illustrating dynamic population coding in a hypothetical population comprising neurons X, Y, and Z. *Left:* The timecourses for neurons X–Z. *Right:* The population response can be depicted as a trajectory through multidimensional state space. The black path represents the combined activity of the three neurons across time. Timepoints  $t=1$ ,  $t=2$ , and  $t=3$  in the trajectory on the right correspond to activity at points  $t=1$ ,  $t=2$ , and  $t=3$  (indicated by the broken lines) in the timecourses on the left.

responses to task-relevant sample stimuli [60]. Another group used concurrent TMS–fMRI [61] to modulate IPFC activity during the delay period of a WM task [62]. They observed enhanced processing in visual cortex of task-relevant information in the presence of distractors, further confirming that IPFC input modulates visual selectivity during WM maintenance.

Taken together, these studies provide direct evidence that top-down IPFC signals modulate sensory activity during WM and that IPFC feedback signals, whether periodic or tonic, may enhance the selectivity of representations in sensory cortex throughout WM maintenance. This idea is in line with results showing that consistent communication between the IPFC and visual cortex across stimulus encoding and maintenance is important for WM [63]. It should be noted that although none of the TMS studies described above drew specific links between top-down IPFC input and persistent activity within the IPFC, they all targeted subregions of the IPFC that typically exhibit sustained activation during WM tasks. Further work is required to explicitly test the relationship between top-down IPFC signals and persistent activity.

### Persistent neural activity revisited

Persistent neural activity, particularly in the IPFC, has become synonymous with WM. However, this equivalence is misleading. First, the IPFC does not appear to be privileged in its ability to generate persistent activity. Particularly when analyses focus on neurons or voxels that are highly stimulus selective, persistent neural activity can be observed nearly everywhere in the brain [8,64–66]. Second, although persistent neural activity is a key mechanism for forming temporal links between sensation and action [54] in various contexts [67–69], there may be other mechanisms by which WM information is actively retained (Figure 2A).

The evidence we have presented thus far underscores the importance of population coding of WM information [40]. As demonstrated in other domains, information can be represented by dynamic spatiotemporal patterns among populations of neurons [70,71]. That is, information encoded in a given state – or set of activation levels across the population – at time  $t$  may be encoded in an entirely different state at time  $t + 1$ . Thus, a representation can be sustained via a dynamic trajectory through state space (where the space is defined by the activity of the neurons in the population) [72] (Figure 2B). Surprisingly, this is not antithetical to the notion of a stable representation; theoretical work suggests that in the case where redundant coding of information is present, time-varying population activation can encode a stable representation [73].

Empirical findings confirm that WM information can be sustained through dynamic population coding: by examining data in short time steps across the delay period, studies have found that the pattern of activation states encoding WM information changes over the course of maintenance. One study decoded electrophysiological data from monkey IPFC during a WM task for object category and showed that decoding was most successful when the decoding algorithm was trained and tested on data from the same point in the trial [38]. Training and testing from temporally

### Box 2. Areas for future research

- A key feature of WM is the ability to maintain task-relevant information in the face of task-irrelevant or distracting information [27,85]. A complete understanding of how WM is implemented in the brain will require new insights into how sustained representations in sensory cortex are affected by incoming sensory input. One study observed that visual WM recruits populations with spatial receptive fields that are distinct from the receptive fields of the neurons activated by the sample item [86]. Although the authors did not examine interference from irrelevant information, some form of global recruitment could potentially prevent distracting items presented in the same spatial location as the sample item from corrupting the WM representation.
- With the studies we review here, we argue that WM relies in part on encoding that is temporally dynamic and spatially distributed. One proposal is that temporal and spatial coordination arises through brain oscillations [87–89]. Direct intracranial recordings in human epilepsy patients [90] and the use of subdural electrode arrays to directly record neural activity across broad regions of cortex in monkeys [91] offer an opportunity to describe the properties of these oscillations in the context of WM [92], simultaneously marrying temporal precision and spatial coverage. These techniques can provide greater insight into regional specificity and interregional integration that are critical for elucidating the neural basis of WM. Additionally, these techniques can facilitate more direct comparisons across monkey and human studies.
- A logical extension of the notion of distributed coding within a region is the idea that brain function arises from large-scale network coordination [93]. There is evidence to suggest that WM relies on such large-scale interregional communication. Visual WM involves top-down signals from the IPFC to parietal cortex [53,94] as well as communication from frontoparietal regions to visual cortex [95]. Additionally, interactions between basal ganglia and the IPFC are thought to mediate the filtering of task-irrelevant information and the updating of task-relevant information in WM [96]. Future work must integrate knowledge about the neural computations performed by these individual regions in the service of WM with an understanding of the functions resulting from interactions between these regions to build a complete picture of WM.

distal times resulted in chance-level decoding. The inability to generalize decoding performance over time suggests that the patterns of activity containing information about the relevant category drifted over the course of the trial. Other studies have documented similar temporally dynamic population codes in monkey IPFC [37,39,74] and one fMRI study recently demonstrated dynamic population coding in visual cortex [24]. In contrast to the fixed-selectivity model, therefore, WM representations may not critically depend on the persistent activity of a fixed set of selectively tuned neurons.

