In press at *Attention & Performance XXV: Mechanisms of Sensory Working Memory*,
Jolicoeur, Martinez-Trujillo, and Lefebvre, Eds.

Neural Bases of the Short-Term Retention of Visual Information

Bradley R. Postle
Depts. of Psychology and Psychiatry, University of Wisconsin-Madison

The sensory-recruitment hypothesis holds that the short-term retention of visual information is accomplished by the same networks responsible for the real-time perceptual analysis of such information. A corollary of this view is that the elevated delay-period activity in frontal and parietal cortex that are often observed during tests of short-term and working memory may not correspond to the information storage, per se, of information. Recent studies applying multivariate analyses to functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) data from humans, and to extracellular recordings from monkeys, have generated findings consistent with both of these ideas. At the level of mechanism, some of these studies have also called into question the longstanding assumption that sustained activity is the basis for the short-term retention of information. In view of this, alternative ideas based on transient synaptic weight-based mechanisms for neural representation are considered. We are at the precipice of exciting, potentially revolutionary breakthroughs in our understanding of how the brain supports short-term and working-memory behaviors.

Total word count: 10,748

1202 West Johnson St.
Madison, WI   53726
USA
tel: +1 608-262-4330
fax: +1 608-262-4029
postle@wisc.edu

Preamble: Defining Concepts and Terminology

To make progress in any scientific endeavor, it is necessary for there to be conceptual clarity about the parameters of the phenomenon being studied, and for there to be terminological clarity such that there is an unambiguous mapping between the concepts and the words that are used to describe them. Although considerable progress has been made in our understanding of the cognitive and neural factors underlying of the behaviors that fall under the rubric of "working memory", this progress has sometimes been hindered by a lack of systematicity at the terminological level that has, in turn, sometimes promoted a lack of conceptual clarity. Within the theme "Mechanisms of Sensory Working Memory", this chapter will focus on the neural processes involved in the short-term retention (STR) of visual information – the neural representation of visual information when that information is no longer present in the environment. Discussions at the 25th International Symposium on Attention and Performance made it clear that this will be done most effectively if we first take the time to disambiguate the STR of information from two related concepts and their labels – short-term memory (STM) and working memory (WM).

STR vs. STM vs. WM

In this chapter, the terms "STM" and "WM" will be used to refer to classes of behavior, and to tasks that measure performance and ability within these classes of behavior. Importantly, they will not be used to refer to hypothetical cognitive or neural mechanisms -- further along in this chapter, it will be illustrated how their use in this latter way can lead to a lack of conceptual clarity and, sometimes, to erroneous inferences about neural functioning. Tests of STM entail the presentation of a small amount of to-be-remembered information (referred to alternately as the "target(s)", or as the "sample(s)"), followed by a delay that can last from 100s of msec to 10s of sec or longer, followed by a second stimulus that requires a response that is related to the first. On tests of delayed response, the second stimulus is a cue to execute the action that had been signaled by the target. On tests of delayed recall, the second stimulus cues the subject to reproduce the target(s). On tests of delayed recognition, the second stimulus (the "probe") requires Y/N recognition, n-alternative forced choice (if > 1 probes are presented), or some other recognition decision about its match to the first stimulus. Thus, STM results from the STR of information that was recently either presented or cued, but is not currently present in the environment, and it is tested by one's ability to guide behavior with that information.

Theoretical bases for the STM-WM distinction

The concept of WM as distinct from STM was proposed and developed by Baddeley and Hitch (1974), in part due to their findings that performance on each of two STM tasks under dual-task conditions could approach levels of performance that one would see when the tasks were administered individually, so long as the two engaged different domains of information[1]. Because, in their early research, the two domains were verbal and visuo-spatial, early versions of their model called for two STM buffers (dubbed the "phonological loop" and the "visuospatial sketchpad", respectively) that could operate independently of each other (Baddeley 1986). The "work" of coordinating the simultaneous engagement of the two buffers was ascribed to a cognitive control mechanism that they dubbed the "Central Executive". Baddeley (1986) has written subsequently that he construed the Central Executive as being akin to Norman and Shallice's (1980) Supervisory Attentional System and, as such, its function

---

[1] At a conference in 2014, Alan Baddeley attributed his adoption of the term "working memory" to a label in a diagrammatic figure of the Atkinson and Shiffrin model (although not appearing in their highly cited (1968) chapter), and explicitly not to Miller, Galanter, and Pribram (1960), which he professed to have not read prior to 1974 (A. Baddeley, personal communication).

wasn't assumed to be restricted to only tasks with an overt memory component[2]. In the context of this "multiple component model" (Baddeley 1986), tasks that only require the STR of information are assumed to only engage one of the buffers, and are thus considered tests of STM (e.g., Logie and Niven 2012). Tasks that require coordination between the buffers, or manipulation of information being held in the buffers -- such as mental navigation in a spatial array, or mental arithmetic -- in contrast, are considered tests of WM. Beyond the context of the multiple component model, many cognitive models that don't relate explicitly to this framework also use a similar convention for categorizing tasks as tests of STM or WM: the former require only the STR of information, the latter require additional cognitive operations that entail the manipulation of the information that is being retained, and/or the flexible use of that information to guide behavior in tasks that are more complicated than simple recognition or recall (e.g., Engle et al. 1999, Daneman and Hannon 2007, Konstantinou and Lavie 2013).

It should be noted that there is a perspective from which this distinction of STM corresponding to the "simple" STR of information vs. WM corresponding to STR plus additional executive control can be called into question, and this will be taken up in the final subsection of this Preamble. First, however, a review of the most compelling basis for maintaining this distinction.

Neurobiological bases for the STM-WM distinction

At the level of neural systems, there's strong evidence that tests of WM require contributions from the prefrontal cortex (PFC) that are not necessary for STM. In some cases this suggests an organization in which WM can be construed as, to oversimplify, "STM + PFC". In others, full double dissociations have been demonstrated. One example of the latter is that in the monkey, the STR of a single object for up to 120 sec (a test of STM) depends on the integrity of the anterior inferotemporal cortex, but not of the mid dorsolateral PFC, whereas the serial selection of 2 different items, separated in time by 10 sec., from a set of 3, 4, or 5 items ("self-ordered choosing", a test of WM) shows the opposite pattern (Petrides 2000). In humans, delayed-recognition of the ordinal position of five randomly ordered letters is not affected by delay-period repetitive transcranial magnetic stimulation (rTMS) of the dorsolateral PFC, but it is impaired by rTMS to parietal cortex. rTMS of dlPFC does impair performance, however, when subjects are required reorder these letters into alphabetical order during the delay period (Postle et al. 2006). Similarly, although large lesions of PFC generally leave verbal and spatial span performance intact, the same patients are impaired on tests of immediate serial recall of items in the reverse of the presention order (i.e., "backward span"; (D'Esposito and Postle 1999)).

As indicated in the previous paragraph, WM differs from STM in that the former entails operations that transform the remembered information, or that require control operations beyond its simple STR. A real-world example of the former is performing mental arithmetic on the remembered total on a restaurant bill, so as to determine the amount of money to leave as a tip. A real-world example of the latter might be keeping track, during a conversation about the grown children of someone who one is meeting for the first time, of which of several children lives in which of several cities. There are three broad classes of WM task. Delay tasks are operationally similar to tests of STM, but require some mental transformation of the remembered information, such as reordering it from order of presentation to some cardinal order (e.g., alphabetical or numerical). Continuous (or "running") tasks entail the serial presentation of multiple items, each of which updates the mental representation of the memory set (as in tests of running span), and, for some tasks, requires a decision that is contingent upon what came earlier in the series of stimuli (e.g., "n-back" and "AX-CPT"[3]). Finally, dual-tasks require switching between

---

[2] Indeed, an early neuroimaging study designed to isolate activity attributable to the Central Executive (and identifying it in the prefrontal cortex) employed a dual-task procedure in which neither of the individual tasks was a memory task (D'Esposito et al., 1995).

[3] This is a "continuous performance task" (CPT) featuring the serial presentation of individual letters, and the instructions that the appearance of the letter *X* requires one type of response if it was

two tasks being performed in parallel, with one task often being retention of a span of items being generated by the other task (e.g., sentence span and operation span).

STM and WM both depend on the STR of information. They differ from long-term memory (LTM) in that they do not depend on the integrity of the medial-temporal lobe diencephalic memory system (hereafter abbreviated MTL) – patients with a pure anterograde amnesia will typically perform normally on tests of STM and on some tests of WM. This means that the STR of information is independent of the processes that encode information into LTM. Importantly, however, there are two things that this does not mean. First, it does not mean that information being held for a test of STM or WM can't also be incidentally processed by the MTL (in neurologically healthy individuals), and thus encoded into LTM (Ranganath and Blumenfeld 2005, Ranganath et al. 2005). Second, it doesn't rule out the possibility that the STR of information may be supported by the temporary activation, via attention, of LTM representations that existed prior to the performance of the STM or WM test (Postle 2007, Lewis-Peacock and Postle 2008).

