



HHS Public Access

Author manuscript

Vision Res. Author manuscript; available in PMC 2017 November 01.

Published in final edited form as:

Vision Res. 2016 November ; 128: 53–67. doi:10.1016/j.visres.2016.09.010.

Neural Mechanisms of Information Storage in Visual Short-Term Memory

John T. Serences

Department of Psychology, Neurosciences Graduate Program, and the Kavli Institute for Mind and Brain, University of California, San Diego

Abstract

The capacity to briefly memorize fleeting sensory information supports visual search and behavioral interactions with relevant stimuli in the environment. Traditionally, studies investigating the neural basis of visual short term memory (STM) have focused on the role of prefrontal cortex (PFC) in exerting executive control over what information is stored and how it is adaptively used to guide behavior. However, the neural substrates that support the actual storage of content-specific information in STM are more controversial, with some attributing this function to PFC and others to the specialized areas of early visual cortex that initially encode incoming sensory stimuli. In contrast to these traditional views, I will review evidence suggesting that content-specific information can be flexibly maintained in areas across the cortical hierarchy ranging from early visual cortex to PFC. While the factors that determine exactly where content-specific information is represented are not yet entirely clear, recognizing the importance of task-demands and better understanding the operation of non-spiking neural codes may help to constrain new theories about how memories are maintained at different resolutions, across different timescales, and in the presence of distracting information.

Introduction

Perception and memory are limited by internal factors such as the finite processing capacity of neural systems, as well as by external factors such as the movement and occlusion of objects in the visual field. Covertly shifting attention and overtly shifting gaze can help to overcome some of these limits; however, occluded objects often remain inaccessible for short periods of time and are thus unavailable for attentive scrutiny, and exploratory eye-movements severely disrupt the continuity of inputs to the retina. As a result, short term memory – or the ability to maintain a coherent representation of sensory information that is no longer present in the visual field – is required to stitch together a useful perceptual representation that persists across discontinuities in the input stream (Goldman-Rakic, 1987; Irwin, 1991; James, 1890; G. A. Miller, Galanter, & Pribham, 1960; Rolfs, 2015).

Contact: University of California, San Diego, Department of Psychology, 9500 Gilman Drive, La Jolla, CA, 92093-0109, jserences@ucsd.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Experimental efforts to understand the cognitive and neural architecture of short term memory (STM) have long been guided by a high degree of cross-talk between experimental psychology and neuroscience (Baddeley, 1986; Baddeley & Hitch, 1974; Fuster & Alexander, 1971; see also: Atkinson & Shiffrin, 1968; G. A. Miller et al., 1960). In one of the most influential early models, Baddeley and Hitch (1974) posited two memory buffers that independently store spatial and verbal information, coupled with a ‘central executive’ that is responsible for gating and manipulating information within these two content-specific buffers. The central-executive component of this model, or the source of control over STM, is thought to be supported largely via circuitry in the PFC. This account is consistent with well documented cognitive control deficits in patients with damage to the PFC (Badre & D’Esposito, 2009; Chao & Knight, 1998; Fuster, Bauer, & Jervey, 1985; G. A. Miller et al., 1960), as well as single-unit recording and functional neuroimaging evidence suggesting that areas of the PFC are involved in maintaining behavioral goals, task-switching, and adaptively manipulating information held in STM (D’Esposito, Postle, & Rypma, 2000; E. K. Miller & Cohen, 2001). Thus, even though some would include other non-PFC structures such as the basal ganglia as crucial nodes in an executive control network, few would dispute the key role played by PFC (e.g. McNab & Klingberg, 2008; E. K. Miller, 2013).

However, understanding the neural substrates that support the maintenance of content-specific information in STM has proven to be more controversial. Early evidence suggests a key role for maintenance in PFC, based on observations of sustained and stimulus-specific spiking activity during memory delays and on evidence from positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies showing that different sub-regions of the PFC can support different types of remembered information (Courtney, Petit, Haxby, & Ungerleider, 1998; Funahashi, Bruce, & Goldman-Rakic, 1989, 1993; Goldman-Rakic, 1995; Mendoza-Halliday, Torres, & Martinez-Trujillo, 2014; Qi et al., 2010; Schumacher et al., 1996; Smith & Jonides, 1999; Smith et al., 1995; Wang, 2001).

However, evidence from other studies using a variety of techniques suggest instead that the storage of information in STM is primarily mediated by more specialized sub-regions of cortex that represent low-level visual features or the identity of whole objects (Awh & Jonides, 2001; Chelazzi, Miller, Duncan, & Desimone, 1993; Curtis & D’Esposito, 2003; D’Esposito, 2007; D’Esposito & Postle, 2015; Harrison & Tong, 2009; Lara & Wallis, 2015; Magnussen, 2000; E. K. Miller, Li, & Desimone, 1993; Pasternak & Greenlee, 2005; Serences, Ester, Vogel, & Awh, 2009; Sreenivasan, Curtis, & D’Esposito, 2014). This view is known as the *sensory-recruitment hypothesis*, and is based on the intuition that neurons in early visual cortex are ideal candidates for storage because they exhibit highly selective tuning for different stimulus features such as orientation, spatial frequency and object identity. In effect, neural responses in visual cortex act as a bank of filters that are specialized to extract precise information about low-level properties of images. Thus, the tuning of neurons in early visual areas might be ideally suited to support both perception as well as mnemonic representations of these same features. This model has two intuitively appealing components. First, recruiting specialized regions of visual cortex to support STM might be a highly efficient way to avoid recoding remembered information using other distal anatomical structures types of neural codes (e.g. Stokes et al., 2013, see section below on *dynamic and activity (spike) silent codes*). Second, the high degree of feature-selectivity

found in many areas of early visual cortex is not typically observed in PFC, and a high degree of selectivity may be critical when trying to remember very subtle distinctions between items stored in STM. On the other hand, others have argued that storing information within early visual cortex would leave memory representations susceptible to overwriting as new sensory stimuli are processed, and that circuits in these regions are not intrinsically wired to instantiate the type of recurrent activity that is often thought to support STM (Bettencourt & Xu, 2016; Riley & Constantinidis, 2015; Stokes, 2015; Wang, 2001). Thus, two general camps have emerged: those who believe that the PFC mediates both control and storage, and those who believe that the PFC largely regulates executive function and that content-specific information is stored primarily in highly-selective regions of early visual cortex.

Here, I will review evidence about the respective roles of PFC and visual cortex in supporting executive control and the storage of content-specific information. For the purpose of focusing on the control/storage distinction, I will not review other important and related topics about the total storage capacity of STM or about the discrete or continuous nature of information in STM (see reviews by Luck & Vogel, 2013; Ma, Husain, & Bays, 2014; van den Berg, Awh, & Ma, 2014; Xu & Chun, 2009 that cover these topics in great detail). Instead, I will argue that the storage of information in STM can vary along a continuum that depends on task demands, and that considering other types of neural codes beyond the classically described sustained spiking in PFC may reveal previously overlooked mechanisms for adaptively storing remembered information.

Sustained activity and executive control functions in the PFC

Given its intuitive appeal, sustained neural activity during memory delay periods has been traditionally viewed as the most widely accepted signature of information storage in STM. A to-be-remembered stimulus (*sample stimulus*) is encoded, and during the retention interval, the sub-set of neurons involved in maintaining a representation of the sample simply spike in a continuous and highly stereotyped manner until the memory probe (*test stimulus*) is presented for comparison. In one early and seminal paper, Fuster and colleagues used a delayed-match-to-sample (DMTS) task in which a monkey had to covertly encode the spatial position of an occluded object (Fuster & Alexander, 1971). The majority of PFC neurons that were identified – as well as neurons in the dorsomedial nucleus of the thalamus that provides input to the PFC – exhibited elevated and sustained spiking activity across memory delay periods that lasted up to 30s. However, these sustained delay-period responses were not selective for the spatial position that the animal was remembering, which led the authors to conclude that sustained spiking in PFC was related to the maintenance of general task rules or behavioral goals as opposed to a spatial memory engram per se.

However, following these initial observations of non-selective sustained responses in PFC, other groups developed variants of the DMTS task in which an animal had to encode a peripheral spatial location that was the target of a saccadic eye movement after a brief 1s-6s delay period. In contrast to the non-selective responses reported by Fuster, many neurons around the principal sulcus, a sub-region of PFC, exhibited a spatially-selective response that carried information about the remembered location (Funahashi et al., 1989). However,

the role of these neurons in supporting spatial STM as opposed to motor planning is not entirely clear, as the remembered position was yoked to the endpoint of the planned saccade. Thus, even in the domain of relatively simple tasks, early unit recording data did not fully distinguish between content-specific memory signals and more general executive control functions related to task-set and motor planning.

Complementing the single-unit data (e.g. Fuster et al., 1971; Funahashi et al., 1989), early work using fMRI in human subjects also showed sustained activation profiles in PFC during memory delays. However, many of the same issues arose regarding whether these sustained activations reflect mnemonic storage or executive functions such as motor planning. In one early study, Courtney and coworkers used fMRI and a DMTS task in which subjects had to remember either the identity or the location of a series of faces (Figure 1; Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998; Courtney, Ungerleider, Keil, & Haxby, 1997). Persistent activity across the memory delay period was observed in different sub-regions of PFC depending on what type of information the subjects were remembering, suggesting a ventral/dorsal split between the storage of object and spatial information, respectively (Courtney, Petit, Haxby, et al., 1998; Courtney, Petit, Maisog, et al., 1998; Courtney et al., 1997; Petit, Courtney, Ungerleider, & Haxby, 1998). In contrast, other investigators argued that the site of delay-period activation in the PFC was more influenced by factors such as covertly planning eye-movements to spatial locations during STM. In support of this view, Curtis and colleagues designed two versions of a DMTS task: in one version subjects were able to plan a saccadic response during the delay period, and in another version subjects were not able to pre-plan the motor response and just had to remember the sensory attributes of the sample stimulus. Many regions of PFC (and parietal cortex) tracked the motor intention of subjects during the memory delay, particularly the dorsal areas that were previously tied to spatial STM (Curtis & D'Esposito, 2003; Curtis, Rao, & D'Esposito, 2004). Thus, the authors proposed that the dorsal/ventral functional division of PFC was not as tightly associated with representing remembered spatial positions and objects, but instead was likely influenced by other factors including planning spatially covert motor plans (Curtis & D'Esposito, 2003, 2004; Curtis et al., 2004; Postle, Berger, Taich, & D'Esposito, 2000; Postle & D'Esposito, 1999).

To avoid the inherent difficulties involved in dissociating spatial WM from spatially-specific response planning, other groups employed non-spatial features such as motion direction, abstract patterns, and oriented gratings to evaluate the significance of sustained activation profiles in PFC. In one line of studies, Postle and colleagues measured fMRI in human subjects while they performed a DMTS task for different directions of motion. However, their approach departed from the traditional method of analyzing the mean timecourse of sustained activation across all voxels (or 'volumetric pixels') within a region. Instead, they used multivariate pattern analysis (MVPA) to examine changes in the *pattern* of activation across voxels within each sub-region of the PFC. By focusing on changes in the large-scale pattern of responses across many voxels, MVPA is far more sensitive to detect whether a brain region is encoding information about a remembered feature during STM (Postle, 2015; Serences & Saproo, 2012; Sprague & Serences, 2015; Tong & Pratte, 2012). This increase in sensitivity arises because a pattern of activation within a region can systematically track changes in the contents of STM even if the mean amplitude of the responses across all

voxels in that region remains perfectly stable. For example, suppose that a spatially intermixed $\frac{1}{2}$ of the voxels in a region increase their response to stimulus A and decrease their response to stimulus B, whereas the other intermixed $\frac{1}{2}$ of the voxels show the opposite pattern of selectivity. As the subject alternately views stimulus A and stimulus B, there would be no net change in the response averaged across all voxels within the region, leading to the erroneous conclusion that no stimulus-specific information was encoded. However, focusing on systematic changes the spatial distribution of the response pattern can reveal content-specific information and lead to the correct inference that the area is sensitive to changes in the features of the remembered stimulus (Figure 2).

Postle and colleagues used this approach to show that information about remembered stimulus features was present in the same areas of early visual cortex that are commonly implicated in the encoding of motion (i.e. the middle temporal area, or MT: see also the next section on the sensory *recruitment hypothesis*). However, even though they observed sustained increases in univariate responses in PFC, which is the traditional signature of information storage during STM, activation patterns in PFC primarily encoded information about a subject's global task-set, such as which feature was currently relevant, as opposed to content-specific information about the feature that was being remembered (Figure 3; Emrich, Riggall, Larocque, & Postle, 2013; Riggall & Postle, 2012). Using a similar analysis technique, Lee and colleagues also reported that the identity of a remembered object could be decoded based on activation patterns in occipital visual cortex, whereas activation patterns in PFC only encoded information related to the task-set of the subject (S. H. Lee, Kravitz, & Baker, 2013). Finally, Christophel and coworkers carried out a series of studies that examined remembering information about several different types of visual features, including abstract patterns that were designed to completely eliminate verbal coding. Consistent with findings by the Postle group, they found evidence for content-specific STM representations within sub-regions of occipital and parietal cortex, but not in the PFC (Christophel, Cichy, Hebart, & Haynes, 2015; Christophel & Haynes, 2014; Christophel, Hebart, & Haynes, 2012). Together, these failures to find evidence for content-specific information in PFC suggest that commonly reported increases in sustained activity are primarily related to executive control functions rather than to storing specific features of remembered stimuli (D'Esposito & Postle, 2015; Postle, 2015).