Although the precise neuronal mechanisms underlying dynamic population coding remain undefined, recent computational work suggests that information can be maintained through rapid short-term changes in synaptic plasticity [75,76]. Thus, activity-dependent changes in neural networks trace dynamic trajectories through state space that reliably and efficiently encode task-related variables [72,76]. Future studies must determine the functional relevance of these mechanisms for the neurobiological implementation of WM.

### Concluding remarks

An understanding of the neural mechanisms underlying WM is critical for gaining insight into the wide range of

goal-directed behaviors supported by WM. In this review, we present a perspective on WM that emphasizes the notion of distributed population activity in encoding WM information. Methodological advances in the past ten years, and in the past few years in particular, have highlighted the sensory nature of sustained WM information in sensory cortices and the high-dimensional nature of information encoded by IPFC activity. Contrary to its imputed role as a storage buffer, we suggest that IPFC activity represents top-down influences on sensory regions. This conceptualization of the IPFC echoes models of hippocampal function that propose that the hippocampus stores pointers that can reactivate cortical memory traces rather than storing the memories themselves [77]. Future work must prioritize further efforts to specify the role of the IPFC in WM, as well as the functions that arise through interactions between the IPFC, sensory regions, and other regions involved in WM (Box 2). An emphasis on sustained representations through multiple neural mechanisms will facilitate the incorporation of these mechanisms into a comprehensive theory of WM.

#### Acknowledgments

This work was supported by grants from the National Institutes of Health (R01 MH63901 to M.D. and R01 EY016407 and R03 MH097206 to C.E.C.).