The role of PFC in the control of interference, on tests of STM and WM

One perspective from which the distinction between STM and WM becomes murky is that of the control of proactive interference (PI) in STM. It has been known at least since the 1960s that every trial after the very first on a test of STM is more prone to error, due to interference from material that was processed on previous trials (Keppel and Underwood 1962). The control of PI in STM has been localized to left inferior PFC via brain imaging (Jonides et al. 1998, D'Esposito et al. 1999), neuropsychology (Thompson-Schill et al. 2002), and rTMS (Feredoes et al. 2006, Feredoes and Postle 2010). Relatedly, findings in the monkey suggest that the delayed-response impairment that follows PFC lesions is better understood as increased susceptibility to interference, and/or poor control of behavior with remembered information, rather than as an increased impairment in the STR, per se, of information (Malmo 1942, Tsujimoto and Postle 2012). Indeed, there is ongoing debate about the extent to which the high correlations that are found between visual STM and general fluid intelligence (reviewed, e.g., in Luck and Vogel 2013) reflect the STR, per se, of information versus such operations as the selection of information into STM (e.g., Linke et al. 2011), the control of the effects of interference in STM (sometimes called "filtering", e.g., Vogel et al. 2005), and/or the retrieval of information in the face of PI (Shipstead and Engle 2013). Independent of this scientific debate, it is undoubtedly the case that in the "real world" outside of the laboratory there are likely very few occasions in which STM isn't supported by some level of PFC-based control. Does this mean that it's not useful to distinguish between STM and WM? The answer to this question is an unequivocal "No". Afterall, "some level of PFC-based control" is involved in virtually all classes of behavior – for example, motor control, language, social behavior, perceptual decision making – yet we nonetheless find it useful to distinguish between these classes of behavior because of important differences between them. Similarly, the arguments laid out earlier in this section provide a conceptual basis, at several levels of analysis, for distinguishing between STM and WM. Indeed, further along, we shall see that failing to do so can lead to erroneous interpretation of neural data.

The Neuroanatomical Bases of the STR of Visual Information

Evidence from lesions and stimulation experiments

There is broad consensus that the STR of information depends on the same networks responsible for the processing of that information in contexts that do not require memory. For language-based information, this implicates regions that are involved in speech production (Leff et al. 2009, Acheson et al. 2011, Koenigs et al. 2011, Richardson et al. 2011); for sensory information, this implicates regions involved in the perception of that sensory modality (e.g., Pasternak and Greenlee 2005). The strongest

---

immediately preceded by the letter *A*, but a different response if it was immediately preceded by any other letter.

evidence for this view comes from the demonstration that perturbations of the critical tissue influences STM in systematic, and selective, ways. For example, the studies of verbal STM cited above have shown that both speech production and verbal STM are sensitive to the integrity of gray matter in left posterior superior temporal gyrus (Leff et al. 2009, Koenigs et al. 2011, Richardson et al. 2011) or to delay-period rTMS of this region (Acheson et al. 2011). Delay-period rTMS of left middle temporal gyrus affects other language-based processes, but neither speech production-specific processes nor verbal STM (Acheson et al. 2011). Within the visual modality, similar evidence has been marshaled for the STR of spatial location with the demonstration that delay-period rTMS of intraparietal sulcus (IPS), superior parietal lobule (SPL) and the frontal eye field (FEF) selectively influences performance on tests of STM for object location but not for object identity. Further, delay-period rTMS of neither dorsolateral prefrontal cortex (PFC) nor postcentral gyrus has this effect (Hamidi et al. 2008). A more nuanced approach, taken for the STR of visual motion, leverages the fact that TMS of visual area MT can produce the percept of a "moving phosphene" – a flash of light that contains coherent motion within the area of the flash. The perceived direction of motion is reproducibly toward the periphery, away from the fovea, in the visual field contralateral to the side of stimulation. Silvanto and Cattaneo (2010) demonstrated that this percept is systematically influenced when TMS is delivered while the subject is engaged in STM for the direction of motion of a target stimulus. When the target motion is in the same direction as the motion of the phosphene, its perception is enhanced. However, when the target motion is in the opposite direction, perception of the delay-period phosphene is reduced. These results indicate that the physiological state of MT varies systematically as a function of the direction of motion being remembered, just as it does, when a stimulus is present, as a function of the direction of motion being perceived.

Evidence from multivariate analyses of fMRI data

Compelling evidence for what has been dubbed the "sensory recruitment" model (e.g., Ester et al. 2013) has also been generated by recent functional magnetic resonance imaging (fMRI) studies of visual STM that have employed multivariate analysis techniques. These represent an important advance over traditional univariate approaches because their superior specificity and sensitivity support tests of whether neural activity in a brain region may support the STR of the stimulus information being retained on a trial-by-trial basis. (Multivariate pattern analysis (MVPA) has been reviewed in many places (e.g., Haynes and Rees 2006, Kriegeskorte et al. 2006, Norman et al. 2006, Pereira et al. 2009); reviews of its applications in STM research can be found in Postle (in press) and Sreenivasan et al.(2014).) Thus, for example, two studies have demonstrated that primary visual cortex (V1) supports the delay period-spanning representation of the color or orientation of target stimuli (Harrison and Tong 2009, Serences et al. 2009). Further, and of relevance for the following section on *The neurophysiological bases of the STR of visual information*, both of these findings were made despite the absence of sustained, elevated levels of signal intensity during the delay period. This pattern of results has been replicated with other classes of stimulus: The STR of motion can be decoded from lateral extrastriate cortex, including area MT+, as well as from medial calcarine and extracalcarine cortex (Riggall and Postle 2012, Emrich, Riggall et al., 2013); the STR of complex visusospatial patterns can be decoded from occipital and parietal cortex (Christophel et al. 2012); and the STR of familiar objects, faces, houses, scenes, and body stimuli can be decoded from the delay-period activity of ventral occipitotemporal cortex (Han et al. 2013, Lee et al. 2013, Nelissen et al. 2013, Sreenivasan et al. 2014).

Many of these studies have also established the specificity of these functions to these regions by failing to decode target identity from other regions. The studies of the STR of visual motion did this, for example, by also applying a standard univariate analysis to identify brain regions that showed elevated activity that was sustained across the delay period. This identified portions of superior and lateral frontal cortex and of intraparietal sulcus (IPS) that are invariably identified with such analyses. When MVPA was applied to these regions, however, it was unable to recover trial-specific stimulus information from these delay-active regions (Riggall and Postle 2012, Emrich, Riggall, et al. 2013) (Figure 1). Note that was it is unlikely the case that the frontal and parietal regions were somehow “less classifiable”, because in the Riggall and Postle (2012) data set, pattern classifiers were able to recover trial-specific task instruction-related information from these regions. That is, MVPA showed the frontal and parietal regions to encode whether the instructions on a particular trial were to remember the speed or the direction of the moving dots that had been presented as the sample stimulus. Analogous patterns of specificity have been reported for STM for other classes of stimuli. With familiar objects, when the task required STM for stimulus features, the STR of stimulus identity could be decoded from ventral occipitotemporal cortex, but not from dorsolateral PFC; when it required STM for stimulus category, however, the reverse was true (Lee et al. 2013). Further, as was the case in the STM-of-motion studies, this pattern was independent of delay-period signal intensity, which did not distinguish among tasks in either region. A study of delayed recognition of faces, scenes, or faces and scenes that decoded at the level of category found evidence for delay-period representation of stimulus category in both dorsolateral PFC and ventral occipitotemporal cortex. However, a clever analysis of the classifier’s “confusion matrix” – that is, the pattern of guesses that it made on trials that were classified erroneously – indicated that that classifier’s performance with signal from ventral occipitotemporal cortex, but not from PFC, was consistent with the STR of a sensory representation (Sreenivasan et al. 2014).

-- Insert Figure 1 about here --

Finally, three studies have linked the precision of the delay-period neural representation of target stimuli in sensory cortex with behavioral estimates of mnemonic precision, showing that “the relative “quality” of … patterns [of activity in sensory cortex] determine the clarity of an individual’s memory” (Ester et al. 2013, p. 754). Ester and colleagues (2013) did so for the STR of orientation by using multivariate encoding models to relate individual differences in delay-period neural representation of target stimuli in visual areas V1 and V2v to individual differences in behavioral performance. More specifically, they showed that the precision of population tuning curves estimated from delay-period signal from these regions predicted the accuracy with which a subject was able to reconstruct the target orientation at the end of the delay period. (See Serences and Saproo (2012) for a nice tutorial review of multivariate encoding models.) Emrich, Riggall, et al. (2013), in work that is detailed in Emrich’s contribution to this volume, did so by varying from trial to trial the number of directions of motion that had to be remembered, and demonstrating a reliable within-subject correlation between the load-related decline in delayed-recall precision and a load-related decline in MVPA decoding performance from calcarine, pericalcarine, and extrastriate regions. Anderson et al. (2014), have demonstrated that the precision of the STR of visual information is reflected in patterns of stimulus-induced activity in the alpha band, measured primarily from posterior electrodes.