The general conclusion that PFC neurons primarily mediate executive control functions is also consistent with data from a variety of other tasks and experimental methodologies. Miller and colleagues reported that single PFC neurons encoded more task-related information about whether a test stimulus was a match or a non-match to the sample, and less information about the low-level stimulus attributes of the remembered stimulus (E. K. Miller, Erickson, & Desimone, 1996). In another task, they trained monkeys to judge whether or not two stimuli were members of the same category, where category membership was flexibly defined based on the currently cued task set. PFC neurons were sensitive to changes in the task set of the subject, even when the physical properties of the stimulus set did not change and the subjects were asked to generalize the application of the categorization rule to novel stimuli (Wallis, Anderson, & Miller, 2001; Wallis & Miller, 2003; White & Wise, 1999). Furthermore, PFC neurons encode information about changes in the structure of a memory task, even when the same sensory stimuli are used (e.g. either a

recognition or a recall task), and PFC neurons also signal the categorical identity of a continuously varying stimulus that is defined by an arbitrary decision boundary (Freedman, Riesenhuber, Poggio, & Miller, 2001, 2002; McKee, Riesenhuber, Miller, & Freedman, 2014; Warden & Miller, 2010). Thus, in many situations, PFC neurons exhibit a high degree of task-selectivity and a relative lack of sensitivity to the low-level properties of the visual stimuli that are being stored in STM.

Collectively, these studies underscore the importance of PFC structures in mediating executive control functions in STM. First, retaining information in STM is often accompanied by sustained increases in both spiking activity and mean fMRI activation levels. In some cases, this sustained activity seems to support relatively general cognitive functions such as maintaining tonic attention on the relevant behavioral task (Fuster & Alexander, 1971). In other cases, the observation of sustained activation has been more closely associated with encoding information about motor intentions or the specific task-set that guides the behavior of the subject (Curtis & D'Esposito, 2003, 2004; Curtis et al., 2004; S. H. Lee et al., 2013; Postle, 2015; Riggall & Postle, 2012; Sreenivasan et al., 2014). Second, there are now many reports that fail to find content-specific information in PFC using MVPA and fMRI (Christophel et al., 2015; Christophel & Haynes, 2014; Christophel et al., 2012; Emrich et al., 2013; Riggall & Postle, 2012).

The sensory recruitment hypothesis of content-specific STM representations

As discussed above, there is strong evidence that delay-period activity in the PFC plays a role in mediating executive control over current task-set as well other factors such as motor control. However, the complexity and variety of responses in PFC motivated the search for more unambiguous the content-specific memory representations, and a natural candidate for such representations is the regions of early visual cortex that initially encode low-level features of sensory stimuli. One of the earliest, and arguably most articulate, formulations of this hypothesis was put forth by William James over 100 years ago (1890) when he wrote that, “The same cerebral process which, when aroused from without by a sense-organ, gives the perception of an object, will give an idea of the same object when aroused by other cerebral processes from within”. This general idea has come to be known as the *sensory recruitment* hypothesis and has re-appeared in many guises over the years, making contact with other literatures such as embodied cognition and the idea that long-term memory retrieval relies on the reactivation of sensory circuits (Barsalou, 2008; Nyberg, Habib, McIntosh, & Tulving, 2000).

Renewed focus on the sensory recruitment hypothesis sprung from a series of single-unit recording and fMRI studies that were carried out in the 1990's, although in many cases the evidence was not initially interpreted in this framework. In one set of studies, Desimone and colleagues demonstrated that object-selective neurons in ventral visual cortex, or inferotemporal (IT) cortex, exhibit sustained and highly-selective responses during DMTS tasks (Chelazzi, Duncan, Miller, & Desimone, 1998; Chelazzi et al., 1993; E. K. Miller, Li, & Desimone, 1991; E. K. Miller et al., 1993; see also: Fuster & Jervey, 1981; Miyashita &

Chang, 1988; Nakamura & Kubota, 1995). Interestingly, when compared to neurons in PFC, responses in IT neurons during STM were more susceptible to disruption when distracting stimuli were presented during a memory delay period (E. K. Miller et al., 1996; E. K. Miller et al., 1993). This disruption of sustained activity by distractors was initially taken to suggest a limited role for IT neurons in STM. Thus, sustained responses in the PFC were assumed to play a relatively important role in supporting content-specific representations in a manner that was more resistant to intervening distractors. However, many of these same studies also demonstrated that the responses of IT neurons were more selective for the remembered object compared to responses in PFC neurons (E. K. Miller et al., 1996). This later result thus suggests that highly-tuned neurons in IT may play a role in remembering the finer details of remembered objects, and suggest that they play a role in maintaining more precise memories, even if those precise memories were subject to overwriting. Roughly analogous results were also reported in early fMRI studies using human subjects: remembering objects that belong to different categories such as faces and non-faces induces sustained increases in the fMRI signal within content-specific object-selective sub-regions of IT cortex (Courtney et al., 1997; O'Craven & Kanwisher, 2000; Ranganath & D'Esposito, 2005; Ranganath, DeGutis, & D'Esposito, 2004).

While these initial content-specific responses were observed in IT, other data suggests that even earlier regions of retinotopically organized visual cortex can also encode information about remembered features. Awh and colleagues had subjects perform a highly demanding spatial STM task and demonstrated that abruptly appearing objects – which are known to capture spatial attention – also disrupt memory when they are presented in non-remembered locations during the delay period (Awh, Anillo-Vento, & Hillyard, 2000; Awh, Jonides, & Reuter-Lorenz, 1998; Yantis & Jonides, 1984). While these behavioral studies are not completely diagnostic of a particular neural substrate, the spatially-selective nature of these effects is consistent with a STM mechanism that relies on activity within the same retinotopically organized regions of visual cortex that support spatial vision. Follow-up studies using fMRI and EEG further demonstrate that remembering the spatial location of an object gives rise to sustained and spatially selective responses in early visual cortex in a manner analogous to modulations observed when spatial attention is deployed to a continuously visible peripheral object (Awh et al., 2000; Awh et al., 1999). Together, these results suggest that at least partially overlapping neural mechanisms sub-serve the initial processing of spatial information as well as the ability to precisely represent an object's location in STM (Awh & Jonides, 2001).

In addition to the domain of spatial memory, studies also suggest that STM for non-spatial features is supported by the same neural populations in early visual cortex that are known to be involved in basic sensory processing. Using psychophysical methods, Magnussen ran a series of studies to determine if the presentation of a sensory stimulus would systematically interfere with a feature-selective mnemonic representation while subjects remembered the spatial frequency of an oriented grating. During the delay period on some blocks of trials, a distractor stimulus was presented, and significant interference was observed when the spatial frequency of the distractor did not match the spatial frequency of the remembered stimulus (Magnussen, Greenlee, Asplund, & Dyrnes, 1991). Rademaker further demonstrated that memory for orientation is systematically biased when an oriented distractor is presented

mid-way through a memory delay period (Rademaker, Bloem, De Weerd, & Sack, 2015), and conceptually similar paradigms reveal similar interactions during the encoding of information in STM and during the initial analysis of low-level visual features (Huang & Sekuler, 2010; Magnussen, 2000; Magnussen & Greenlee, 1992, 1999; Nemes, Parry, Whitaker, & McKeefry, 2012). These results thus demonstrate that presenting a distractor can interfere with STM in a manner consistent with a co-localized, and thus interacting, representation of both mnemonic and sensory information in early visual cortex.

More recently, studies have followed up on these psychophysical results using MVPA and fMRI to more directly assess the role of early visual cortex in maintaining content-specific representations of information in STM. As noted in the previous section, the use of MVPA represents a significant advance over univariate approaches, particularly in early areas of visual cortex that do not typically show sustained increases in delay-period activation (e.g. Figure 3; Riggall & Postle, 2012). The lack of a sustained mean response during the delay period in these early visual areas may be due to several factors, but studies of feature-based attention offer a ready example of why using MVPA is so important in this case. When attention is deployed to a feature such as a direction of motion, the response of cells that are tuned to the attended feature are enhanced and the response of cells that are not tuned to the attended feature are suppressed (Martinez-Trujillo & Treue, 2004; Scolar, Byers, & Serences, 2012). Thus, if a similar combination of response enhancement and suppression is maintained across a memory delay period, there would be little overall change in the mean response amplitude averaged across all voxels because the response enhancement and response suppression would cancel out. However, if different voxels within a visual area exhibit even a slight bias in their response to different feature values, then using MVPA to examine the spatial pattern of response modulations should reveal systematic changes related to the storage of different features in STM (Postle, 2015; Serences & Saproo, 2012; Tong & Pratte, 2012).

Using MVPA, several studies now demonstrate that mnemonic representations for basic visual features such as orientation and color can be observed as early as primary visual cortex (V1, Emrich et al., 2013; Harrison & Tong, 2009; Serences et al., 2009). To address the concern that these modulatory patterns were simply a passive echo of a sensory response evoked by the presentation of the sample stimulus, Harrison and Tong (2009) presented two sequential oriented gratings and then post-cued one of the gratings as the relevant item to remember over the delay period. This method balanced the amount of sensory information that was evoked by the remembered and the non-remembered gratings, so any sustained modulation of activation patterns during the delay period could be unequivocally linked to a memory representation as opposed to a passive sensory trace. Information about the remembered feature was observed across the delay period, whereas information about the non-remembered feature was not. In another study, Serences et al. (2009) manipulated whether the color or the orientation of a sample grating was relevant on each trial. Response patterns in V1 only encoded information about the relevant feature dimension (color or orientation) during the memory delay, suggesting that selective attention governs the information content of activation patterns in early visual cortex during STM. In subsequent studies, this link between feature-based attention and STM was further bolstered by the observation that memory-related activation patterns in early visual areas are modulated in a

spatially global manner that is also characteristic of feature-based attention (Ester, Serences, & Awh, 2009; but see: Pratte & Tong, 2014). More strikingly, several studies now demonstrate that shifting attention toward and away from a feature-selective mnemonic representation leads to a systematic rise and fall in the amount of information that early visual areas encode about the identity of currently relevant and irrelevant stimuli (Emrich et al., 2013; LaRocque, Lewis-Peacock, Drysdale, Oberauer, & Postle, 2013; Larocque, Lewis-Peacock, & Postle, 2014). Together, these observations build on earlier work in the domain of spatial STM and bolster the notion that the contents of STM are maintained via a process that shares a representational format with selective attention to continuously visible sensory stimuli.

While the above studies establish that modulations in early visual areas track the information content of STM for spatial positions and features, establishing a direct link between these modulations and behavior is more challenging. Although not causal, several studies have now demonstrated that the information content of response patterns in early visual cortex is positively correlated with behavioral performance on STM tasks (Bettencourt & Xu, 2016; Emrich et al., 2013; Ester, Anderson, Serences, & Awh, 2013). Ester et al. (2013) asked subjects to remember a single oriented grating across a 10s delay period. As opposed to using MVPA, Ester used an inverted encoding model (IEM) to reconstruct a representation of the remembered stimulus orientation on each trial. Like MVPA, this method also exploits multivariate activation patterns, but instead of using activation patterns to infer a categorical stimulus label on each trial, IEMs provide a continuous and graded reconstruction of the remembered feature by modelling the response of each voxel using a set of functions that are meant to mimic the response properties of neurons in that visual area (i.e. basis functions; see Figure 4A–D). After the feature-selectivity of each voxel is estimated using this model, a novel activation pattern is combined with information about the selectivity of each voxel to reconstruct an image of the stimulus, where the ‘image’ of the stimulus is in the native feature space (i.e. orientation space in this example; Figure 4E, reviewed in Sprague, Saproo, & Serences, 2015). This approach is therefore more constrained than MVPA in that it requires an explicit model of how stimulus features are processed and represented in a given visual area (whereas MVPA is more flexible and can exploit any consistent patterns in the signal). However, IEMs can be more powerful because quantifying the fidelity of reconstructed images can support more precise and graded inferences about how different experimental manipulations impact the quality of internal representations (Brouwer & Heeger, 2009, 2011; Naselaris & Kay, 2015; Naselaris, Kay, Nishimoto, & Gallant, 2011; Serences & Saproo, 2012; Sprague et al., 2015; Tong & Pratte, 2012)

Using an IEM, Ester (2013) reconstructed representations of remembered orientations based on activation patterns in V1 and V2 and demonstrated a between-subject correlation between the precision of the orientation reconstructions and the precision of behavioral recall performance at the end of the delay period (Figure 4E–F). Later, Sprague used a similar IEM method to reconstruct representations of remembered locations and demonstrated a systematic degradation in the fidelity of STM reconstructions as subjects remembered more locations, consistent with the well-documented drop in memory precision with increasing memory load (Emrich et al., 2013; Ma et al., 2014; Sprague, Ester, & Serences, 2014; Zhang & Luck, 2008).

Moving beyond these correlational studies, others have experimentally perturbed neural activity to determine if these modulations in early visual cortex are causally involved in STM. In one study, subjects remembered the direction of a peripheral motion stimulus while transcranial magnetic stimulation (TMS) was applied over visual cortex to induce a visual ‘phosphene’, or the percept of a transiently moving cloud of light. When subjects simultaneously remembered a direction of motion, their perception of the TMS-induced phosphene was systematically biased in the direction of the remembered motion stimulus. Thus, this finding again supports overlap in the mechanisms that support visual perception and visual STM (Silvanto & Cattaneo, 2010). Similarly, TMS over the region of occipital cortex that encodes an array of sample stimuli disrupts encoding and storage in STM, particularly when memory load is high (van de Ven, Jacobs, & Sack, 2012). When combined with other psychophysical and neuroimaging evidence, these data lend causal support for theories that emphasize a common role of early visual areas in supporting perception and in maintaining content-specific information during STM.