#### References

- Fuster, J.M. and Alexander, G.E. (1971) Neuron activity related to short-term memory. *Science* 173, 652–654
- Funahashi, S. *et al.* (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349
- Goldman-Rakic, P.S. (1995) Cellular basis of working memory. *Neuron* 14, 477–485
- Curtis, C.E. and D'Esposito, M. (2003) Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423
- Georgopoulos, A. *et al.* (1986) Neuronal population coding of movement direction. *Science* 233, 1416–1419
- Graf, A.B.A. *et al.* (2011) Decoding the activity of neuronal populations in macaque primary visual cortex. *Nat. Neurosci.* 14, 239–245
- Pasternak, T. and Greenlee, M. (2005) Working memory in primate sensory systems. *Nat. Rev. Neurosci.* 6, 97–107
- Ranganath, C. and D'Esposito, M. (2005) Directing the mind's eye: prefrontal, inferior and medial temporal mechanisms for visual working memory. *Curr. Opin. Neurobiol.* 15, 175–182
- Offen, S. *et al.* (2009) The role of early visual cortex in visual short-term memory and visual attention. *Vision Res.* 49, 1352–1362
- Duda, R.O. *et al.* (2001) *Pattern Classification*. (2nd edn), John Wiley & Sons
- Hung, C.P. *et al.* (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866
- Norman, K.A. *et al.* (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430
- Tong, F. and Pratte, M.S. (2012) Decoding patterns of human brain activity. *Annu. Rev. Psychol.* 63, 483–509
- Brouwer, G.J. and Heeger, D.J. (2009) Decoding and reconstructing color from responses in human visual cortex. *J. Neurosci.* 29, 13992–14003
- Harrison, S.A. and Tong, F. (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635
- Serences, J.T. *et al.* (2009) Stimulus-specific delay activity in human primary visual cortex. *Psychol. Sci.* 20, 207
- Ester, E.F. *et al.* (2013) A neural measure of precision in visual working memory. *J. Cogn. Neurosci.* 25, 754–761
- Xing, Y. *et al.* (2013) Decoding working memory of stimulus contrast in early visual cortex. *J. Neurosci.* 33, 10301–10311
- Albers, A.M. *et al.* (2013) Shared representations for working memory and mental imagery in early visual cortex. *Curr. Biol.* 23, 1427–1431
- Linden, D.E.J. *et al.* (2011) Mapping brain activation and information during category-specific visual working memory. *J. Neurophysiol.* 107, 628–639
- Christophel, T.B. *et al.* (2012) Decoding the contents of visual short-term memory from human visual and parietal cortex. *J. Neurosci.* 32, 12983–12989
- Woloszyn, L. and Sheinberg, D.L. (2009) Neural dynamics in inferior temporal cortex during a visual working memory task. *J. Neurosci.* 29, 5494–5507
- Han, X. *et al.* (2013) Multi-voxel pattern analysis of selective representation of visual working memory in ventral temporal and occipital regions. *Neuroimage* 73, 8–15
- Sreenivasan, K.K. *et al.* (2014) Distributed and dynamic storage of working memory stimulus information in extrastriate cortex. *J. Cogn. Neurosci.* (in press)
- 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
- Luck, S.J. and Vogel, E.K. (2013) Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn. Sci.* 17, 391–400
- Vogel, E.K. *et al.* (2005) Neural measures reveal individual differences in controlling access to working memory. *Nature* 438, 500–503
- Scolari, M. and Serences, J.T. (2010) Basing perceptual decisions on the most informative sensory neurons. *J. Neurophysiol.* 104, 2266–2273
- 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
- van de Ven, V. *et al.* (2012) Topographic contribution of early visual cortex to short-term memory consolidation: a transcranial magnetic stimulation study. *J. Neurosci.* 32, 4–11
- Silvanto, J. and Soto, D. (2011) Causal evidence for subliminal percept-to-memory interference in early visual cortex. *Neuroimage* 59, 840–845
- Cattaneo, Z. *et al.* (2009) Contrasting early visual cortical activation states causally involved in visual imagery and short-term memory. *Eur. J. Neurosci.* 30, 1393–1400
- Constantinidis, C. *et al.* (2001) The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nat. Neurosci.* 4, 311–316
- Warden, M.R. and Miller, E.K. (2010) Task-dependent changes in short-term memory in the prefrontal cortex. *J. Neurosci.* 30, 15801–15810
- Romo, R. *et al.* (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399, 470–473
- Wallis, J.D. *et al.* (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411, 953–956
- Stokes, M.G. *et al.* (2013) Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78, 364–375
- Meyers, E.M. *et al.* (2008) Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.* 100, 1407–1419
- Barak, O. *et al.* (2010) Neuronal population coding of parametric working memory. *J. Neurosci.* 30, 9424–9430
- Jun, J.K. *et al.* (2010) Heterogeneous population coding of a short-term memory and decision task. *J. Neurosci.* 30, 916–929
- Machens, C.K. *et al.* (2010) Functional, but not anatomical, separation of “what” and “when” in prefrontal cortex. *J. Neurosci.* 30, 350–360
- Rigotti, M. *et al.* (2010) Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Front. Comput. Neurosci.* 4, 24
- Rigotti, M. *et al.* (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590
- Sussillo, D. and Abbott, L.F. (2009) Generating coherent patterns of activity from chaotic neural networks. *Neuron* 63, 544–557
- Averbeck, B.B. and Seo, M. (2008) The statistical neuroanatomy of frontal networks in the macaque. *PLoS Comput. Biol.* 4, e1000050
- Postle, B.R. (2006) Working memory as an emergent property of the mind and brain. *Neuroscience* 139, 23–38
- 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
- Lee, S-H. *et al.* (2013) Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat. Neurosci.* 16, 997–999