Because function follows from structure, the above-summarized evidence for *where* in the brain stimulus information is retained during tests of STM constitutes a necessary and important first step for addressing the broader question of *how* the STR of visual information may be accomplished. The next section reviews the evolution in thinking about the mechanisms that may underlie the STR of visual information.

The Neurophysiological Bases of Working Memory

The thesis of this section is that conflation of the concepts of the STR of information and of WM has led to mistaken assumptions about the neural bases of the former. Thus, disentangling this confused state of affairs, and specifying some of what we know about the neural bases of WM performance, is a necessary step for making progress in understanding the STR of visual information.

Rate Coding Models

There is a seductively powerful, intuitive tendency to ascribe a first-order rate-code interpretation to the information-processing function of a neuron. To experience this first hand, do an internet search for “Hubel and Wiesel simple cell” and watch a movie of one of their early, pioneering experiments. What one sees is the projection screen from the perspective of the anesthetized cat, and, like the cat's fanciful homunculus, one also listens in on the report from one V1 neuron while also observing the full range of movements and shape-changes of the bar of light while it sweeps about the screen, both inside and outside of the neuron's receptive field. The selectivity of this neuron's responses seems clear: It only responds to a luminance contrast-defined edge within a narrow range of orientations, and only at a specific location on the screen. The inference of the function being carried out by this neuron follows automatically: *The neuron is acting as a detector of this narrowly defined feature, and only when the feature appears in this small region of the visual field*. When this process is repeated across several neurons, and each is observed to have different tuning properties, one can then formulate a model about the function of the region from which these neurons were sampled: *The primary visual cortex functions as banks of feature detectors that report the presence of a finite number of elemental features to higher levels of cortex; integration across receptive fields at these higher levels leads to object recognition*. This would be a reasonable, if grossly oversimplified, one-sentence summary of "how primary visual cortex works". Note, however, that the initial intuitive interpretation of the function of a V1 simple cell is necessarily univariate, in that one only has access to the activity of one unit in the brain and, further, to only one dimension along which that unit's activity is varying (i.e., firing rate). Thus, it was not only "lucky" for Hubel and Wiesel that the edge of a glass slide became stuck, at just the right angle, in their projector. It was also lucky for them that the system that they were studying turned out to be amenable to this intuitive analysis. The same would not have been true, for example, if they had been seeking to "crack the code" of olfactory perception (Wilson and Mainen 2006). And, of primary relevance for this chapter, the preponderance of evidence indicates that same is also not true for the functions of neurons of the PFC. Nonetheless, for historical as well as expository reasons, a review of how we have come to know that the STR of visual information is not supported by a rate code in highly specialized PFC neurons will be an instructive way to begin this section.

As reviewed elsewhere (e.g., Postle 2006), in the decades prior to the advent of extracellular recording from neurons in the PFC, the first of which were reported in 1971, the consensus view from lesion experiments was that the dorsolateral PFC is necessary for performance on tasks of WM, as defined in the *Preamble* of this chapter, but not for the STR, per se, of information. For example, the seminal volume from Warren and Akert (1964) on *The Frontal Granular Cortex and Behavior,* which followed from a 1962 symposium featuring the who's-who of the day, emphasizes the role for PFC in learning novel tasks, controlling perseveration on tasks that periodically reverse reward contingencies (including delayed alternation in monkeys (Mishkin 1964) and the Wisconsin Card Sorting Task in humans (Milner 1964)), and guiding behavior with information held in STM. With regard to the delayed-response deficit that is sometimes observed after lesions of the PFC, Pribram and colleagues (1964) summarized studies from the 1930s and 1940s that "found experimentally that aspects other than the trace of memory [in our parlance, the STR of information] were involved by the frontal procedure [i.e., were compromised by PFC lesions]: action at the time of stimulus presentation and distractability were found to be important. These results were interpreted to indicate that the delayed-response task tested for one trial learning and retroactive inhibition rather than for memory-trace decay" (p. 29). Consistent with this review of the (then) two decades-old data from monkeys, in that same volume, Milner also reported on the performance of PFC-lesioned patients on tests of 60-sec delayed recognition. When tested with novel nonsense shapes, with no reuse of stimuli across trials (i.e., an "open" stimulus set), performance of these patients was intact. With other types of stimuli, however, the patients were tested on "closed sets", in which items were reused across trials, and with these their performance was impaired. Thus, the impairment of the patients was interpreted as one of impaired control of PI, rather than of impaired STR of information. These decades of accumulated knowledge were largely neglected, however, with the advent of single-unit electrophysiology.

In 1971, the groups of Joaquin Fuster and Garrett Alexander at UCLA, and of Kisou Kubota and

Hiroaki Niki at Kyoto University, published the results of recordings from neurons in the PFC while monkeys performed delayed response and delayed alternation, respectively. During delayed-response performance, Fuster and Alexander (1971) found that many neurons in the PFC displayed elevated firing rates that spanned the duration of the delay period, which varied in length, unpredictably, within a range of 15 to 65 sec. During delayed-alternation performance, Kubota and Niki (1971) observed two classes of task-related activity: neurons with elevated activity during the delay ("D"); and neurons that became active just prior to, and during, the response period ("E"; presumably because their activity predicted the onset of activity in the electromyogram (EMG) that was recording from muscles in the animal's arm). Interestingly, the response profile of many E cells changed quantitatively when the animals performed a simple alternation task with no delay period interposed between responses – pre-response bursts were of lower intensity (i.e., slower firing rate), and they preceded the motor response by a smaller period of time. Of particular relevance for our present purposes, both sets of authors provided interpretations of their data that was compatible with the preexisting understanding of the neuropsychology of the PFC. Fuster and Alexander (1971):

> *"The temporal pattern of firing frequency observed in prefrontal and thalamic units during cue and delay periods suggest the participation of these units in the acquisition and temporary storage of sensory information which are implicated in delay response performance. Their function, however, does not seem to be the neural coding of information contained in the test cues, at least according to a frequency code, for we have not found any unit showing differential reactions to the two positions of the reward.*
>
> *It is during the transition from cue to delay that apparently the greatest number of prefrontal units discharge at firing levels higher than the intertrial baseline… We believe that the excitatory reactions of neurons in MD and granular frontal cortex during delayed response trials are specifically related to the focusing of attention by the animal on information that is being or has been placed in temporary memory storage for prospective utilization."* (p. 654)

Kubota and Niki (1971):

> *... in the delay task, periprincipal E unit may be causally coupled with the initiation of the voluntary lever pressing. During delay task the E unit activity may be coupled with the initiation and sustaining of […] lever pressing rather than [a] memory retrieval process.*
>
> *... D units are hardly correlated with the memory storage (1) or retrieval of the memory for the lever pressing (2). Frequency of D unit during delay phases is not apparently different between right and left lever pressings. However, this interpretation does not exclude the possibility that the activities of neurons in the prefrontal cortex represent a memory function related to the choice of the correct performance on the basis of immediate past experience, i.e., remembering the spatially directed response on the preceding trial"* (p. 346).

Despite these interpretations from the authors themselves, however, what seems to have captured the attention of many who read these reports was that they suggested a physiological correlate of the first of Hebb's (1949) dual traces, the "reverberatory" mechanism for "a transient 'memory' of [a] stimulus" (p. 61) that would sustain a representation until it could be encoded into a more permanent state via synaptic strengthening

The attribution of explicitly mnemonic interpretations to the findings of Fuster and Alexander (1971) and of Kubota and Niki (1971) became more prevalent with the subsequent explosion interest in the construct of WM that arose from the introduction of the multiple-component model of Baddeley and Hitch (1974). Particularly influential was the view of Goldman-Rakic (1987, 1992), that the sustained delay-period activity in the PFC of the monkey and the storage buffers of the multiple-component model of WM from cognitive psychology were cross-species homologues of the same fundamental mental phenomenon: the STR of information in the service of STM and WM. The basic research protocol employed to test this idea, taken from visual neurophysiology research, proceeded in two stages: First, determine the tuning properties of a neuron; Second, study its delay-period activity during trials when the animal is remembering that neuron's preferred stimulus vs. when it is remembering a non-preferred stimulus. This was exemplified by the studies of oculomotor delayed response by Funahashi and colleagues (1989, 1990), in which neurons in the PFC showed elevated firing rates that were delay-spanning, and retinotopically specific. The model that emerged from these data was of a population of PFC neurons, each tuned for a narrow region of retinotopically defined space, supporting the STR of a location in space via a rate code.

The flaw with this reasoning, however, was that it conflated two constructs from the multiple component model: the storage buffers and the Central Executive. In the parlance of this chapter, it conflated concepts of the STR of information and of WM. It is the Central Executive that would be expected to support performance on tests of WM, as defined at the beginning of this chapter, and that might reasonably be localized, to a first order of approximation, to the PFC. However, even though the delay-spanning activity of PFC neurons was often referred to as "spatial WM activity", conceptually it was being interpreted as a neural correlate of a storage buffer, rather than as a neural correlate of the Central Executive. This, therefore, propagated the idea that the STR of information is accomplished via the rate code-based activity of memory-specialized neurons in the PFC.