Evidence linking executive control signals in PFC with content-specific signals in early visual areas

As reviewed in the previous sections, the modal model of STM holds that PFC primarily mediates executive control functions such as maintaining information about task-set. In turn, these executive control signals are thought to provide biasing input to the earlier areas of visual cortex that maintain content-specific mnemonic representations (Postle and D’Esposito, 2015). This theoretical framework is supported by a number of empirical observations made using both human and non-human primate model systems. Pribham and coworkers (1964) reviewed early work suggesting that damage to sub-regions of the PFC reliably produced deficits in tasks that required some element of delay between a stimulus and response, including delayed alternation tasks that were often used to more specifically assay STM (Pribham, Ahumada, Hartog, & Roos, 1964). Knight and colleagues later demonstrated that lesions in the lateral PFC (IPFC) gave rise to deficits in a more traditional DMTS task, and that EEG markers of distraction from irrelevant sensory stimuli were especially elevated in patients with damage to IPFC compared to controls (Chao & Knight, 1998). This later finding suggests that PFC damage selectively impairs the ability to suppress distracting information from overwriting content-selective representations stored in earlier visual areas. Finally, Fuster and colleagues showed that reversible inactivation of the PFC led to a reduction in stimulus-specific spiking activity in IT cortex during the delay period of a DMTS task (Fuster et al., 1985). Similar studies have also shown that reversibly interrupting PFC function using theta-burst TMS in human subjects leads to impaired performance on a DMTS task that is accompanied by corresponding declines in the precision of content-specific representations in visual cortex (Feredoes, Heinen, Weiskopf, Ruff, & Driver, 2011; Higo, Mars, Boorman, Buch, & Rushworth, 2011; T. G. Lee & D’Esposito, 2012; B. T. Miller, Vytlačil, Fegen, Pradhan, & D’Esposito, 2011; Zanto, Rubens, Thangavel, & Gazzaley, 2011). Thus, perturbations in PFC can directly impact the quality of stimulus-specific codes in earlier regions of occipital and ventral visual cortex, consistent with the putative role of PFC in exerting executive control over content-specific memory buffers in early visual cortex.

More than control signals: evidence in favor of content-specific mnemonic representations in PFC

While the general framework outlined above holds that there is a functional split between PFC control-regions and content-specific early visual areas, much evidence instead suggests a more flexible and nuanced account. As discussed above, some early reports show that sustained spiking activity in PFC might support spatial STM (e.g. Funahashi et al., 1989). However, as noted previously, some of these findings were difficult to interpret because of simultaneous changes in task-demands and motor preparation.

That said, more recent experimental efforts are better able to dissociate content-specific representations from other executive functions, and they suggest that neural activity in PFC can also encode content-specific representations of remembered stimuli. In one study, monkeys were trained to remember a spatial position of a sample stimulus and then to make a saccade to one of two positions. Importantly, the correct saccade vector could not be anticipated until the animal interpreted a set of cues that were presented at the end of the delay period. Persistent spiking activity that was systematically yoked with a particular remembered location was still observed in this context, even though the animal could not easily pre-plan the required motor response (Funahashi, Chafee, & Goldman-Rakic, 1993; Qi et al., 2010; Qi, Meyer, Stanford, & Constantinidis, 2011). Another study required remembering and comparing a single feature with a test stimulus that was presented in an unpredictable location. As in the spatial post-cuing task employed by Qi et al. (2010), this experimental manipulation dissociated the contents of memory from the planning of a motor response, yet robust population coding for the remembered feature was still observed within sub-regions of IPFC (Mendoza-Halliday et al., 2014). Finally, the magnitude of sustained spiking in PFC has been shown to correlate with the remembered physical intensity of a sample stimulus (Constantinidis, Williams, & Goldman-Rakic, 2002). This later result suggests that sustained activity in PFC reflects the quality of the internal representation of the remembered stimulus, independent of (or in addition to) the intention to respond via a spatially directed movement.

In addition to these single-unit studies, recent fMRI studies also support content-specific representations in sub-regions of the PFC. Ester et al (2015) used the same multivariate IEM analysis approach described earlier to show that orientation-selective information was present in many sub-regions of both parietal and frontal cortex during a memory delay period (Figure 5; Ester, Sprague, & Serences, 2015). At first glance, this demonstration of content-specific representations in PFC is hard to reconcile with earlier null results in PFC that were obtained with standard MVPA approaches (see e.g. Christophel et al., 2012; Emrich et al., 2013; S. H. Lee et al., 2013). However, there are also hints in some of these previous studies suggesting that more information may be present in PFC than typically thought. Several groups have now used MVPA and IEM approaches to demonstrate that information about remembered spatial locations is present within sub-regions of the PFC around the superior precentral sulcus (sPCS, or the putative human frontal eye fields, Jerde & Curtis, 2013; Jerde, Merriam, Riggall, Hedges, & Curtis, 2012; Sprague et al., 2014). Moreover, Lewis-Peacock et al. (2012) reported significant classification accuracy for

different categories of visual stimuli in PFC, even though there was considerably less information in PFC compared to other regions in occipital and parietal cortex (Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012). Together these results suggest that some degree of content-specific information is present within sub-regions of the PFC. However, MVPA, combined with the relatively coarse spatial scale of fMRI, may not be the ideal tool to reveal these modulations as the spatial scale of feature-selective sub-populations of neurons in PFC is much finer than the scale observed in early visual cortex (see: Mendoza-Halliday et al., 2014). That said, Ester's result showing feature-selective representations in PFC does suggest that some methods – like the IEM approach they employed – might be more sensitive as it effectively ignores variance in the data that is not systematically related to variations along the feature dimension that is being explicitly modeled. However, additional studies that employ a variety of converging methods and stimulus sets will be needed to establish the stability and generalizability of STM representations in human PFC.

Finally, recent single unit recording studies demonstrate that neurons in the PFC exhibit highly complex tuning profiles that may prove problematic for many methods that attempt to dissociate control and storage operations in PFC. For example, Warden and Miller, 2010 demonstrated that many PFC neurons respond to variables related to the current task set as well as to content-selective information about the remembered stimuli (Mante, Sussillo, Shenoy, & Newsome, 2013; Rigotti et al., 2013; Warden & Miller, 2010). This type of neural 'multiplexing' suggests the same cells can be selective for sensory attributes of a remembered item as well as for other more abstract factors that govern the behavior of an animal. Thus, the observation that a particular cell seems to exhibit selectivity for encoding variables related to executive control does not rule out a role in simultaneously encoding content-specific information about remembered features. Moreover, this mixed selectivity might undermine the utility of conventional MVPA techniques – in both fMRI studies and in multi-unit physiology studies – if neurons in PFC are simultaneously encoding information about both executive control signals and remembered features. For example, if neural responses to different types of information combine linearly, then standard MVPA should still be able to uniquely associate a pattern of responses with a particular task-set or a particular remembered feature. However, if neural responses combine non-linearly, then standard MVPA approaches will not be as sensitive to learn the proper associations. That said, the multiplexing capacity of PFC neurons is still an active area of investigation, and it is too early to tell exactly how this may impact the interpretation of spiking and fMRI activity. However, it does seem clear that PFC neurons exhibit relatively complex response patterns, and considering this complexity will be important when trying to reconcile competing accounts regarding the role of PFC in mediating executive functions versus maintaining content-specific representations in STM.

Evidence against content-specific representations in early visual cortex

While the evidence reviewed in the preceding section argues against the notion that PFC is solely an executive control center, other recent evidence challenges the companion idea that content-specific representations in visual cortex are the primary mechanism of storage. Bettencourt and Xu (2016) argue against sensory-recruitment on the logical grounds that distractors are almost always present during everyday perception, and should therefore wipe

out mnemonic representations that are maintained in early visual cortex. Instead, they suggest that areas of parietal and frontal cortex might support WM in a format that is more insulated from distractor interference. Consistent with this theory, Bettencourt and Xu (2016) show that presenting distractors during a memory delay period interferes with mnemonic representations in visual cortex. However, this interference effect only happens when the distractors are expected; unexpected distractors did not lead to a loss of information coding in early visual cortex (Bettencourt & Xu, 2016). Interestingly, the degree of distracting input that might occur during STM is highly unpredictable in everyday perception, and this situation is arguably more closely related to the experiment in which unexpected distractors failed to interfere with mnemonic representations in early visual cortex (their Experiment 3). Moreover, the authors overall interpretation that early visual areas do not play a role in STM is at odds with other studies showing that expected distractors bias mnemonic representations in a systematic manner and thus implicate a partially shared code for perception and for STM (Awh et al., 2000; Awh & Jonides, 2001; Awh et al., 1998; Magnussen, 2000; Magnussen & Greenlee, 1999; Rademaker et al., 2015). Nevertheless, the study raises the important, and usually neglected, point that the expectation or presence of distracting information can systematically bias the anatomical substrates that support mnemonic representations.

To simultaneously evaluate the role of IPFC and early visual cortex in maintaining information in STM, Mendoza-Halliday et al (2014) used multi-unit electrophysiology and a task that required remembering the direction of a moving-dot field. They found that patterns of spiking activity in motion selective middle temporal area (MT) only encoded information while the sample stimulus was physically present in the display – information about the remembered directions dropped precipitously as soon as the sample stimulus was extinguished and the memory delay period began (Figure 6A–B; Mendoza-Halliday et al., 2014). In contrast, spiking activity patterns in the medial superior temporal area (MST), which is immediately adjacent to area MT, and in IPFC were able to discriminate the remembered direction across the entire delay period, consistent with the operation of content-specific STM buffers. Based on this failure to observe spike-based coding of remembered features in area MT, the authors argue against the sensory recruitment hypothesis. At the same time, their data also provide compelling evidence for sustained mnemonic representations in sub-regions of the PFC, which simultaneously argues against a restricted role of PFC in mediating executive control.

Alternate mechanisms of storage – dynamic and activity (spike) silent codes for short-term memory

Until recently, most studies focused on temporally stable activity patterns as the primary signature of information storage in STM (e.g. Fuster, 2000; E. K. Miller et al., 1996; Wang, 2001). However, there is increasing evidence that a variety of alternate neural codes can support content-specific representations, and that understanding these novel codes may reveal new insights about how and where information is stored in STM. In one of the most clear-cut empirical examples to date, Stokes and coworkers trained non-human primates (NHPs) to remember three cue-target pairings. Multi-unit activity patterns were then

recorded from PFC while subjects tried to discriminate the cued memory target from a stream of distractors. The firing rate of single neurons was highest following the presentation of the cued target, but quickly drifted back to baseline and did not exhibit sustained activity across the trial. More interestingly, the multivariate pattern of activity across PFC neurons showed a high degree of stimulus selectivity during the initial encoding of the memory cue. Importantly, however, the stimulus-selective response pattern was not static, and instead underwent a continuous transition across the first part of the trial – initially there was little cross-timepoint transfer indicating a highly dynamic code. However, the code continuously evolved and by the end of the trial there was relatively better cross-timepoint transfer (also still not complete transfer; see Figure 7 and also: Myers et al., 2015). This finding stands in contrast to previous observations that a relatively sustained and stable code is the hallmark of information storage in STM, and suggests instead that stable mnemonic representations can be maintained even if the underlying neural code is highly dynamic over time. Moreover, one potential advantage of a dynamic code that evolves farther and farther away from the initial sensory-evoked pattern is that it would be more easily separable from the pattern of activity evoked by newly perceived perceptual stimuli (Stokes, 2015). Thus, in theory at least, temporally dynamic codes could provide a mechanism that insulates STM representations from interference by incoming sensory inputs. While this framework has many appealing features, future studies should evaluate whether these dynamic codes reflect a general mechanism of STM, or whether they are primarily recruited when a relatively small set of highly-trained stimulus-response mappings are used and the need to maintain a highly precise representation of a remembered sensory stimulus in STM is not required.

Other models propose an even more radical departure from traditional accounts by suggesting that mnemonic representations can be retained for short periods of time even in the complete absence of ongoing spiking activity. Mongillo et al. (2008) developed a biologically plausible neural network model and demonstrated that the stimulus-specific pattern of spiking during the presentation of a sample stimulus would alter the concentration of pre-synaptic calcium (CA⁺) and would leave a primed pathway in the network that could facilitate the evaluation of a subsequent test stimulus (Mongillo, Barak, & Tsodyks, 2008). Similarly, Sugase-Miyamoto and colleagues proposed that object-selective neurons in IT cortex rapidly adjust their synaptic weights based on the patterns of inputs evoked by a sample stimulus (Sugase-Miyamoto, Liu, Wiener, Optican, & Richmond, 2008). When a subsequent stimulus is presented, the pattern of evoked responses could be combined with the adjusted synaptic weights to determine a correlation between the new stimulus and the remembered sample stimulus. On this account, IT neurons form matched filters and their net output upon the presentation of the test stimulus is sufficient to determine the probability that the sample and the test stimuli are the same (Sugase-Miyamoto et al., 2008). Importantly, both of these models show how the properties of a network can be rapidly reconfigured via short-term plasticity to support a memory trace that does not depend on sustained spiking during the delay period, providing a metabolically inexpensive and almost entirely passive means to store remembered information (see Stokes, 2015) for a more focused review of activity-silent mechanisms).

Finally, sub-threshold changes in membrane potential within areas of visual cortex can track changes in the low-level properties of a remembered stimulus, thus suggesting another form of relatively silent STM code. In the same paper that failed to find spiking codes for remembered items in MT, Mendoza-Halliday et al. (2014) demonstrated that local field potentials (or LFPs) in MT were reliably modulated by the direction of the remembered stimulus (Figure 6C). LFPs can be driven by many sources, including bottom-up input, feedback from higher regions, and local connections. Importantly though, there was no bottom-up input available to MT during the memory delay period, so the LFP modulations were most likely driven by top-down input. Consistent with this account, there was significant spike-field synchrony between IPFC and MT during the delay period, suggesting that IPFC exerts a feature-selective biasing influence on LFPs in MT (with the spike-field coherence peaking in the alpha-low-beta range; see also: Liebe, Hoerzer, Logothetis, & Rainer, 2012).