- 49 Freedman, D.J. *et al.* (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316
- 50 Freedman, D.J. *et al.* (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.* 23, 5235–5246
- 51 Chen, A.J.-W. *et al.* (2012) Goal-directed attention alters the tuning of object-based representations in extrastriate cortex. *Front. Hum. Neurosci.* 6, 187
- 52 Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202
- 53 Crowe, D.A. *et al.* (2013) Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nat. Neurosci.* 16, 1484–1491
- 54 Fuster, J.M. (2001) The prefrontal cortex – an update: time is of the essence. *Neuron* 30, 319–333
- 55 Ekstrom, L.B. *et al.* (2008) Bottom-up dependent gating of frontal signals in early visual cortex. *Science* 321, 414–417
- 56 Noudoost, B. and Moore, T. (2011) Control of visual cortical signals by prefrontal dopamine. *Nature* 474, 372–375
- 57 Higo, T. *et al.* (2011) Distributed and causal influence of frontal operculum in task control. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4230–4235
- 58 Miller, B.T. *et al.* (2011) The prefrontal cortex modulates category selectivity in human extrastriate cortex. *J. Cogn. Neurosci.* 23, 1–10
- 59 Lee, T.G. and D'Esposito, M. (2012) The dynamic nature of top-down signals originating from prefrontal cortex: a combined fMRI–TMS study. *J. Neurosci.* 32, 15458–15466
- 60 Zanto, T.P. *et al.* (2011) Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nat. Neurosci.* 14, 656–661
- 61 Driver, J. *et al.* (2009) Concurrent brain-stimulation and neuroimaging for studies of cognition. *Trends Cogn. Sci.* 13, 319–327
- 62 Feredoes, E. *et al.* (2011) Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17510–17515
- 63 Cohen, J.R. *et al.* (2012) Correspondence between stimulus encoding- and maintenance-related neural processes underlies successful working memory. *Cereb. Cortex* <http://dx.doi.org/10.1093/cercor/bhs339>
- 64 Super, H. (2001) A neural correlate of working memory in the monkey primary visual cortex. *Science* 293, 120–124
- 65 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
- 66 Zelano, C. *et al.* (2009) A specialized odor memory buffer in primary olfactory cortex. *PLoS ONE* 4, e4965
- 67 Curtis, C.E. and Lee, D. (2010) Beyond working memory: the role of persistent activity in decision making. *Trends Cogn. Sci.* 14, 216–222
- 68 Deco, G. *et al.* (2013) Brain mechanisms for perceptual and reward-related decision-making. *Prog. Neurobiol.* 103, 194–213
- 69 Jerde, T.A. *et al.* (2012) Prioritized maps of space in human frontoparietal cortex. *J. Neurosci.* 32, 17382–17390
- 70 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
- 71 Afshar, A. *et al.* (2011) Single-trial neural correlates of arm movement preparation. *Neuron* 71, 555–564
- 72 Buonomano, D.V. and Maass, W. (2009) State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125
- 73 Druckmann, S. and Chklovskii, D.B. (2012) Neuronal circuits underlying persistent representations despite time varying activity. *Curr. Biol.* 22, 2095–2103
- 74 Meyers, E.M. *et al.* (2012) Incorporation of new information into prefrontal cortical activity after learning working memory tasks. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4651–4656
- 75 Mongillo, G. *et al.* (2008) Synaptic theory of working memory. *Science* 319, 1543–1546
- 76 Sugase-Miyamoto, Y. *et al.* (2008) Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS Comput. Biol.* 4, e1000073
- 77 Teyler, T.J. and DiScenna, P. (1986) The hippocampal memory indexing theory. *Behav. Neurosci.* 100, 147
- 78 Jimura, K. and Poldrack, R.A. (2012) Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia* 50, 544–552
- 79 Freeman, J. *et al.* (2011) Orientation decoding depends on maps, not columns. *J. Neurosci.* 31, 4792–4804
- 80 Todd, M.T. *et al.* (2013) Confounds in multivariate pattern analysis: theory and rule representation case study. *Neuroimage* 77, 157–165
- 81 Serences, J.T. and Saproo, S. (2012) Computational advances towards linking BOLD and behavior. *Neuropsychologia* 50, 435–446
- 82 Kay, K. *et al.* (2008) Identifying natural images from human brain activity. *Nature* 452, 352–355
- 83 Sprague, T.C. and Serences, J.T. (2013) Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nat. Neurosci.* 16, 1879–1887
- 84 Rossi, S. (2009) Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin. Neurophysiol.* 120, 2008–2039
- 85 Burgess, G.C. *et al.* (2011) Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *J. Exp. Psychol. Gen.* 140, 674–692
- 86 Ester, E.F. *et al.* (2009) Spatially global representations in human primary visual cortex during working memory maintenance. *J. Neurosci.* 29, 15258–15265
- 87 Canolty, R.T. and Knight, R.T. (2010) The functional role of cross-frequency coupling. *Trends Cogn. Sci.* 14, 506–515
- 88 Wang, X. (2010) Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* 90, 1195
- 89 Roux, F. and Uhlhaas, P.J. (2013) Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends Cogn. Sci.* 18, 16–25
- 90 Jacobs, J. and Kahana, M.J. (2010) Direct brain recordings fuel advances in cognitive electrophysiology. *Trends Cogn. Sci.* 14, 162–171
- 91 Rubehn, B. *et al.* (2009) A MEMS-based flexible multichannel ECoG-electrode array. *J. Neural Eng.* 6, 036003
- 92 Siegel, M. *et al.* (2009) Phase-dependent neuronal coding of objects in short-term memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21341
- 93 Bressler, S.L. and Menon, V. (2010) Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* 14, 277–290
- 94 Edin, F. *et al.* (2009) Mechanism for top-down control of working memory capacity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6802–6807
- 95 Chadick, J.Z. and Gazzaley, A. (2011) Differential coupling of visual cortex with default or frontal-parietal network based on goals. *Nat. Neurosci.* 14, 830–832
- 96 Hazy, T.E. *et al.* (2007) Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 362, 1601–1613