Beginning in the late 1990s, many studies have been performed that have shown that sustained delay-period activity in PFC neurons is better understood as supporting one or more aspects of the control of memory and/or behavior (what might be portrayed as the functions of the Central Executive), rather than the STR, per se, of information. For example, PFC neurons have been shown to not be specially tuned for any particular kind of information, but, instead, to modify their response properties to reflect changing environmental exigencies (Rao et al. 1997, Duncan and Miller 2002, Fuster 2002, Duncan 2010). The activity of these neurons isn't restricted to delay periods, or even to tasks that require the STR of information, in that dorsolateral PFC neurons with delay-period activity in the oculomotor delayed-response task also exhibit similar sustained activity during the "delay" period of a visually guided saccade task (i.e., when the target remains visible throughout the trial, and so no memory is needed, Tsujimoto and Sawaguchi 2004). These neurons can dynamically change during a single delay period from retrospectively representing the location of the sample stimulus to prospectively representing the target of the impending saccade, on a task in which the saccade must be made to a location that is a rotated transformation of the sample location (Takeda and Funahashi 2002, Takeda and Funahashi 2004, Takeda and Funahashi 2007). In a task that dissociates the focus of spatial attention from the focus of spatial memory, the majority of DLPFC neurons are seen to track the former (Lebedev et al. 2004). Additional evidence against the view that the PFC acts as a STM buffer – from extracellular electrophysiology, neuroimaging, and perturbation studies using lesions, rTMS, and neuropharmacological interventions -- has been reviewed elsewhere, for example, by Postle (2006) and Tsujimoto and Postle (2012).

Given these more recent findings, how can one then understand the activity of PFC neurons that had seemed to be supporting the STR of visual information with a rate code? One answer comes from the recent emphasis on the multivariate nature of neural signals, to which we now turn.

Dynamic and Distributed Coding in the PFC supports WM Performance

A simple statement of the rationale behind multivariate analyses of neural signals, which are similar in principle to those introduced in the earlier section of this chapter on *Evidence from multivariate analyses of fMRI data*, is that the brain supports the simultaneous activity of many, many millions of processing elements all at the same time, and that the overall influence of this multiplicity of activity is very likely to be different from what one could deduce by studying any single processing element in isolation. This has been demonstrated very convincingly at the level of neuronal activity in the PFC via multivariate reanalyses of extant datasets that retrospectively assemble all of the records of single neurons recorded while the animal performed identical trials, and analyzes them all together as though all had been recorded in parallel during a single "virtual trial". In one such study, Meyers and colleagues (2008) reanalyzed a data set in which monkeys had viewed images of 3D computer-generated animals and judged whether each was more cat-like or dog-like. The original study had compared neuronal activity in PFC vs. IT, finding that IT seemed "more involved in the analysis of currently viewed shapes, whereas the PFC showed stronger category signals, memory effects, and a greater tendency to encode information in terms of its behavioral meaning" (Freedman et al. 2003, p. 5235). The multivariate reanalysis, although generally confirming these broader patterns, uncovered novel, surprising information about the interpretation of sustained, elevated firing rates: The representation of stimulus category information in a delayed-match-to-category task "is coded by a nonstationary pattern of activity that changes over the course of a trial with individual neurons … containing information on much shorter time scales than the population as a whole" (Meyers et al., 2008; p. 1407). That is, this information was not carried for extended periods in time by any individual neuron, regardless of whether or not its activity level was elevated at a sustained level across the delay period. Thus, a first-order, intuitive rate-coding interpretation of the activity of a neuron will often lead to a faulty understanding of how its activity is contributing to the representation of behaviorally relevant information.

A second example of a multivariate reanalysis of a single-unit data set, also led by Meyers (Meyers et al. 2012), illustrates this point very clearly. Whereas the original univariate ("rate code-based") analysis of these data indicated that there was elevated sustained activity in feature-tuned neurons, which could be interpreted within the framework of a rate-coding model (Meyer et al. 2011), the multivariate analyses of the same neurons revealed new information that was incompatible with such a model. First, it revealed that, at the population level, the PFC transitioned during the trial from representing the target stimulus to representing the trials's status as a "Match" or "Nonmatch" trial. Second, the first-order activity profile of any particular neuron could not be interpreted at face value, because "task-relevant information in several neurons was present for only short periods of time relative to the duration of the … delay period" (p. 4652). This led to the conclusion that "task-relevant information is incorporated into PFC by interleaving/overlapping new information into ongoing dynamic activity that is carrying information about other variables and consequently the absolute firing rate level of a single neuron at a particular time point is often highly ambiguous if the context of the larger population is not taken into account" (Meyers et al. 2012, p. 4652). Other aspects of these studies are considered in Constantinidis's contribution to this volume.

With this understanding of how rate-coding interpretations of the delay-period activity of PFC neurons are being reinterpreted in the context of dynamic, distributed coding models, we can now turn our attention to mechanisms that may underlie the STR of visual information.

The Neurophysiological Bases of the STR of Visual Information

Distributed Patterns of Activity in Sensory Cortex

The literature reviewed in the previous section helps to reconcile the MVPA-based fMRI findings about *The Neuroanatomical Bases of the STR of Visual Information* with the results from earlier studies of single-unit activity in posterior regions. For example, recordings from MT in monkeys performing delayed-recognition of the direction of motion have failed to find delay-spanning elevated activity in directionally tuned neurons (i.e., no evidence for a sustained rate code, Zaksas and Pasternak 2006), yet MVPA analyses of fMRI data are able to detect a delay-spanning representation of the target stimulus from MT+ (Riggall and Postle 2012, Emrich, Riggall et al., 2013). It may well be that, as with the representation of stimulus information in PFC (e.g., Meyers et al. 2012, Stokes et al. 2013), inferior temporal cortex (Meyers et al. 2008), and posterior parietal cortex (Crowe et al. 2010), the STR of motion in MT+ is supported by a dynamic, distributed code that is not evident from the inspection of the activity of single units. A related possibility is considered in some detail in the chapter in this volume by Martinez-Trujillo, which includes evidence that the STR of motion information may be supported by distributed patterns in local field potentials (LFPs) in MT. That is, the delay-period neural code in MT may be subthreshold, in the sense that it is maintained in patterns of oscillating membrane potentials, but not in spiking activity that could be read out by other brain regions. Consistent with this oscillatory account of the STR of information are the EEG findings of Anderson and colleagues (2014), that the STR of visual orientation information, including its precision, is reflected in patterns of stimulus-induced activity in the alpha band. Because the blood oxygenation level-dependent (BOLD) signal to which fMRI is sensitive corresponds more closely to dynamics of the LFP than to firing rates (e.g., Logothetis et al. 2001), it is plausible that successful delay-period decoding of target motion from MT+ with fMRI (Riggall and Postle 2012, Emrich, Riggall et al., 2013) may be due to distributed patterns in the LFP in this region.

For the remainder of this chapter, however, we will consider evidence that the MVPA decoding of delay-period activity may not reflect the STR of information, per se, but rather the allocation of attention to this information.

Does the STR of sensory information fundamentally depend on a weight-based scheme?

It is unequivocally the case that MVPA techniques feature superior sensitivity and specificity than univariate analysis methods (e.g., Lewis-Peacock and Postle 2012). These facts embolden the articulation of the question posed in the title of this subsection, which derives from the last set of data to be reviewed in this chapter.

The first of three studies to be reviewed here, Lewis-Peacock et al. (2012), was an fMRI study of a multistep delayed-recognition task (adapted from Oberauer 2005) presenting two sample stimuli, then retrocues informing the subject which sample was relevant for each of the two successively presented memory probes. More specifically, each trial began with the presentation of two sample stimuli, always selected from two of three categories (oriented lines, words, and pronounceable pseudowords), one in the top half of the screen and one in the bottom half. After offset of the stimulus display and an initial delay period, a retrocue indicated which sample was relevant for the first recognition probe, followed by a second delay, followed by an initial Y/N recognition probe (and response). Critically, up until this point in the trial, both items needed to be kept in STM, because the first probe was followed by a second retrocue that, with equal probability, would indicate that the same item (a "repeat" trial) or the previously uncued item (a "switch" trial) would be tested by the trial-ending second recognition probe. Thus, the first delay was assumed to require the active retention of two items, whereas the second delay would feature the retention of an "attended memory item" (AMI) and an "unattended memory item" (UMI). (Behavioral data indicate that subjects remove items from the focus of attention following the first retrocue, even though they know that $p = .5$ that the uncued memory item will be cued by the second retrocue (Oberauer 2005, LaRocque et al. 2013).) The third delay would only require the retention of an AMI, because it was certain that memory for the item not cued by the second retrocue would never be tested. This design therefore allowed us to unconfound the STR of information from the effects of attending to information in STM. (Note that the vast majority of tasks assessing STM and WM do confound these two factors. Exceptions include the dual-tasks paradigms summarized in this chapter's *Preamble*, the study of Lebedev et al. (2004) that was cited in the section on *The Neurophysiological Bases of the STR of Visual Information*, and the studies reviewed in this subsection.)