Given conventional wisdom in the field about the exclusive importance of spikes, content-specific patterns of LFP modulation may support a highly adaptive type of code that can simultaneously provide a high-fidelity memory representation while also protecting that representation from overwriting by intervening distractors (see also: Bisley, Zaksas, Droll, & Pasternak, 2004; Zaksas & Pasternak, 2006). Consider the three areas studied by Mendoza-Halliday: sensory area MT, association area MST, and the IPFC. On this account, the sample stimulus evokes a feature-selective spiking pattern in MT, and when the sample stimulus is removed from view, spiking activity in MT drops off quickly such that spiking patterns no longer encode information about the remembered stimulus feature. At the same time, however, sustained and content-specific patterns of spiking persist in IPFC, and this persistent spiking in PFC may provide important top-down feedback to MT that induces a locally spike-silent but content-specific change in LFPs (Figure 8). Finally, under this coding scheme, the presentation of a matching test stimulus at the end of the delay period may more easily trigger the re-emergence of the same spiking pattern that was present in MT during the encoding of the sample stimulus, as this pattern is held in a 'primed' state during the STM delay period via the continuous top-down modulation of LFPs (see also: Lui & Pasternak, 2011 for conceptually related ideas). Importantly, because this tonic top-down input to MT is generated via sustained spiking activity in distal IPFC, the content-specific patterns of sub-threshold membrane potentials in MT could co-exist with and persevere through bouts of local spiking in MT that are evoked by other distracting stimuli presented during the delay period. Thus, tonic top-down input may allow a relatively early visual area like MT to retain high-fidelity information about a remembered stimulus in a non-spiking format that could nevertheless facilitate content-specific spiking in response to a test stimulus in an adaptive manner that would support behavioral performance. This model is appealing because tonic top-down biasing signals from the PFC to early visual areas – where neurons are more precisely tuned to low-level visual features – may provide a means of maintaining a more precise code for low-level sensory features that could support the comparison of sample and test stimuli while simultaneously decreasing the probability that new bottom-up input would overwrite this locally spike-silent mnemonic code.

Aside from the potential theoretical importance of this kind of locally-spike-silent code in early visual areas, this account may also help to explain why fMRI studies commonly find

that modulations of activation patterns in early visual areas such as V1 predict the contents of STM whereas single unit recording studies typically find little evidence for sustained spiking activity in the same early visual areas (Mendoza-Halliday et al., 2014). That said, there are several caveats that may bear on such attempts to reconcile the relative lack of spiking in early visual cortex during STM and the relatively strong fMRI responses. First, single-unit studies usually record from a small set of specific cell types such as large excitatory pyramidal neurons. This might lead researchers to miss other subtle spiking modulations that occur in these regions during STM, and may explain at least some of the apparent discrepancies between single-unit and fMRI data. However, it seems unlikely that this type of issue would result in researchers missing highly significant and widespread single-unit spiking modulations, so if single-unit modulations are really so sparse during STM, then the implication is that the BOLD signal may simply be more sensitive to small changes in mean spike rates integrated over large expanses of cortex (or, as mentioned above, that fMRI is extremely sensitive to changes in sub-threshold modulations).

These potentially methodological differences between single-unit and fMRI studies aside, fMRI is known to be sensitive to a host of factors, including both spiking activity and LFPs (that influence vasodilation via indirect mechanisms: Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Logothetis & Wandell, 2004; Hillman, 2014; O'Herron et al., 2016). Thus tools such as MVPA applied in early visual areas may turn out to be driven in large part by LFPs induced by tonic top-down control signals that are generated by distal spiking activity in other cortical areas (such as the LPFC; Figure 8; Mendoza-Halliday et al. 2014). If this account is correct, then complementary information about both sub and supra-threshold modulations can be usefully assessed by combining information about local spiking activity using NHP models and information from fMRI studies in humans that can simultaneously measure response modulations across entire visual areas. By using these tools in tandem and employing the same behavioral paradigms in both NHPs and human subjects, researchers will hopefully be able to make more rapid and definitive statements about the importance of each of the candidate codes that might support STM.

Flexible coding of content-specific representations

Motivated in large part by the early work of Baddeley, most studies of STM have focused on dissociating the neural substrates of executive control and information storage (Baddeley, 1986; Baddeley & Hitch, 1974). This framework has proven remarkably successful, and the role of PFC in mediating executive control now seems well established. However, as more evidence is gathered suggesting that content-specific representations are observed in many different areas of cortex, it is becoming increasingly hard to definitively assign storage operations exclusively to either the PFC or to early visual areas. Unfortunately, the factors that govern exactly how and where content-specific information will be expressed in a given context are not yet entirely clear. Due to this uncertainty, and in an effort to foster discussions about how to productively move forward, I list below several avenues for future inquiry that may help to clarify—and perhaps start to reconcile – some of the discrepancies between current theories of information storage in STM.

First, there is the obvious, but often unacknowledged, fact that even subtly different STM tasks can impose very different demands on the precision of the mnemonic representations that are required to support successful behavior. In turn, these differences in task-demands may have a significant influence on where in cortex a content-specific memory representation is expressed. At one end of the spectrum, a subject might have to remember whether a stimulus belongs to one clearly defined category or another (e.g. was the remembered color red or blue? Or was it a face or a house?). In situations like this a relatively coarse and categorical representation should be adequate to support high behavioral performance, and a correspondingly coarse neural code based on activity in a relatively nonselective neuronal population might suffice. By a similar logic, coarse categorical memories for only a small set of objects might be more easily supported by a dynamically evolving code that transitions over the course of a delay period, as has been documented in PFC (Stokes et al., 2013). At the other end of the spectrum, a very demanding STM task might force subjects to use a more analog memory representation in order to differentiate between highly similar stimuli (e.g. is the orientation of the test stimulus $\pm 2^\circ$ from the sample stimulus). In this situation, the response properties of neurons in early visual cortex – which are highly selective for specific sensory features – seem ideally suited to support fine-grained discriminations between sample and test stimuli. In practice, most tasks fall somewhere along the continuum and future work would benefit from more formally considering the precision required by a given STM task. For example, models have been developed to describe different neural coding schemes that are required to support fine versus coarse discriminations, and these models might be used as a basis to determine the a priori feasibility of finding content-specific representations in a particular area of cortex (Butts & Goldman, 2006; Jazayeri & Movshon, 2006, 2007; Navalpakkam & Itti, 2007; Regan & Beverley, 1985; Scolari & Serences, 2009). Importantly, more carefully considering task-demands may play an important role in reconciling findings from studies that use human and NHPs. For the most part, studies using NHPs use relatively coarse discriminations, whereas studies using humans often, although not always, require more fine-grained discriminations. Thus, at least some of the variability in the brain structures that are thought to mediate the storage of information in STM might be attributable to changes in task demands. Finally, some might argue that equating task difficulty between tasks used in NHPs and tasks used in humans should be sufficient to ensure generalizability. However, equating difficulty alone does not ensure that analogous strategies are being used to solve the behavioral tasks, and by extension, very different neural mechanisms might be employed by each species (Birman & Gardner, 2015).

A second, and related point, concerns the impact of training and its relationship to species-specific differences in how information is represented in STM. NHPs are typically trained for far longer than their human-subject counterparts (many months at the low end for NHPs vs usually about 30 minutes to an hour in a typical human study). In turn, extended training might give rise to qualitative differences in how a STM task is performed (Birman & Gardner, 2015). In the domain of perceptual learning, for example, some reports suggest that activity in areas of parietal and frontal cortex is attenuated with training, consistent with the idea that demands on attentional control are reduced as perceptual expertise develops (Mukai et al., 2007; Sigman et al., 2005). Analogously, sufficiently long periods of training

that support the development of perceptual expertise may cause a task to transition from a categorical to an analog or more continuous representation, and this might influence where and how content-specific representations are stored. Thus, the amount of training that is typically required in NHP studies, along with the inability to self-report strategies, makes it challenging to evaluate exactly how NHPs are performing a given task. That said, recording from animals while they are still in the learning-phase of a complex cognitive task might provide a rich set of benchmark data to determine if or when spiking data collected in NHPs maps onto the patterns of modulation that are observed in human subjects.

Third, studies have begun to investigate how the presence of distractors in a delay period might alter the locus of memory representations, especially since data and theoretical considerations suggest that distractor resistance might be higher in PFC than in regions of early visual cortex (E. K. Miller et al., 1996; Riley & Constantinidis, 2015; Stokes, 2015). One solution is that the locus of mnemonic representations is flexible and determined by the expectation of distractors (Bettencourt & Xu, 2016). Alternatively, spiking activity is likely not the sole currency of STM, and other more dynamic codes or tonic feedback from PFC that produces sub-threshold modulations in early visual areas might provide alternate mechanisms for preserving high-fidelity information even in the presence of bottom-up interference. Understanding how the neural mechanisms that mediate STM can maintain distractor-resistant representations is thus a critical question for future research, and one that should be guided by considering how the visual system might be able to disentangle co-existing representations of remembered items and distractors (Orhan & Ma, 2015).

Finally, new knowledge about the different neural mechanisms that support memory could potentially be used to generate more specific predictions about the behavioral operating characteristics of STM in different experimental settings. The classic signature of STM, sustained spiking, is certainly intuitive and is one straightforward way to encode information for extended periods of time. However, both theoretical and empirical studies suggest that there are other mechanisms and computational principles that can support STM. For instance, the sub-threshold potentiation of feature-selective neurons in early visual cortex (Mendoza-Halliday et al., 2014) or the short-term modification of synaptic weights form two locally spike-silent codes that may provide a relatively efficient way to store information while minimizing metabolic expenditure (Sugase-Miyamoto et al. 2008; Mongillo et al., 2008). Importantly, each of these mechanisms might express different behavioral phenotypes if STM tasks are designed to target the timescales and hypothesized operating characteristics of each putative neural mechanism. While the feasibility of trying to dissect different mechanisms in this manner may or may not yield useful insights, explicitly acknowledging that there are multiple means of supporting STM might productively move the field away from the implicit assumption that all instances of storing information for less than ~20s relies on the operation of a common mechanism.

Conclusions

In addition to the suggestions outlined above, anecdotal evidence gathered primarily via informal discussions with colleagues suggests that another exciting movement is starting to take hold in the field of memory research. Dating back to the start of the cognitive revolution

in the 1950's, the field has been guided by the traditional Von Neumann computing architecture in which storage and computation are explicitly separated. However, the community is coming to recognize that separating storage and computation is highly inefficient: processing capacity has now grown to the point where computing speed is no longer limited solely by CPU speed, but instead by the power limitations imposed by systems that implement memory and computation in separate modules that must continuously exchange information. Hints that the human brain may have evolved from the very start in a manner that sidesteps this limitation are also becoming increasingly evident. For example, some of the models reviewed above suggest that short-term plasticity within visual cortex may simultaneously enable a population of feature-selective cells to store information and to implement a match/non-match decision-rule (e.g. Sugase-Miyamoto et al., 2008). In addition, others report that there is a high degree of local integration between the neural populations that provide executive control signals and neural populations that encode content-specific information about remembered stimuli. If true, then the utility of viewing computation and storage as separate modules may be limited. Instead many cognitive functions might be supported by more integrated neural mechanisms (e.g. 'multiplexing cells' in PFC), thus obviating some of the need to implement long-range coordination between distributed functional modules. While data supporting this type of architecture are sparse, remaining open to this non-traditional framework may help to gradually change how we understand the organization of the neural systems that support complex cognitive functions such as short-term memory.

Acknowledgments

Thanks to Thomas Sprague, Rosanne Rademaker, Timothy Gentner, Bradley Postle, Vy Vo and Kimye Zanessa. This work was supported by grants from the Nation Eye Institute (R01-EY025872, and R21-EY024733), and a Scholar Award from the James S. McDonnell Foundation.