Prior to performing the multistep delayed-recognition task, subjects were first scanned while performing a simple one-delay delayed-recognition task, and the data from this *Phase 1* scan were used to train the classifier that was then applied to the data from the multistep task. For Phase 1, subjects were trained to indicate whether the probe stimulus matched the sample according to a category-specific criterion -- synonym judgment for words, rhyme judgment for pseudowords, and an orientation judgment for line segments. Our rationale was that by training the classifiers (separately for each subject) on data from the delay period of this task, we'd be training it on patterns of brain activity related to the STR of just a single representational code: phonological (pseudoword trials), semantic (word trials), or visual (line trials). This, in turn, would provide the most unambiguous decoding of delay periods entailing the STR of two AMIs (Delay 1) vs. of one AMI and one UMI (Delay 2) vs. of one AMI (Delay 3).

As illustrated in Figure 2, for all trial types, classifier evidence for both trial-relevant categories rose steeply at trial onset and remained at the same elevated level until the onset of the first retrocue. This indicated that both items were encoded and sustained in the focus of attention across the initial memory delay, while it was equiprobable that either could be relevant for the first memory response. Following onset of the first retrocue, however, classifier evidence for the two memory items diverged. Postcue brain activity patterns were classified as highly consistent with the category of the cued item, whereas evidence for the uncued item dropped precipitously, becoming indistinguishable from the classifier's evidence for the stimulus category not presented on that trial (i.e., not different from baseline). If the second cue was a repeat cue, classifier evidence for the already-selected memory item remained elevated and that of the uncued item remained indistinguishable from baseline (Figure 2, Repeat). If, in contrast, the second cue was a switch cue, classifier evidence for the previously uncued item was reinstated, while evidence for the previously cued item dropped to baseline (Figure 2, Switch).

-- Insert Figure 2 about here --

These results raise the possibility that only AMIs are held in a neurally active state. Importantly, despite the apparent loss of sustained activity, UMIs were nonetheless easily remembered on "cue-switch" trials, meaning that they remained "in" STM throughout the trial. Thus, it may be that the STR of information does not depend on the active neural representation of that item. Possible alternative mechanisms, such as the STR of information in a synaptic weight-based format, will be considered toward the end of this section. First, however, we have to consider possible alternative reasons for what amounts to a null finding of multivariate evidence for an active neural representation of UMIs.

One important caveat about the Lewis-Peacock et al. (2012) findings is that, because they were derived from fMRI data, the failure to detect the active representation of UMIs may have been because the STR of the UMI is accomplished via a mechanism to which the BOLD signal is not sensitive. One candidate for such a mechanism is neuronal oscillations, as considered in the earlier subsection on *Distributed Patterns of Activity in Sensory Cortex*. Although the point there was that an oscillatory code may be used as an alternative to a spiking code, and that the BOLD signal is sensitive to dynamics in the LFP, it is nonetheless the case that the BOLD signal is most reliably sensitive to LFP dynamics within certain ranges of frequencies, the gamma band (≥ 30 Hz) being the most important. Thus, if the STR of information were supported by neuronal oscillations at a lower frequency (e.g., Jensen et al. 2002, Sauseng et al. 2009, Uhlhaas et al. 2009, Fuentemilla et al. 2010, Palva and Palva 2011), fMRI might not be expected to pick this up reliably. Therefore, LaRocque and colleagues (2013) designed a follow-up study to replicate the critical features of Lewis-Peacock et al. (2012), with the exception that neural activity was concurrently measured with the electroencephalogram (EEG), rather than with fMRI. EEG is sensitive to neuronal oscillations across a broad, physiologically relevant range of frequencies, and its efficacy for studying STM and WM is highlighted, for example, in Anderson et al. (2014).

The results from the EEG study of LaRocque et al. (2013) replicated the principal finding from fMRI (Lewis-Peacock et al. 2012): MVPA of the EEG signal failed to find evidence that information that was outside the focus of attention, but nonetheless in STM (i.e., UMIs), was retained in an active state. An additional analysis also ruled out the possibility that a neural representation is represented differently when being retained as a UMI vs. when being retained as an AMI. (If this were the case, MVPA of data trained on AMIs from Phase 1 might be expected to fail to detect UMIs during Phase 2).

This was achieved by implementing MVPA by training and testing on data from Delay 2 (i.e., following the first retrocue, when there was one AMI and one UMI) via the leave-one-out cross-validation procedure. These results, illustrated in Figure 3, qualitatively replicated those from the train-on-Phase-1-test-on-Phase-2 analysis.

-- Insert Figure 3 about here --

Although the findings from the Lewis-Peacock et al. (2012) and the LaRocque et al. (2013) studies both failed to find evidence that UMIs are maintained in an active state, there is a caveat that applies to both of them, due to the fact that they employed decoding at the level of category (i.e., semantics vs. phonology vs. visual features), rather than at the level of remembered item. Because of this, they cannot rule our the possibility that the sustained activation of the neural representation of an individual item may not be detectable by a classifier trained to discriminate among categories of stimuli. The most likely reason for this would boil down to sensitivity: Absent the boost of selective attention, the activity presumed to underlie the STR of an individual item may simply not be detectable via MVPA of fMRI and EEG data. Although we can never rule out this possibility by conducting additional studies with fMRI or EEG, Joshua LaRocque and colleagues (2013) are in the process of carrying out the strongest possible test of this idea with fMRI, by replicating the procedure from Lewis-Peacock et al. (2012), but doing so at the level of individual items all drawn from the same stimulus category. Instead of drawing, for each trial, on two stimuli from each of three possible categories, each trial in the LaRocque et al. (2013) study is drawing from two of three possible directions of motion. The results from this study, with an *n* of 8 subjects, look qualitatively identical to the results from Lewis-Peacock et al. (2012) that are illustrated in Figure 1. That is, MVPA evidence for an active neural representation for an item drops to baseline as soon as a retrocue indicates that that item will not be relevant for the impending memory probe. Thus, across 2+ studies, we have consistently failed to find evidence for the active neural representation of UMIs in STM. Although this record can be construed as a string of null results, each was produced with a method boasting superior sensitivity and specificity than any study of STM that preceded it. Further, multivariate analyses of extracellular recordings from the PFC have also failed to find evidence for sustained activity (c.f. the contribution by Martinez-Trujillo to this volume). This confluence of evidence impels us to consider mechanisms other than sustained activity that may support the STR of visual information.

There is near-universal acceptance among neuroscientists that LTM for past episodes is encoded in the brain via distributed patterns of synaptic weights that, when activated, produce the retrieval of information about that episode. Less often considered, but just as computationally plausible, is that information that must be held for seconds, in STM, might also be encoded in a distributed pattern of synaptic weights. Such a mechanism has also been inferred from multivariate studies of extracellular activity in PFC (Barak et al. 2010, Barak et al. 2013, Stokes et al. 2013). For example, Stokes and colleagues (2013) offered the following account of activity recorded during performance of a delayed paired-association task, in which the target could be followed by one or more neutral or foil stimuli before it's associate appeared:

> *"If patterned activity leaves behind a patterned change in the synaptic weights of the network (i.e., hidden state), then subsequent stimulation will be patterned according to the recent stimulation history of a network... Thus, any driving input to the system will trigger a systematic population response that could be used to decode the recent stimulation history of the network. Exactly this phenomenon is seen in our data during the presentation of the neutral stimulus ... Although this stimulus was fixed across trials, the population response was patterned according to the identity of the previous cue, providing a more reliable readout of the memory content than the population response observed during the relatively quiescent delay period."* (p. 372)

Relatedly, Sugase-Miyamoto et al. (2008) have proposed that regions of inferior temporal cortex support visual STM for objects by acting as a “matched filter” that holds a “static copy” of the target in a pattern of synaptic weights. The physiological mechanism(s) that would implement such a transient pattern of weights would be different from those that implement MTL-mediated LTM consolidation. Plausible mechanisms include GluR1-dependent short-term potentiation, which effectively lasts in a tissue slice for a few 10s of sec (Erickson, Maramara, & Lisman, 2009), and transient presynaptic increases in calcium ion concentration, which is the basis for a computational simulation of STM without sustained activity (Mongillo, Barak, & Tsodyks, 2008).