References

- Atkinson, RC.; Shiffrin, RM. Human memory: A proposed system and its control processes. In: Spence, KW.; Spence, JT., editors. *The psychology of learning and motivation*. New York: Academic Press; 1968. p. 89-195.
- Awh E, Anllo-Vento L, Hillyard SA. The role of spatial selective attention in working memory for locations: evidence from event-related potentials. *J Cogn Neurosci*. 2000; 12(5):840–847. [PubMed: 11054925]
- Awh E, Jonides J. Overlapping mechanisms of attention and spatial working memory. *Trends Cogn Sci*. 2001; 5(3):119–126. [PubMed: 11239812]
- Awh E, Jonides J, Reuter-Lorenz PA. Rehearsal in spatial working memory. *J Exp Psychol Hum Percept Perform*. 1998; 24(3):780–790. [PubMed: 9627416]
- Awh E, Jonides J, Smith EE, Buxton RB, Frank LR, Love T, Gmeindl L. Rehearsal in spatial working memory: Evidence from neuroimaging. *Psychol Sci*. 1999; 10(5):433–437.
- Baddeley, A. *Working Memory*. New York: Oxford University Press; 1986.
- Baddeley, A.; Hitch, GJ. Working memory. In: Bower, G., editor. *Psychology of Learning and Motivation*. New York: Academic Press; 1974.
- Badre D, D'Esposito M. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat Rev Neurosci*. 2009; 10(9):659–669. [PubMed: 19672274]
- Barsalou LW. Grounded cognition. *Annu Rev Psychol*. 2008; 59:617–645. [PubMed: 17705682]
- Bettencourt KC, Xu Y. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nat Neurosci*. 2016; 19(1):150–157. [PubMed: 26595654]

- Birman D, Gardner JL. Parietal and prefrontal: categorical differences? *Nat Neurosci.* 2015; 19(1):5–7.
- Bisley JW, Zaksas D, Droll JA, Pasternak T. Activity of neurons in cortical area MT during a memory for motion task. *J Neurophysiol.* 2004; 91(1):286–300. [PubMed: 14523065]
- Brouwer GJ, Heeger DJ. Decoding and reconstructing color from responses in human visual cortex. *J Neurosci.* 2009; 29(44):13992–14003. [PubMed: 19890009]
- Brouwer GJ, Heeger DJ. Cross-orientation suppression in human visual cortex. *J Neurophysiol.* 2011; 106(5):2108–2119. [PubMed: 21775720]
- Butts DA, Goldman MS. Tuning curves, neuronal variability, and sensory coding. *PLoS Biol.* 2006; 4(4):e92. [PubMed: 16529529]
- Chao LL, Knight RT. Contribution of human prefrontal cortex to delay performance. *J Cogn Neurosci.* 1998; 10(2):167–177. [PubMed: 9555105]
- Chelazzi L, Duncan J, Miller EK, Desimone R. Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol.* 1998; 80(6):2918–2940. [PubMed: 9862896]
- Chelazzi L, Miller EK, Duncan J, Desimone R. A neural basis for visual search in inferior temporal cortex. *Nature.* 1993; 363(6427):345–347. [PubMed: 8497317]
- Christophel TB, Cichy RM, Hebart MN, Haynes JD. Parietal and early visual cortices encode working memory content across mental transformations. *Neuroimage.* 2015; 106:198–206. [PubMed: 25463456]
- Christophel TB, Haynes JD. Decoding complex flow-field patterns in visual working memory. *Neuroimage.* 2014; 91:43–51. [PubMed: 24480302]
- Christophel TB, Hebart MN, Haynes JD. Decoding the contents of visual short-term memory from human visual and parietal cortex. *J Neurosci.* 2012; 32(38):12983–12989. [PubMed: 22993415]
- Constantinidis C, Williams GV, Goldman-Rakic PS. A role for inhibition in shaping the temporal flow of information in prefrontal cortex. *Nat Neurosci.* 2002; 5(2):175–180. [PubMed: 11802172]
- Courtney SM, Petit L, Haxby JV, Ungerleider LG. The role of prefrontal cortex in working memory: examining the contents of consciousness. *Philos Trans R Soc Lond B Biol Sci.* 1998; 353(1377):1819–1828. [PubMed: 9854254]
- Courtney SM, Petit L, Maisog JM, Ungerleider LG, Haxby JV. An area specialized for spatial working memory in human frontal cortex. *Science.* 1998; 279(5355):1347–1351. [PubMed: 9478894]
- Courtney SM, Ungerleider LG, Keil K, Haxby JV. Transient and sustained activity in a distributed neural system for human working memory. *Nature.* 1997; 386(6625):608–611. [PubMed: 9121584]
- Curtis CE, D’Esposito M. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci.* 2003; 7(9):415–423. [PubMed: 12963473]
- Curtis CE, D’Esposito M. The effects of prefrontal lesions on working memory performance and theory. *Cogn Affect Behav Neurosci.* 2004; 4(4):528–539. [PubMed: 15849895]
- Curtis CE, Rao VY, D’Esposito M. Maintenance of spatial and motor codes during oculomotor delayed response tasks. *J Neurosci.* 2004; 24(16):3944–3952. [PubMed: 15102910]
- D’Esposito M. From cognitive to neural models of working memory. *Philos Trans R Soc Lond B Biol Sci.* 2007; 362(1481):761–772. [PubMed: 17400538]
- D’Esposito M, Postle BR. The cognitive neuroscience of working memory. *Annu Rev Psychol.* 2015; 66:115–142. [PubMed: 25251486]
- D’Esposito M, Postle BR, Rypma B. Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Exp Brain Res.* 2000; 133(1):3–11. [PubMed: 10933205]
- Emrich SM, Riggall AC, Larocque JJ, Postle BR. Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *J Neurosci.* 2013; 33(15):6516–6523. [PubMed: 23575849]
- Ester EF, Anderson DE, Serences JT, Awh E. A neural measure of precision in visual working memory. *J Cogn Neurosci.* 2013; 25(5):754–761. [PubMed: 23469889]
- Ester EF, Serences JT, Awh E. Spatially global representations in human primary visual cortex during working memory maintenance. *J Neurosci.* 2009; 29(48):15258–15265. [PubMed: 19955378]

- Ester EF, Sprague TC, Serences JT. Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*. 2015; 87(4):893–905. [PubMed: 26257053]
- Feredoes E, Heinen K, Weiskopf N, Ruff C, Driver J. 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*. 2011; 108(42):17510–17515. [PubMed: 21987824]
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*. 2001; 291(5502):312–316. [PubMed: 11209083]
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. *J Neurophysiol*. 2002; 88(2):929–941. [PubMed: 12163542]
- Funahashi S, Bruce CJ, Goldman-Rakic PS. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol*. 1989; 61(2):331–349. [PubMed: 2918358]
- Funahashi S, Bruce CJ, Goldman-Rakic PS. Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic “scotomas”. *J Neurosci*. 1993; 13(4):1479–1497. [PubMed: 8463830]
- Funahashi S, Chafee MV, Goldman-Rakic PS. Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature*. 1993; 365(6448):753–756. [PubMed: 8413653]
- Fuster JM. Memory networks in the prefrontal cortex. *Prog Brain Res*. 2000; 122:309–316. [PubMed: 10737067]
- Fuster JM, Alexander GE. Neuron activity related to short-term memory. *Science*. 1971; 173(3997):652–654. [PubMed: 4998337]
- Fuster JM, Bauer RH, Jervey JP. Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Res*. 1985; 330(2):299–307. [PubMed: 3986545]
- Fuster JM, Jervey JP. Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science*. 1981; 212(4497):952–955. [PubMed: 7233192]
- Goldman-Rakic, PS. Circuitry of the prefrontal cortex and the regulation of behavior by representational knowledge. In: Plum, F.; Mountcastle, VB., editors. *Handbook of Physiology: The Nervous System*. Bethesda, MD: American Physiological Society; 1987. p. 373-417.
- Goldman-Rakic PS. Cellular basis of working memory. *Neuron*. 1995; 14(3):477–485. [PubMed: 7695894]
- Harrison SA, Tong F. Decoding reveals the contents of visual working memory in early visual areas. *Nature*. 2009; 458(7238):632–635. [PubMed: 19225460]
- Higo T, Mars RB, Boorman ED, Buch ER, Rushworth MF. Distributed and causal influence of frontal operculum in task control. *Proc Natl Acad Sci U S A*. 2011; 108(10):4230–4235. [PubMed: 21368109]
- Hillman EMC. Coupling Mechanism and Significance of the BOLD Signal: A Status Report. *Annual Review of Neuroscience*. 2014; 37:161–181.
- Huang J, Sekuler R. Distortions in recall from visual memory: two classes of attractors at work. *J Vis*. 2010; 10(2):2421–2427.
- Irwin DE. Information integration across saccadic eye movements. *Cogn Psychol*. 1991; 23(3):420–456. [PubMed: 1884598]
- James, W. *The Principles of Psychology*. New York: Henry Hold and Company; 1890.
- Jazayeri M, Movshon JA. Optimal representation of sensory information by neural populations. *Nat Neurosci*. 2006; 9(5):690–696. [PubMed: 16617339]
- Jazayeri M, Movshon JA. A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*. 2007; 446(7138):912–915. [PubMed: 17410125]
- Jerde TA, Curtis CE. Maps of space in human frontoparietal cortex. *J Physiol Paris*. 2013; 107(6):510–516. [PubMed: 23603831]
- Jerde TA, Merriam EP, Riggall AC, Hedges JH, Curtis CE. Prioritized maps of space in human frontoparietal cortex. *J Neurosci*. 2012; 32(48):17382–17390. [PubMed: 23197729]
- Lara AH, Wallis JD. The Role of Prefrontal Cortex in Working Memory: A Mini Review. *Front Syst Neurosci*. 2015; 9:173. [PubMed: 26733825]

- LaRocque JJ, Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR. Decoding attended information in short-term memory: an EEG study. *J Cogn Neurosci*. 2013; 25(1):127–142. [PubMed: 23198894]
- Larocque JJ, Lewis-Peacock JA, Postle BR. Multiple neural states of representation in short-term memory? It's a matter of attention. *Front Hum Neurosci*. 2014; 8:5. [PubMed: 24478671]
- Lee SH, Kravitz DJ, Baker CI. Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat Neurosci*. 2013; 16(8):997–999. [PubMed: 23817547]
- Lee TG, D'Esposito M. The dynamic nature of top-down signals originating from prefrontal cortex: a combined fMRI-TMS study. *J Neurosci*. 2012; 32(44):15458–15466. [PubMed: 23115183]
- Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR. Neural evidence for a distinction between short-term memory and the focus of attention. *J Cogn Neurosci*. 2012; 24(1):61–79. [PubMed: 21955164]
- Liebe S, Hoerzer GM, Logothetis NK, Rainer G. Theta coupling between V4 and prefrontal cortex predicts visual short-term memory performance. *Nat Neurosci*. 2012; 15(3):456–462. S451-452. [PubMed: 22286175]
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. *Nature*. 2001; 412(6843):150–157. [PubMed: 11449264]
- Logothetis NK, Wandell BA. Interpreting the BOLD signal. *Annu Rev Physiol*. 2004; 66:735–769. [PubMed: 14977420]
- Luck SJ, Vogel EK. Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn Sci*. 2013; 17(8):391–400. [PubMed: 23850263]
- Lui LL, Pasternak T. Representation of comparison signals in cortical area MT during a delayed direction discrimination task. *J Neurophysiol*. 2011; 106(3):1260–1273. [PubMed: 21676932]
- Ma WJ, Husain M, Bays PM. Changing concepts of working memory. *Nat Neurosci*. 2014; 17(3):347–356. [PubMed: 24569831]
- Magnussen S. Low-level memory processes in vision. *Trends Neurosci*. 2000; 23(6):247–251. [PubMed: 10838593]
- Magnussen S, Greenlee MW. Retention and disruption of motion information in visual short-term memory. *J Exp Psychol Learn Mem Cogn*. 1992; 18(1):151–156. [PubMed: 1532017]
- Magnussen S, Greenlee MW. The psychophysics of perceptual memory. *Psychol Res*. 1999; 62(2–3): 81–92. [PubMed: 10472196]
- Magnussen S, Greenlee MW, Asplund R, Dyrnes S. Stimulus-specific mechanisms of visual short-term memory. *Vision Res*. 1991; 31(7–8):1213–1219. [PubMed: 1891813]
- Mante V, Sussillo D, Shenoy KV, Newsome WT. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*. 2013; 503(7474):78–84. [PubMed: 24201281]
- Martinez-Trujillo JC, Treue S. Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol*. 2004; 14(9):744–751. [PubMed: 15120065]
- McKee JL, Riesenhuber M, Miller EK, Freedman DJ. Task dependence of visual and category representations in prefrontal and inferior temporal cortices. *J Neurosci*. 2014; 34(48):16065–16075. [PubMed: 25429147]
- McNab F, Klingberg T. Prefrontal cortex and basal ganglia control access to working memory. *Nat Neurosci*. 2008; 11(1):103–107. [PubMed: 18066057]
- Mendoza-Halliday D, Torres S, Martinez-Trujillo JC. Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat Neurosci*. 2014; 17(9):1255–1262. [PubMed: 25108910]
- Miller BT, Vytlačil J, Fegen D, Pradhan S, D'Esposito M. The prefrontal cortex modulates category selectivity in human extrastriate cortex. *J Cogn Neurosci*. 2011; 23(1):1–10. [PubMed: 20586702]
- Miller EK. The “working” of working memory. *Dialogues Clin Neurosci*. 2013; 15(4):411–418. [PubMed: 24459408]
- Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*. 2001; 24:167–202. [PubMed: 11283309]
- Miller EK, Erickson CA, Desimone R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci*. 1996; 16(16):5154–5167. [PubMed: 8756444]