Conclusion

In summary, a review of decades-old lesion data, combined with the results of recent neurostimulation studies and multivariate analyses of fMRI and electrophysiological studies in human and nonhuman primates, support the following propositions about of the STR of visual information:

- It depends on the same neural circuits that are necessary for the real-time perception of visual features, objects, and locations (i.e., “sensory recruitment”)
- It may not depend on the sustained activity of neural representations; instead, it may depend on transiently configured patterns of synaptic weights.
- Rather than reflecting the STR of information, per se, the active delay-period representation of target stimuli may be attributable to selective attention, a mechanism that is often experimentally confounded with the STR of information.
- Therefore, although there may, indeed, be multiple states in which information can exist in STM (e.g., Cowan 1988, Oberauer 2002, Olivers et al. 2011), psychological states of “activation”, such as “activated LTM”, may not correspond to a neurophysiological state of activation.

Future research

There are many important topics for future research. One is how systems that act as the source of attentional control select and interact with neural representations of sensory information (e.g., Nelissen et al. 2013). A second is, assuming that some variant of the structural-trace model of the STR of visual information can be confirmed, an understanding the circumstances under which such a trace will vs. will not be “overwritten” by trial-irrelevant information that is presented during the delay period (Sugase-Miyamoto et al. 2008, Stokes et al. 2013). A third is understanding the function of the (metabolically expensive) increases in STM and WM task-related activity that is commonly observed in PFC and IPS, even when MVPA suggests that these regions aren’t directly engaged in stimulus representation (e.g., Woolgar et al. 2011).

References

Acheson, D. J., M. Hamidi, J. Binder and B. R. Postle (2011). "A common neural substrate for language production and verbal working memory." Journal of Cognitive Neuroscience **23**: 1358-1367.

Anderson, D. E., J. T. Serences, E. K. Vogel and E. Awh (2014). "Induced alpha rhythms track the content and quality of visual working memory representations with high temporal precision." The Journal of Neuroscience **34**: 7587-7599.

Atkinson, R. C. and R. M. Shiffrin, Eds. (1968). Human memory: A proposed system and its control processes. The psychology of learning and motivation. New York, Academic Press.

Baddeley, A. D. (1986). Working Memory. London, Oxford University Press.

Baddeley, A. D. and G. J. Hitch (1974). Working Memory. The Psychology of Learning and Motivation. G. H. Bower. New York, Academic Press. **8:** 47-89.

Barak, O., D. Sussillo, R. Romo, M. Tsodyks and L. F. Abbott (2013). "From fixed points to chaos: three models of delayed discrimination." Prog Neurobiol **103**: 214-222.

Barak, O., M. Tsodyks and R. Romo (2010). "Neuronal population coding of parametric working memory." J Neurosci **30**(28): 9424-9430.

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

Cowan, N. (1988). "Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system." Psychological Bulletin **104**: 163-171.

Crowe, D. A., B. B. Averbeck and M. V. Chafee (2010). "Rapid sequences of population activity patterns dynamically encode task-critical spatial information in parietal cortex." The Journal of Neuroscience **30**: 11640-11653.

D'Esposito, M., J. A. Detre, D. C. Alsop, R. K. Shin, S. Atlas and M. Grossman (1995). "The neural basis of the central executive system of working memory." Nature **378**: 279-281.

D'Esposito, M. and B. R. Postle (1999). "The dependence of span and delayed-response performance on prefrontal cortex " Neuropsychologia **37**: 1303-1315.

D'Esposito, M., B. R. Postle, J. Jonides and E. E. Smith (1999). "The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI." Proceedings of the National Academy of Sciences, USA **96**: 7514-7519.

Daneman, M. and B. Hannon (2007). What do working memory span tasks like reading span really measure? . The cognitive neuroscience of working memory N. Osaka, R. H. Logie and M. D'Esposito. Oxford, Oxford University Press
: 21-42.

Duncan, J. (2010). "The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour." Trends Cogn Sci **14**(4): 172-179.

Duncan, J. and E. K. Miller (2002). Cognitive focus through adaptive neural coding in the primate prefrontal cortex. Principles of Frontal Lobe Function. D. Stuss and R. Knight. Oxford, Oxford University Press**:** 278-291.

Emrich, S. M., A. C. Riggall, J. J. Larocque and B. R. Postle (2013). "Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory." J Neurosci **33**(15): 6516-6523.

Engle, R. W., S. W. Tuholski, J. E. Laughlin and A. R. A. Conway (1999). "Working memory, short-term memory, and general fluid intelligence: a latent-variable approach." Journal of Experimental Psychology: General **128**: 309-331.

Ester, E. F., D. E. Anderson, J. T. Serences and E. Awh (2013). "A neural measure of precision in visual working memory." J Cogn Neurosci **25**(5): 754-761.

Feredoes, E. and B. R. Postle (2010). "Prefrontal control of familiarity and recollection in working memory." Journal of Cognitive Neuroscience **22**: 323-330.

Feredoes, E., G. Tononi and B. R. Postle (2006). "Direct evidence for a prefrontal contribution to the control of proactive interference in verbal working memory." Proceedings of the National Academy of Science (USA) **103**: 19530-19534.

Freedman, D. J., M. Riesenhuber, T. Poggio and E. K. Miller (2003). "A comparison of primate prefrontal and inferior temporal cortices during visual categorization." J Neurosci **23**(12): 5235-5246.

Fuentemilla, L., W. D. Penny, N. Cashdollar, N. Bunzeck and E. Düzel (2010). "Theta-coupled periodic replay in working memory." Current Biology **20**: 606-612.

Funahashi, S., C. J. Bruce and P. S. Goldman-Rakic (1989). "Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex." Journal of Neurophysiology **61**: 331-349.

Funahashi, S., C. J. Bruce and P. S. Goldman-Rakic (1990). "Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms." Journal of Neurophysiology **63**: 814-831.

Fuster, J. M. (2002). Physiology of executive functions: the perception-action cycle. Principles of Frontal Lobe Function. D. T. Stuss and R. T. Knight. Oxford, Oxford University Press**:** 96-108.

Fuster, J. M. and G. E. Alexander (1971). "Neuron activity related to short-term memory." Science **173**: 652-654.

Goldman-Rakic, P. S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational memory. Handbook of Neurobiology. V. B. Mountcastle, F. Plum and S. R. Geiger. Bethesda, American Physiological Society**:** 373-417.

Goldman-Rakic, P. S. (1992). "Working memory and the mind." Scientific American **267**: 110-117.

Hamidi, M., G. Tononi and B. R. Postle (2008). "Evaluating frontal and parietal contributions to spatial working memory with repetitive transcranial magnetic stimulation " Brain Research **1230**: 202-210.

Han, X., A. C. Berg, H. Oh, D. Samaras and H. C. Leung (2013). "Multi-voxel pattern analysis of selective representation of visual working memory in ventral temporal and occipital regions." Neuroimage **73**: 8-15.

Harrison, S. A. and F. Tong (2009). "Decoding reveals the contents of visual working memory in early visual areas." Nature **458**: 632-635.

Haynes, J.-D. and G. Rees (2006). "Decoding mental states from brain activity in humans." Nature Reviews Neuroscience **7**: 523-534.

Hebb, D. O. (1949). The Organization of Behavior: A Neuropsychological Theory. New York, NY, John Wiley & Sons, Inc.

Jensen, O., J. Gelfand, J. Kounios and J. E. Lisman (2002). "Oscillations in the Alpha Band (9-12 Hz) increase with memory load during retention in a short-term memory task." Cerebral Cortex **12**: 877-882.

Jonides, J., E. E. Smith, C. Marshuetz, R. A. Koeppe and P. A. Reuter-Lorenz (1998). "Inhibition of verbal working memory revealed by brain activation." Proceedings of the National Academy of Sciences **95**: 8410-8413.

Keppel, G. and B. J. Underwood (1962). "Proactive inhibition in short-term retention of single items." Journal of Verbal Learning and Verbal Behavior **1**: 153-161.

Koenigs, M., D. J. Acheson, A. K. Barbey, J. Solomon, B. R. Postle and J. Grafman (2011). "Areas of left perisylvian cortex mediate auditory-verbal short-term memory." Neuropsychologia **49**(13): 3612-3619.

Konstantinou, N. and N. Lavie (2013). "Dissociable roles of different types of working memory load in visual detection." Journal of Experimental Psychology: Human Perception and Performance **39**: 919 – 924.