- Miller EK, Li L, Desimone R. A neural mechanism for working and recognition memory in inferior temporal cortex. *Science*. 1991; 254(5036):1377–1379. [PubMed: 1962197]
- Miller EK, Li L, Desimone R. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J Neurosci*. 1993; 13(4):1460–1478. [PubMed: 8463829]
- Miller, GA.; Galanter, E.; Pribham, KH. *Plans and the Structure of Behavior*. New York: Hold, Rinehart and Winston; 1960.
- Miyashita Y, Chang HS. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*. 1988; 331(6151):68–70. [PubMed: 3340148]
- Mongillo G, Barak O, Tsodyks M. Synaptic theory of working memory. *Science*. 2008; 319(5869): 1543–1546. [PubMed: 18339943]
- Mukai I, Kim D, Fukunaga M, Japee S, Marrett S, Ungerleider LG. Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. *J Neurosci*. 2007; 27(42): 11401–11411. [PubMed: 17942734]
- Myers NE, Rohenkohl G, Wyart V, Woolrich MW, Nobre AC, Stokes MG. Testing sensory evidence against mnemonic templates. *Elife*. 2015; 4:e09000. [PubMed: 26653854]
- Nakamura K, Kubota K. Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *J Neurophysiol*. 1995; 74(1):162–178. [PubMed: 7472321]
- Naselaris T, Kay KN. Resolving Ambiguities of MVPA Using Explicit Models of Representation. *Trends Cogn Sci*. 2015; 19(10):551–554. [PubMed: 26412094]
- Naselaris T, Kay KN, Nishimoto S, Gallant JL. Encoding and decoding in fMRI. *Neuroimage*. 2011; 56(2):400–410. [PubMed: 20691790]
- Navalpakkam V, Itti L. Search goal tunes visual features optimally. *Neuron*. 2007; 53(4):605–617. [PubMed: 17296560]
- Nemes VA, Parry NR, Whitaker D, McKeefry DJ. The retention and disruption of color information in human short-term visual memory. *J Vis*. 2012; 12(1):26.
- Nyberg L, Habib R, McIntosh AR, Tulving E. Reactivation of encoding-related brain activity during memory retrieval. *Proc Natl Acad Sci U S A*. 2000; 97(20):11120–11124. [PubMed: 11005878]
- O’Craven KM, Kanwisher N. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J Cogn Neurosci*. 2000; 12(6):1013–1023. [PubMed: 11177421]
- O’Herron P, Chhatbar PY, Levy M, Shen Z, Schramm AE, Lu Z, Kara P. Neural correlates of single-vessel haemodynamic responses in vivo. *Nature*. 2016; 534:378–382. [PubMed: 27281215]
- Orhan AE, Ma WJ. Neural population coding of multiple stimuli. *J Neurosci*. 2015; 35(9):3825–3841. [PubMed: 25740513]
- Pasternak T, Greenlee MW. Working memory in primate sensory systems. *Nat Rev Neurosci*. 2005; 6(2):97–107. [PubMed: 15654324]
- Petit L, Courtney SM, Ungerleider LG, Haxby JV. Sustained activity in the medial wall during working memory delays. *J Neurosci*. 1998; 18(22):9429–9437. [PubMed: 9801381]
- Postle, BR. Activation and information in working memory research. In: Duarte, A.; Shields, M.; Addis, DR., editors. *The Wiley-Blackwell Handbook on the Cognitive Neuroscience of Memory*. Oxford: Wiley-Blackwell; 2015.
- Postle BR, Berger JS, Taich AM, D’Esposito M. Activity in human frontal cortex associated with spatial working memory and saccadic behavior. *J Cogn Neurosci*. 2000; (12 Suppl 2):2–14. [PubMed: 11506643]
- Postle BR, D’Esposito M. “What”-Then-Where” in visual working memory: an event-related fMRI study. *J Cogn Neurosci*. 1999; 11(6):585–597. [PubMed: 10601740]
- Pratte MS, Tong F. Spatial specificity of working memory representations in the early visual cortex. *J Vis*. 2014; 14(3):22.
- Pribham, KH.; Ahumada, A.; Hartog, J.; Roos, L. A progress report on the neurological processes disturbed by frontal lesions in primates. In: Warren, JM.; Akert, K., editors. *The Frontal Cortex and Behavior*. New York: McGraw-Hill; 1964.
- Qi XL, Katsuki F, Meyer T, Rawley JB, Zhou X, Douglas KL, Constantinidis C. Comparison of neural activity related to working memory in primate dorsolateral prefrontal and posterior parietal cortex. *Front Syst Neurosci*. 2010; 4:12. [PubMed: 20514341]

- Qi XL, Meyer T, Stanford TR, Constantinidis C. Changes in prefrontal neuronal activity after learning to perform a spatial working memory task. *Cereb Cortex*. 2011; 21(12):2722–2732. [PubMed: 21527786]
- Rademaker RL, Bloem IM, De Weerd P, Sack AT. The impact of interference on short-term memory for visual orientation. *J Exp Psychol Hum Percept Perform*. 2015; 41(6):1650–1665. [PubMed: 26371383]
- Ranganath C, D’Esposito M. Directing the mind’s eye: prefrontal, inferior and medial temporal mechanisms for visual working memory. *Curr Opin Neurobiol*. 2005; 15(2):175–182. [PubMed: 15831399]
- Ranganath C, DeGutis J, D’Esposito M. Category-specific modulation of inferior temporal activity during working memory encoding and maintenance. *Brain Res Cogn Brain Res*. 2004; 20(1):37–45. [PubMed: 15130587]
- Regan D, Beverley KI. Postadaptation orientation discrimination. *J Opt Soc Am A*. 1985; 2(2):147–155. [PubMed: 3973752]
- Riggall AC, Postle BR. The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *J Neurosci*. 2012; 32(38):12990–12998. [PubMed: 22993416]
- Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, Fusi S. The importance of mixed selectivity in complex cognitive tasks. *Nature*. 2013; 497(7451):585–590. [PubMed: 23685452]
- Riley MR, Constantinidis C. Role of Prefrontal Persistent Activity in Working Memory. *Front Syst Neurosci*. 2015; 9:181. [PubMed: 26778980]
- Rolfs M. Attention in Active Vision: A Perspective on Perceptual Continuity Across Saccades. *Perception*. 2015; 44(8–9):900–919. [PubMed: 26562908]
- Schumacher EH, Lauber E, Awh E, Jonides J, Smith EE, Koeppel RA. PET evidence for an amodal verbal working memory system. *Neuroimage*. 1996; 3(2):79–88. [PubMed: 9345478]
- Scolari M, Byers A, Serences JT. Optimal deployment of attentional gain during fine discriminations. *J Neurosci*. 2012; 32(22):7723–7733. [PubMed: 22649250]
- Scolari M, Serences JT. Adaptive allocation of attentional gain. *J Neurosci*. 2009; 29(38):11933–11942. [PubMed: 19776279]
- Serences JT, Ester EF, Vogel EK, Awh E. Stimulus-specific delay activity in human primary visual cortex. *Psychol Sci*. 2009; 20(2):207–214. [PubMed: 19170936]
- Serences JT, Saproo S. Computational advances towards linking BOLD and behavior. *Neuropsychologia*. 2012; 50(4):435–446. [PubMed: 21840553]
- Sigman M, Pan H, Yang Y, Stern E, Silbersweig D, Gilbert CD. Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron*. 2005; 46(5):823–835. [PubMed: 15924867]
- Silvanto J, Cattaneo Z. Transcranial magnetic stimulation reveals the content of visual short-term memory in the visual cortex. *Neuroimage*. 2010; 50(4):1683–1689. [PubMed: 20079448]
- Smith EE, Jonides J. Storage and executive processes in the frontal lobes. *Science*. 1999; 283(5408):1657–1661. [PubMed: 10073923]
- Smith EE, Jonides J, Koeppel RA, Awh E, Schumacher EH, Minoshima S. Spatial versus Object Working Memory: PET Investigations. *J Cogn Neurosci*. 1995; 7(3):337–356. [PubMed: 23961865]
- Sprague TC, Ester EF, Serences JT. Reconstructions of information in visual spatial working memory degrade with memory load. *Curr Biol*. 2014; 24(18):2174–2180. [PubMed: 25201683]
- Sprague TC, Saproo S, Serences JT. Visual attention mitigates information loss in small- and large-scale neural codes. *Trends Cogn Sci*. 2015; 19(4):215–226. [PubMed: 25769502]
- Sprague, TC.; Serences, JT. Using human neuroimaging to examine top-down modulations of visual perception. In: Forstmann, BU.; Wagenmakers, EJ., editors. *An introduction to model-based cognitive neuroscience*. New York: Springer-Verlag; 2015.
- Sreenivasan KK, Curtis CE, D’Esposito M. Revisiting the role of persistent neural activity during working memory. *Trends Cogn Sci*. 2014; 18(2):82–89. [PubMed: 24439529]

- Stokes MG. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci*. 2015; 19(7):394–405. [PubMed: 26051384]
- Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J. Dynamic coding for cognitive control in prefrontal cortex. *Neuron*. 2013; 78(2):364–375. [PubMed: 23562541]
- Sugase-Miyamoto Y, Liu Z, Wiener MC, Optican LM, Richmond BJ. Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS Comput Biol*. 2008; 4(5):e1000073. [PubMed: 18464917]
- Super H, Spekreijse H, Lamme VA. A neural correlate of working memory in the monkey primary visual cortex. *Science*. 2001; 293(5527):120–124. [PubMed: 11441187]
- Tong F, Pratte MS. Decoding patterns of human brain activity. *Annu Rev Psychol*. 2012; 63:483–509. [PubMed: 21943172]
- van de Ven V, Jacobs C, Sack AT. Topographic contribution of early visual cortex to short-term memory consolidation: a transcranial magnetic stimulation study. *J Neurosci*. 2012; 32(1):4–11. [PubMed: 22219265]
- van den Berg R, Awh E, Ma WJ. Factorial comparison of working memory models. *Psychol Rev*. 2014; 121(1):124–149. [PubMed: 24490791]
- Wallis JD, Anderson KC, Miller EK. Single neurons in prefrontal cortex encode abstract rules. *Nature*. 2001; 411(6840):953–956. [PubMed: 11418860]
- Wallis JD, Miller EK. From rule to response: neuronal processes in the premotor and prefrontal cortex. *J Neurophysiol*. 2003; 90(3):1790–1806. [PubMed: 12736235]
- Wang XJ. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci*. 2001; 24(8):455–463. [PubMed: 11476885]
- Warden MR, Miller EK. Task-dependent changes in short-term memory in the prefrontal cortex. *J Neurosci*. 2010; 30(47):15801–15810. [PubMed: 21106819]
- White IM, Wise SP. Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res*. 1999; 126(3):315–335. [PubMed: 10382618]
- Xu Y, Chun MM. Selecting and perceiving multiple visual objects. *Trends Cogn Sci*. 2009; 13(4):167–174. [PubMed: 19269882]
- Yantis S, Jonides J. Abrupt visual onsets and selective attention: evidence from visual search. *J Exp Psychol Hum Percept Perform*. 1984; 10(5):601–621. [PubMed: 6238122]
- Zaksas D, Pasternak T. Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J Neurosci*. 2006; 26(45):11726–11742. [PubMed: 17093094]
- Zanto TP, Rubens MT, Thangavel A, Gazzaley A. Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nat Neurosci*. 2011; 14(5):656–661. [PubMed: 21441920]
- Zhang W, Luck SJ. Discrete fixed-resolution representations in visual working memory. *Nature*. 2008; 453(7192):233–235. [PubMed: 18385672]

Highlights

- Prefrontal cortex (PFC) controls what information is stored in working memory (WM)
- Stored information is maintained in early areas of visual cortex
- Likely more flexible: content-specific WM representations in PFC and visual cortex
- This flexibility is perhaps achieved by employing spiking and non-spiking codes

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

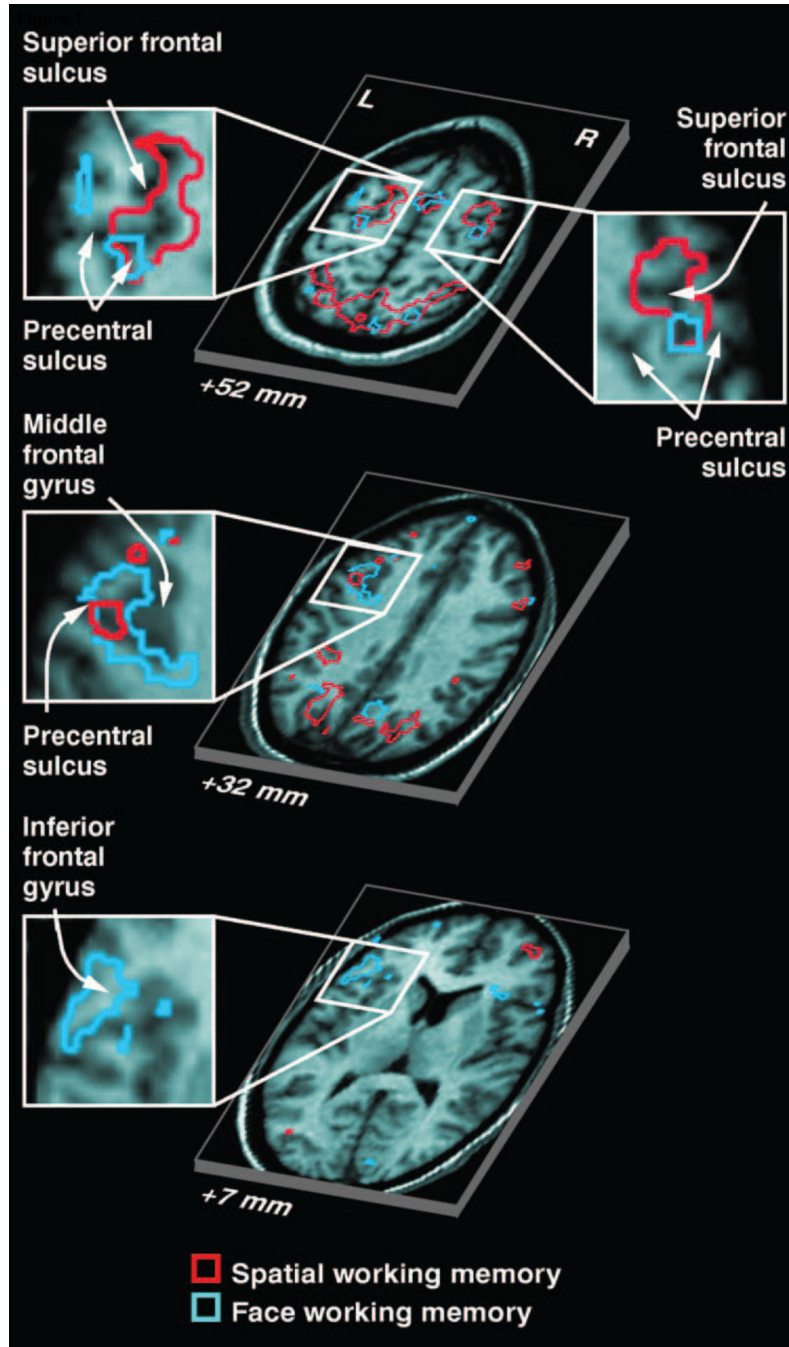
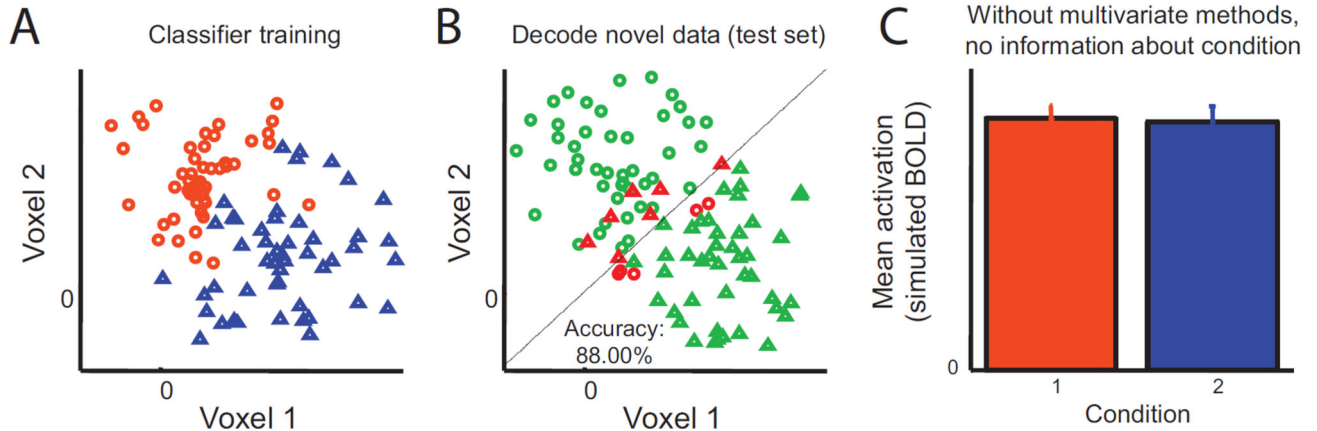


Figure 1. (A) Areas in PFC that showed elevated delay period activation as measured with fMRI during memory for either face identity or a spatial location. Partially segregated areas were observed that responded more during STM for one type of content versus the other (From Courtney et al., 1998). Reprinted with permission from the authors and the original publisher.

Multivariate decoding

Region level activation

**Figure 2.**

A hypothetical dataset in which activation across a “region” of 2 voxels carries information which can discriminate between 2 conditions, but in which the mean activation level is equivalent across the two voxels. (A) Scatter plot of activation in each voxel for each trial, color-coded based on their condition. In order to identify which of two conditions (condition 1, orange circles, condition 2, blue triangles) the activation from a given trial corresponds to, a linear classifier can be trained to find a line which best discriminates conditions 1 and 2. Note that a line is used because this is a 2-dimensional space; a hyperplane would be used in a more realistic situation where the pattern extended across many voxels. (B) An independent test set is then used to evaluate how well this decision rule discriminates between the two conditions. Trials in the test set known to be from condition 1 (circles) and condition 2 (triangles) are color-coded based on whether they are accurately classified (green is correct, red is incorrect). (C) Without this type of multivariate analysis, this 2-voxel activation pattern would be assumed to carry no information about which condition a trial belongs to. Adapted from Sprague and Serences (2015). Reprinted with permission from the authors and the original publisher.

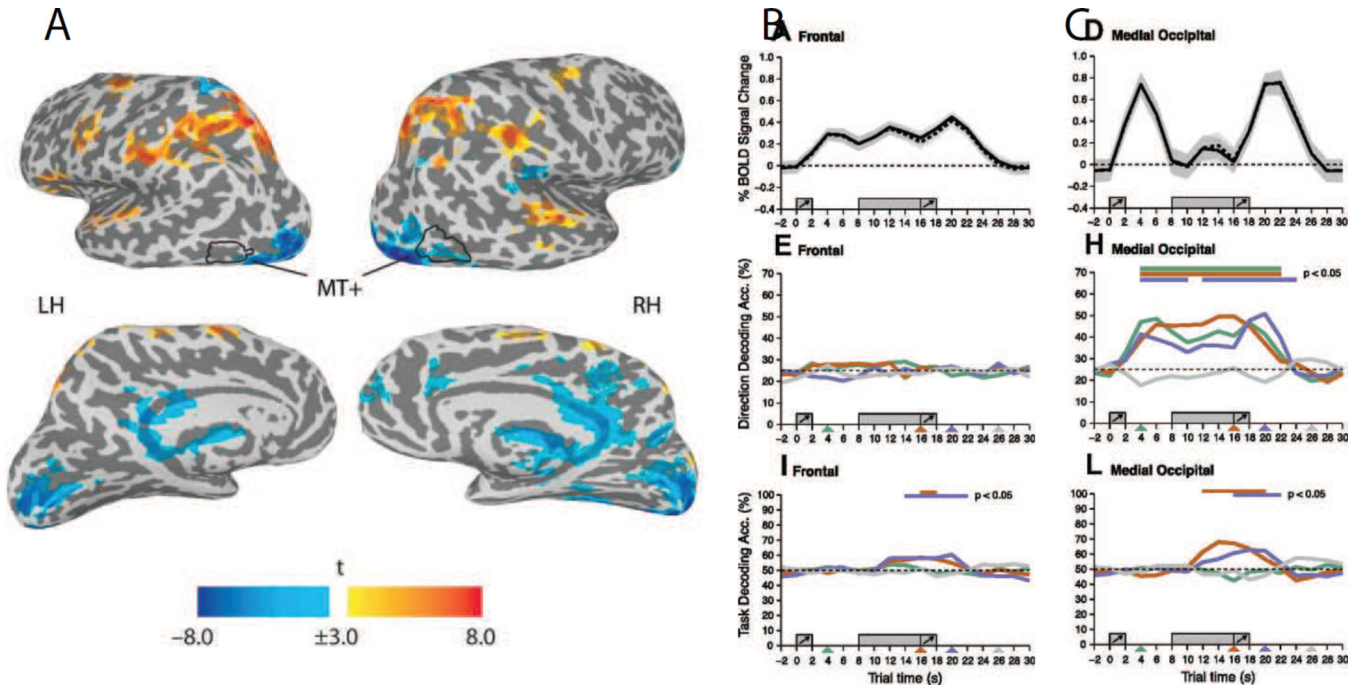
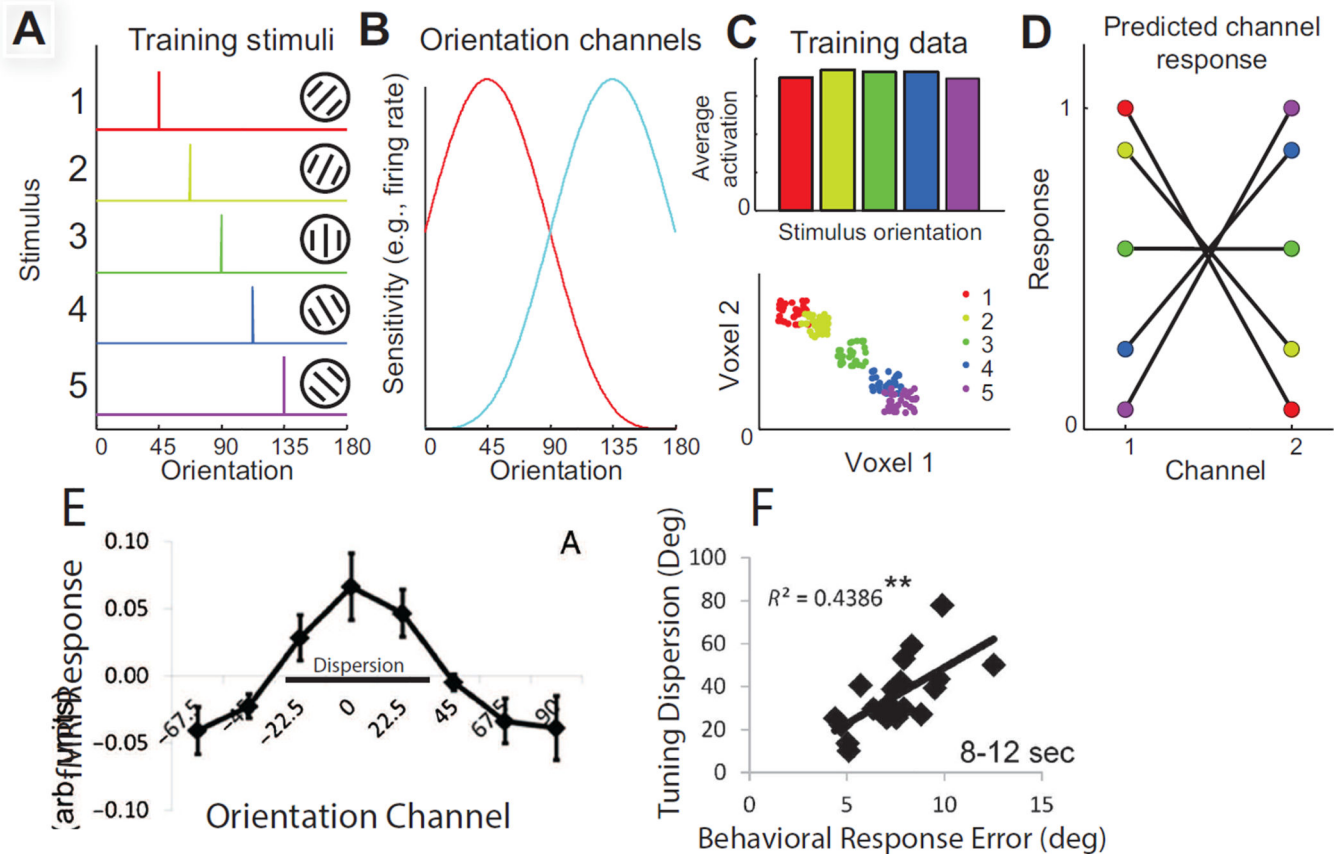


Figure 3.

(A) Areas showing significant univariate fMRI responses during the delay period of a STM task for motion direction or for motion speed. Areas in yellow-to-red showed a sustained response increase during the delay period, whereas areas in blue showed a decreased response during the delay period. (B) Top panel: timecourse of activation showing a sustained response during the delay period in regions of interest (ROIs) in the PFC. Middle panel: MVPA-based decoding accuracy for remembered features (motion direction or speed) in the PFC ROI. Chance is ~25%. Bottom panel: decoding accuracy for the task rule indicating whether subjects should remember the motion direction or the motion speed (chance decoding is 50%). (C) Analogous data from a medial occipital cortex that does not show an increased response during the delay period. Note that decoding accuracy for the contents of memory was near chance in PFC and well-above chance in occipital visual cortex. In contrast, both regions encoded information about the task-set (i.e. whether subjects should remember motion direction or speed). These data show a striking dissociation between sustained increases in activation and the amount of information encoded about specific memoranda. All error bars reflect SEM across subjects. Adapted from Riggall and Postle, 2012. Reprinted with permission from the authors and the original publisher.

**Figure 4.**

(A–D) Schematic of inverted encoding models (IEMs) for reconstructing representations of the stimulus. The response in each voxel is modeled using linear regression to estimate the magnitude of the response in different information channels that correspond to hypothesized response properties of underlying neural populations (here two orientation selective tuning functions are shown, and these would be used as the basis set to model the response of each voxel). Once the weight on each channel is computed using an independent training set of data, novel test data collected in each voxel can be mapped back into the space of the information channels, which effectively forms a reconstruction of the remembered stimulus. Illustrated here is a schematic with just two voxels and a 2-channel encoding model for stimulus orientation. (A) shows the set of stimuli used to train the encoding model, and (B) shows the orientation selective basis functions. (C) The response to each of the 5 oriented stimuli is then measured in each voxel, and used to estimate the response in each information channel (shown in D). As in MVPA, you can have an equal average response to each of the 5 stimuli within the region of interest, but there can still be a pattern of activation across the voxels that carries information about the remembered feature (lower panel, C). (E) Reconstruction of a remembered orientation from areas V1 and V2 during the delay period of a recall STM task. (F) Correlations between the bandwidth (dispersion or inverse of precision) of individual subject orientation reconstructions and their behaviorally assessed memory recall performance. All error bars reflect SEM across subjects. Panels A-D from

Sprague and Serences, 2015, and panels E-F from Ester et al., 2013. Reprinted with permission from the authors and the original publisher.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

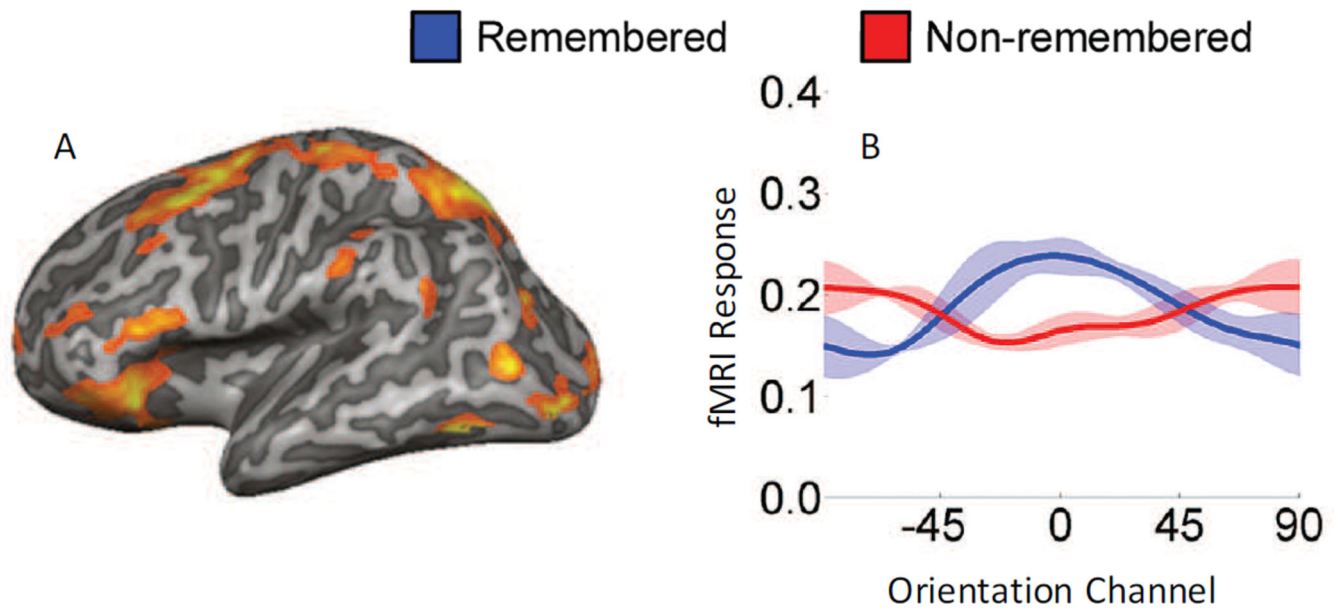


Figure 5.