Kriegeskorte, N., R. Goebel and P. A. Bandettini (2006). "Information-based functional brain mapping." Proceedings of the National Academy of Science (USA) **103**: 3863-3868.
Kubota, K. and H. Niki (1971). "Prefrontal cortical unit activity and delayed alternation performance in monkeys." J Neurophysiol **34**(3): 337-347.
LaRocque, J. J., J. A. Lewis-Peacock, A. Drysdale, K. Oberauer and B. R. Postle (2013). "Decoding attended information in short-term memory: An EEG study." Journal of Cognitive Neuroscience **25**: 127-142.
LaRocque, J. J., A. C. Riggall, S. M. Emrich and B. R. Postle (2013). "Active representations of individual items in short-term memory: A matter of attention, not retention." Annual Meeting of the Society for Neuroscience: 507.506.
Lebedev, M. A., A. Messinger, J. D. Kralik and S. P. Wise (2004). "Representation of attended versus remembered locations in prefrontal cortex." PLoS Biology **2**(11): e365.
Lebedev, M. A., A. Messinger, J. D. Kralik and S. P. Wise (2004). "Representation of attended versus remembered locations in prefrontal cortex." PloS Biology **2**: 1919-1935.
Lee, S. H., D. J. Kravitz and C. I. Baker (2013). "Goal-dependent dissociation of visual and prefrontal cortices during working memory." Nat Neurosci **16**(8): 997-999.
Leff, A. P., T. M. Schofield, J. T. Crinion, M. L. Seghier, A. Grogan, D. W. Green and C. J. Price (2009). "The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: evidence from 210 patients with stroke." Brain **132**(Pt 12): 3401-3410.
Lewis-Peacock, J. A., A. Drysdale, K. Oberauer and B. R. Postle (2012). "Neural evidence for a distinction between short-term memory and the focus of attention." Journal of Cognitive Neuroscience **23**: 61-79.
Lewis-Peacock, J. A. and B. R. Postle (2008). "Temporary activation of long-term memory supports working memory." The Journal of Neuroscience **28**: 8765-8771.
Lewis-Peacock, J. A. and B. R. Postle (2012). "Decoding the internal focus of attention." Neuropsychologia **50**: 470-478.
Linke, A. C., A. Vicente-Grabovetsky, D. J. Mitchell and R. Cusack (2011). "Encoding strategy accounts for individual differences in change detection measures of VSTM." Neuropsychologia **49**: 1476-1486.
Logie, R. H. and E. Niven (2012). Working memory: An ensemble of functions in on-line cognition. From Mental Imagery to Spatial Cognition and Language: Essays in honour of Michel Denis. . V. Gyselinck and F. Pazzaglia. Hove, Psychology Press**:** 77-105.
Logothetis, N. K., J. Pauls, M. Augath, T. Trinath and A. Oeltermann (2001). "Neurophysiological investigation of the basis of the fMRI signal." Nature **412**: 150-157.
Luck, S. J. and E. K. Vogel (2013). "Visual working memory capacity: from psychophysics and neurobiology to individual differences. ." Trends in Cognitive Sciences **17**: 391-400.
Malmo, R. B. (1942). "Interference factors in delayed response in monkey after removal of the frontal lobes." Journal of Neurophysiology **5**: 295-308.
Meyer, T., X. L. Qi, T. R. Stanford and C. Constantinidis (2011). "Stimulus selectivity in dorsal and ventral prefrontal cortex after training in working memory tasks." J Neurosci **31**(17): 6266-6276.
Meyers, E. M., D. J. Freedman, G. Kreiman, E. K. Miller and T. Poggio (2008). "Dynamic population coding of category information in inferior temporal and prefrontal cortex." J Neurophysiol **100**(3): 1407-1419.
Meyers, E. M., D. J. Freedman, G. Kreiman, E. K. Miller and T. Poggio (2008). "Dynamic population coding of category information in inferior temporal and prefrontal cortex." Journal of Neurophysiology **100**: 1407-1419.
Meyers, E. M., X. L. Qi and C. Constantinidis (2012). "Incorporation of new information into prefrontal cortical activity after learning working memory tasks." Proc Natl Acad Sci U S A **109**(12): 4651-4656.

Milner, B. (1964). Some effects of frontal lobectomy in man. The Frontal Granular Cortex and Behavior. J. M. Warren and K. Akert. New York, McGraw-Hill**:** 313-334.

Mishkin, M. (1964). Perseveration of central sets after frontal lesionss in monkeys. The Frontal Granular Cortex and Behavior. J. M. Warren and K. Akert. New York, McGraw-Hill**:** 219-237.

Nelissen, N., M. Stokes, A. C. Nobre and M. F. Rushworth (2013). "Frontal and Parietal Cortical Interactions with Distributed Visual Representations during Selective Attention and Action Selection." J Neurosci **33**(42): 16443-16458.

Norman, D. A. and T. Shallice (1980). Attention to action: Willed and automatic control of behavior, University of California San Diego.

Norman, K. A., S. M. Polyn, G. J. Detre and J. V. Haxby (2006). "Beyond mind-reading: multi-voxel pattern analysis of fMRI data." Trends in Cognitive Sciences **10**: 424-430.

Oberauer, K. (2002). "Access to information in working memory: exploring the focus of attention." Journal of Experimental Psychology: Learning, Memory, and Cognition **28**: 411-421.

Oberauer, K. (2005). "Control of the contents of working memory—A comparison of two paradigms and two age groups." Journal of Experimental Psychology: Learning, Memory, & Cognition **31**: 714-728.

Olivers, C. N. L., J. Peters, R. Houtkamp and P. R. Roelfsema (2011). "Different states in visual working memory: when it guides attention and when it does not." Trends in Cognitive Sciences **15**: 327-334.

Palva, S. and J. M. Palva (2011). "Functional roles of alpha-band phase synchronization in local and large-scale cortical networks." Frontiers in Psychology **2**: doi: 10.3389/fpsyg.2011.00204.

Pasternak, T. and M. W. Greenlee (2005). "Working memory in primate sensory systems." Nature Reviews Neuroscience **6**: 97-107.

Pereira, F., T. Mitchell and M. M. Botvinick (2009). "Machine learning classifiers and fMRI: a tutorial overview." NeuroImage **45**: S199-S209.

Petrides, M. (2000). "Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory." Journal of Neuroscience **20**: 7496-7503.

Postle, B. R. (2006). "Working memory as an emergent property of the mind and brain." Neuroscience **139**: 23-38.

Postle, B. R. (2007). "Activated long-term memory"? The bases of representation in working memory. The Cognitive Neuroscience of Working Memory. N. Osaka, R. H. Logie and M. D'Esposito. Oxford, U.K., Oxford University Press**:** 333-350.

Postle, B. R. (in press). Activation and information in working memory research. The Wiley-Blackwell Handbook on the Cognitive Neuroscience of Memory. A. Duarte, M. Barense and D. R. Addis. Oxford, U.K., Wiley-Blackwell.

Postle, B. R., F. Ferrarelli, M. Hamidi, E. Feredoes, M. Massimini, M. Peterson, A. Alexander and G. Tononi (2006). "Repetitive transcranial magnetic stimulation dissociates working memory manipulation from retention functions in prefrontal, but not posterior parietal, cortex." Journal of Cognitive Neuroscience **18**: 1712-1722.

Pribram, K. H., A. Ahumada, J. Hartog and L. Roos (1964). A progress report on the neurological processes disturbed by frontal lesions in primates. The Frontal Granular Cortex and Behavior. J. M. Warren and K. Akert. New York, McGraw-Hill Book Company**:** 28-55.

Ranganath, C. and R. S. Blumenfeld (2005). "Doubts about double dissociation between short-and long-term memory." Trends in Cognitive Sciences **9**: 374-380.

Ranganath, C., M. X. Cohen and C. J. Brozinsky (2005). "Working memory maintenance contributes to long-term memory formation: neural and behavioral evidence." Journal of Cognitive Neuroscience **17**: 994-1010.

Rao, S. C., G. Rainer and E. K. Miller (1997). "Integration of what and where in the primate prefrontal cortex." Science **276**: 821-824.

Richardson, F. M., S. Ramsden, C. Ellis, S. Burnett, O. Megnin, C. Catmur, T. M. Schofield, A. P. Leff and C. J. Price (2011). "Auditory short-term memory capacity correlates with gray matter density in the left posterior STS in cognitively normal and dyslexic adults." J Cogn Neurosci **23**(12): 3746-3756.

Riggall, A. C. and B. R. Postle (2012). "The relationship between working memory storage and elevated activity as measured with funtional magnetic resonance imaging." The Journal of Neuroscience **32**: 12990-12998.

Sauseng, P., W. Klimesch, K. F. Heise, W. P. Gruber, E. Holz, A. A. Karim, M. Glennon, C. Gerloff, N. Birbaumer and F. Hummel (2009). "Brain oscillatory substrates of visual short-term memory capacity." Current Biology **19**: 1846-1852.

Serences, J. T., E. F. Ester, E. K. Vogel and E. Awh (2009). "Stimulus-specific delay activity in human primary visual cortex." Psychological Science **20**: 207-214.

Serences, J. T. and S. Saproo (2012). "Computational advances towards linking BOLD and behavior." Neuropsychologia **50**(4): 435-446.

Shipstead, Z. and R. W. Engle (2013). "Interference within the focus of attention: working memory tasks reflect more than temporary maintenance." Journal of Experimental Psychology: Learning, Memory, & Cognition **39**: 277-289.

Silvanto, J. and Z. Cattaneo (2010). "Transcranial magnetic stimulation reveals the content of visual short-term memory in the visual cortex." Neuroimage **50**(4): 1683-1689.

Sreenivasan, K., J. Vytlacil and M. D'Esposito (2014). "Distributed and dynamic storage of working memory stimulus information in extrastriate cortex." Journal of Cognitive Neuroscience **26**: 1141-1153.