(A) Sample data from one hemisphere from one subject showing areas that encoded information about remembered orientations in a STM recall task. (B) Reconstructions of remembered orientations from a region in the right dorsal lateral PFC showing a content specific representation of the remembered orientation, but relatively little information about a simultaneously presented non-remembered orientation. Error bars reflect within-subject SEM.

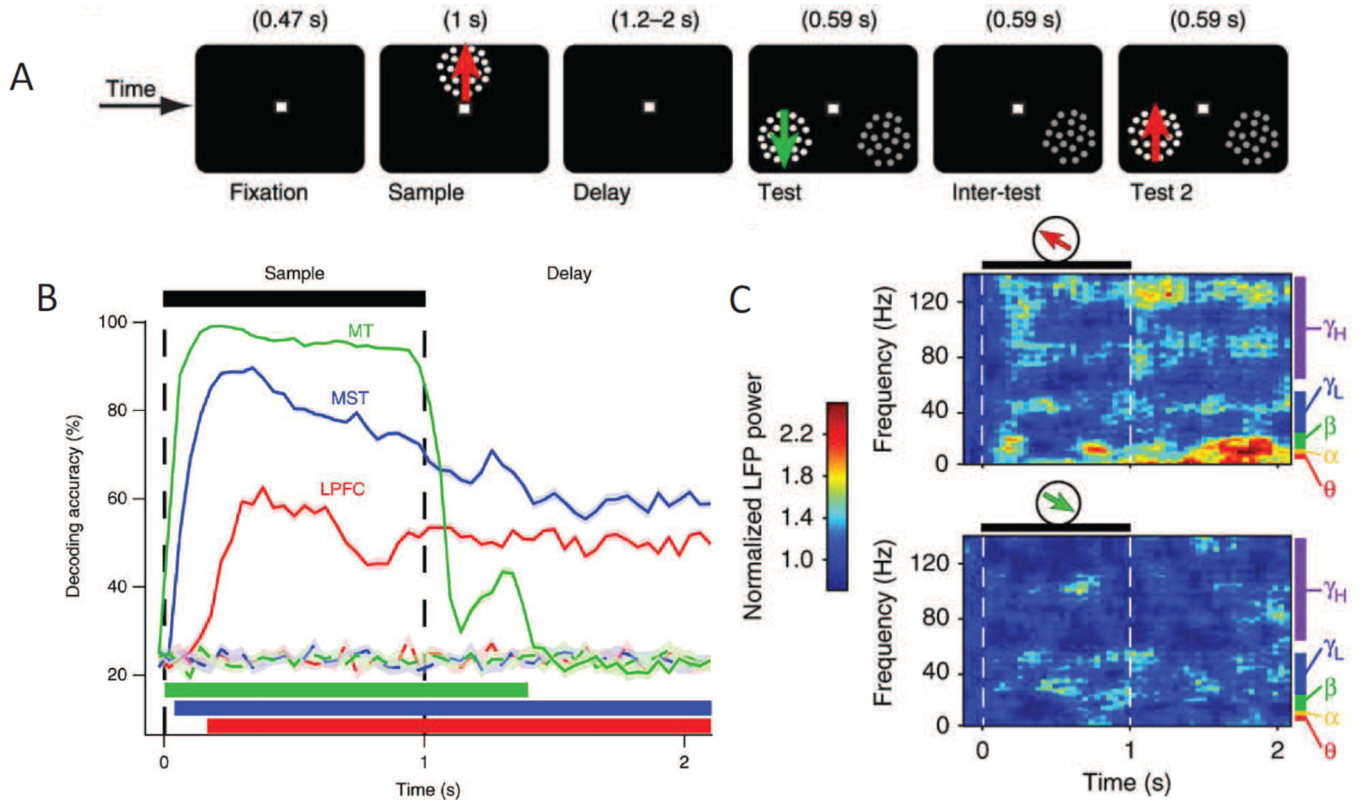


Figure 6. (A) Delayed-match to sample task in which the direction of a sample had to be remembered and then used to guide a saccadic eye movement to the matching test stimulus at the end of the delay period. Note that the monkeys could not plan a saccade before the end of the delay period because the location of the matching test stimulus was unknown in advance. (B) Multivariate patterns of spiking activity in MT, MST and IPFC all decoded the direction of the sample stimulus while it was present on the screen, but only spiking patterns in MST and IPFC decoded the remembered direction for the duration of the delay period. (C) Even though spiking activity in MT was not direction selective during the delay period, there was a significant direction selective modulation of LFPs in MT, especially in lower frequency bands. Error bars reflect SEM. Adapted from Mendoza-Halliday et al., 2014. Reprinted with permission from the authors and the original publisher.

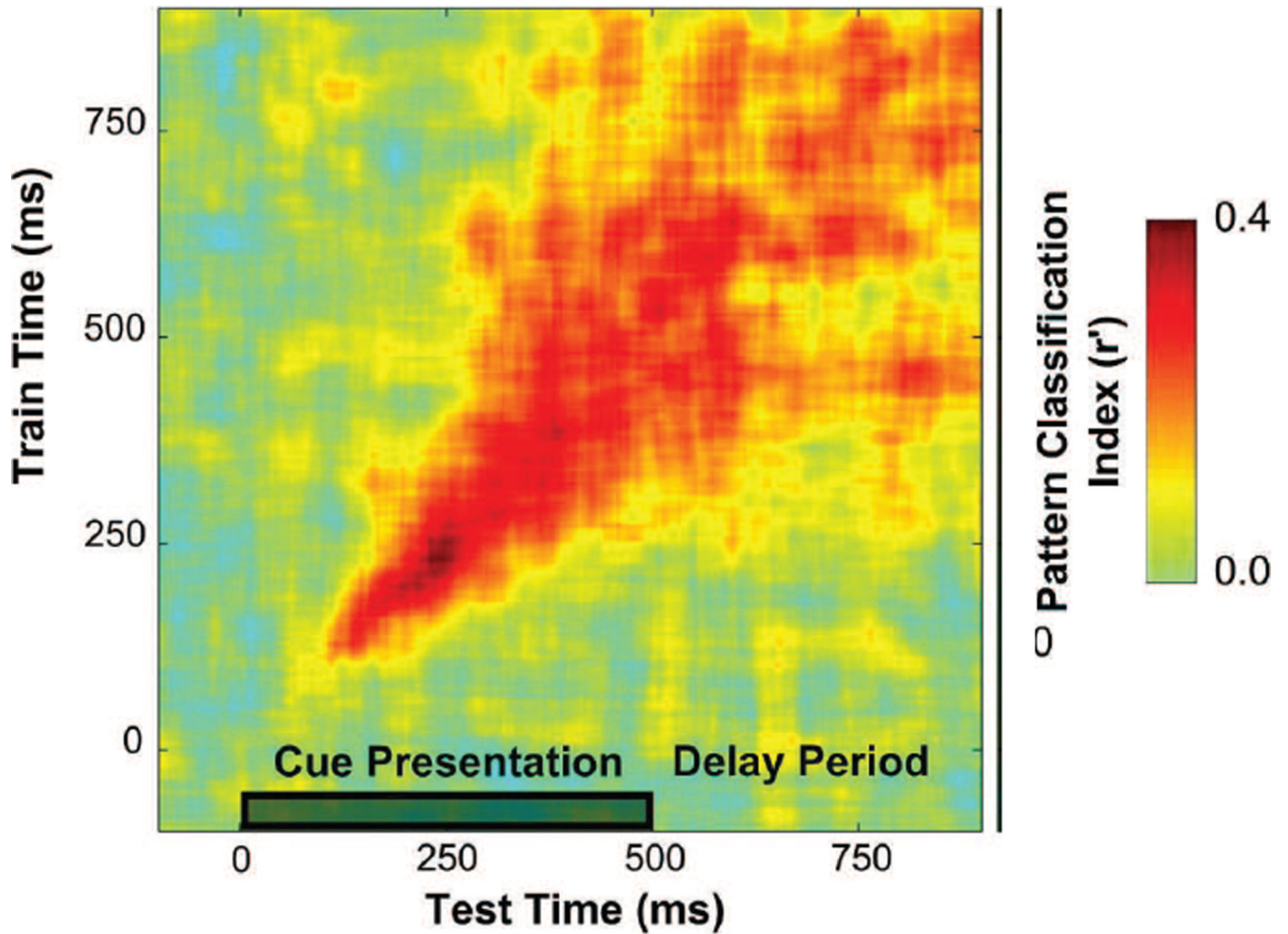


Figure 7.

Index of population-level information about a memory cue when a classifier is cross-trained and tested on different time bins. Early in the trial, during the presentation of the memory cue, there is a highly selective pattern as evidenced by relatively strong on-diagonal classification accuracy but poor off-diagonal classification accuracy. However, after several hundred msec, the pattern stabilizes into a more generalizable code (as indexed by more off-diagonal information). This dynamic transition in the multivariate code over the course of a trial runs counter to the traditional notion that sustained and stable spiking activity over the entire course of the delay period is the signature pattern of a code for STM. Reprinted with permission from the authors and the original publisher.

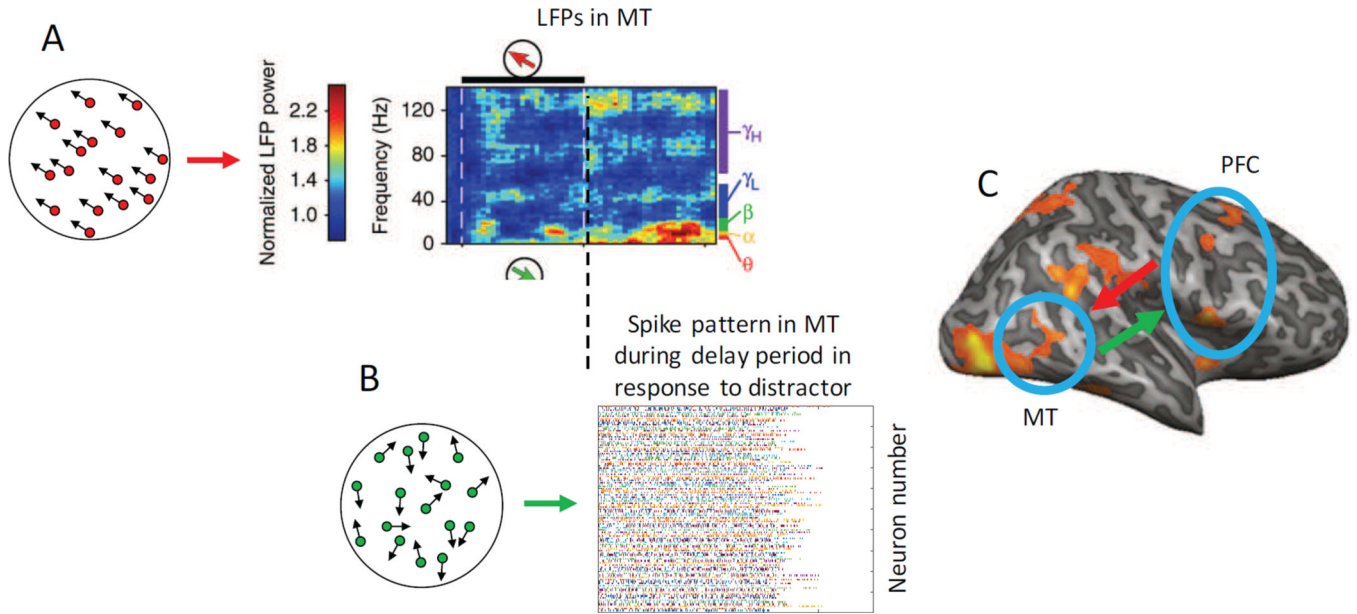


Figure 8.

(A) The presentation of a memory sample stimulus evokes an initial content-specific pattern of spiking activity in MT, followed by a cessation of content-specific spiking activity during the delay period (see Figure 6). However, sustained feedback signals in the form of spike-field coherence from PFC to MT induce a stable sub-spiking-threshold LFP modulation in MT that exhibits selectivity for the remembered motion direction, particularly in the lower frequency bands (shown in red, data from Mendoza-Halliday et al., 2014). (B) During the delay period, distracting stimuli might still trigger spiking activity in MT that is unrelated to the identity of the remembered motion direction. However, the sustained top-down input from PFC might maintain neurons in MT in a ‘primed’ state that will in turn make content-specific spiking patterns more likely to re-emerge upon the presentation of the test stimulus. (C) Bi-directional flow of information between visual areas and PFC under this general account. Bottom-up spiking activity from the presentation of any sensory stimulus (sample, test, or distractor) feeds-forward into PFC, which in turn provides a sustained top-down signal that biases MT neurons in a feature-selective manner that is consistent with the relevant remembered motion direction. This effectively holds MT in a primed state in preparation for the presentation of the test stimulus at the end of the delay interval. Using this architecture, the tonic top-down biasing signals to MT might attenuate the potentially negative overwriting effects of intervening distractors, even in the absence of a sustained content-specific spiking code in MT. Some panels reprinted with permission from the author and the original publisher (Mendoza-Halliday et al., 2014).