Sreenivasan, K. K., C. E. Curtis and M. D'Esposito (2014). "Revisiting the role of persistent neural activity in working memory." Trends in Cognitive Sciences **18**: 82-89.

Stokes, M. G., M. Kusunoki, N. Sigala, H. Nili, D. Gaffan and J. Duncan (2013). "Dynamic coding for cognitive control in prefrontal cortex." Neuron **78**(2): 364-375.

Sugase-Miyamoto, Y., Z. Liu, M. C. Wiener, L. M. Optican and B. J. Richmond (2008). "Short-term memory trace in rapidly adapting synapses of inferior temporal cortex." PLoS Comput Biol **4**(5): e1000073.

Takeda, K. and S. Funahashi (2002). "Prefrontal task-related activity representing visual cue location or saccade direction in spatial working memory tasks." Journal of Neurophysiology **87**(1): 567-588.

Takeda, K. and S. Funahashi (2004). "Population vector analysis of primate prefrontal activity during spatial working memory." Cerebral Cortex **14**(12): 1328-1339.

Takeda, K. and S. Funahashi (2007). "Relationship between prefrontal task-related activity and information flow during spatial working memory performance." Cortex **43**(1): 38-52.

Thompson-Schill, S., J. Jonides, C. Marshuetz, E. E. Smith, M. D'Esposito, I. P. Kan, R. T. Knight and D. Swick (2002). "Effects of frontal lobe damage on interference effects in working memory." Cognitive, Affective, and Behavioral Neuroscience **2**: 109-120.

Tsujimoto, S. and B. R. Postle (2012). "The prefrontal cortex and delay tasks: a reconsideration of the "mnemonic scotoma"." Journal of Cognitive Neuroscience **24**: 627-635.

Tsujimoto, S. and T. Sawaguchi (2004). "Properties of delay-period neuronal activity in the primate prefrontal cortex during memory- and sensory-guided saccade tasks." European Journal of Neuroscience **19**(2): 447-457.

Uhlhaas, P. J., G. Pipa, B. Lima, L. Melloni, S. Neuenschwander, D. Nikolic and W. Singer (2009). "Neural synchrony in cortical networks: history, concept and current status." Frontiers in Integrative Neuroscience **3**: doi: 10.3389/neuro.3307.3017.2009.

Vogel, E. K., A. W. McCollough and M. G. Machizawa (2005). "Neural measures reveal individual differences in controlling access to working memory." Nature **438**: 368-387.

Warren, J. M. and K. Akert, Eds. (1964). The Frontal Granular Cortex and Behavior. New York, McGraw-Hill Book Company.

Wilson, R. I. and Z. F. Mainen (2006). "Early events in olfactory processing." Annu Rev Neurosci **29**: 163-201.

Woolgar, A., A. Hampshire, R. Thompson and J. Duncan (2011). "Adaptive coding of task-relevant information in human frontoparietal cortex." J Neurosci **31**(41): 14592-14599.

Zaksas, D. and T. Pasternak (2006). "Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task." The Journal of Neuroscience **26**: 11726-11742.

Figure Legends

Figure 1.a. Behavioral task from Riggall and Postle (2012). Subjects maintained the direction and speed of a sample motion stimulus over a 15 sec-long delay period. Midway through the delay period, they were cued as to the dimension on which they would be making an upcoming comparison against the remembered sample, either direction or speed. At the end of the delay period, they were presented with a probe motion stimulus and had to indicate with a button press whether it did or did not match the sample stimulus on the cued dimension.

b. BOLD and MVPA time courses from four ROIs, Sample presentation occurred at 0 sec, and at 8 sec subjects were cued that either the direction or speed of sample motion would be tested on that trial. (*A-D*) Average ROI BOLD activity. Data from direction-cued trials use solid lines and speed-cued trials use dashed-lines, bands cover average standard error across subjects. (*E-H*) ROI stimulus-direction decoding results and (*I-L*) ROI trial-dimension decoding results. Each waveform represents the mean direction-decoding accuracy across subjects ($n = 7$) for a classifier trained with data limited to a single time point in the trial and then tested on all time points in the hold out trials (e.g., the green line illustrates the decoding time course from a classifier trained on only data from time point 4, indicated by the small green triangle along the x-axis.) Horizontal bars along the top indicate points at which the decoding accuracy for the corresponding classifier was significantly above chance ($p < 0.05$, permutation test). Schematic icons of trial events are shown at the appropriate times along the x-axis. Data are unshifted in time.

Figure 2.Images from Experiment 2 of Lewis-Peacock et al. (2012). (A) In the first phase, participants

performed short-term recognition of a pseudoword (phonological STM), a word (semantic STM), or two lines (visual STM). (B) In the second phase, during the same scanning session, participants performed short-term recognition with two stimuli (between-category combinations of pseudowords, words, and lines). On half of the trials, the same memory item was selected as behaviorally relevant by the first and second cues (repeat trials), and on the other half of trials the second cue selected the previously uncued item (switch trials).

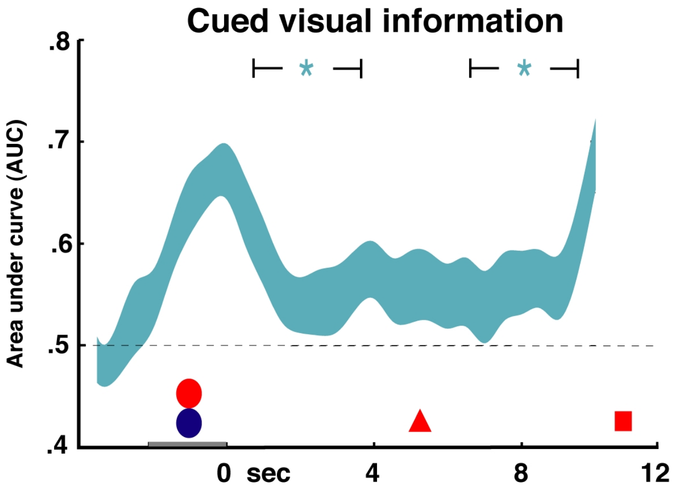
(C) Classifier decoding from Experiment 2 of Lewis-Peacock et al. (2012). Results are shown separately for repeat (left) and switch (right) trials. Classifier evidence values for phonological, semantic, and visual were relabeled and collapsed across all trials into three new categories: *cued* (red, the category of the memory item selected by the first cue), *other* (blue, the category of the other memory item), and *irrel* (gray, the trial-irrelevant category). The colored shapes along the horizontal axis indicate the onset of the targets (red and blue circles, 0 sec), the first cue (red triangle, 10 sec), the first recognition probe (red square, 18 sec), the second cue (red or blue triangle, 22 sec), and the final recognition probe (red or blue square, 30 sec). Data for each category are shown as ribbons whose thickness indicate ±1 SEM across participants, interpolated across the 23 discrete data points in the trial-averaged data. Statistical comparisons of evidence values focused on within-subject differences. For every 2-sec interval throughout the trial, color-coded circles along the top of each graph indicate that the classifier's evidence for the *cued* or *other* categories, respectively, was reliably stronger ($p < .002$, based on repeated measures t tests, corrected for multiple comparisons) than the evidence for the *irrel* category. (Classification was performed at the whole-brain level, no in restricted ROIs. "Importance maps" showing which brain regions contributed most importantly to the discrimination of each stimulus category are presented in Figure 2 of Lewis-Peacock and Postle (2012).)

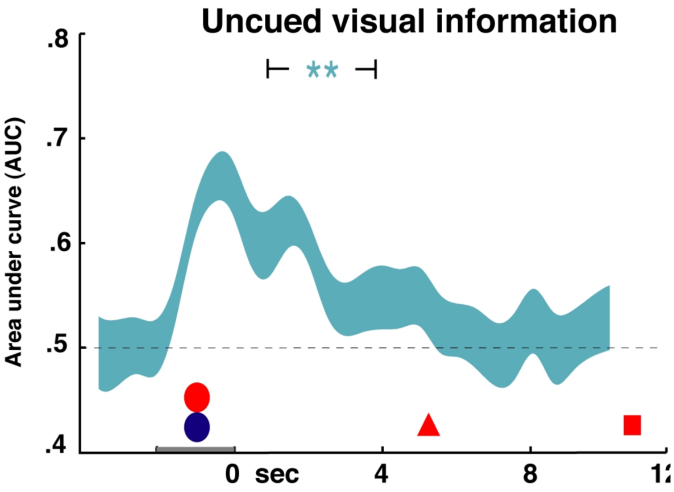
Reprinted with permission from Jarrod A. Lewis-Peacock, Andrew T. Drysdale, Klaus Oberauer, and Bradley R. Postle, 'Neural Evidence for a Distinction between Short-term Memory and the Focus of

Attention', Journal of Cognitive Neuroscience, 24:1 (January, 2012), pp. 61-79. © 2012 by the Massachusetts Institute of Technology.

Figure 3. Performance of MVPA cross validation decoding of trials, from LaRocque et al. (2013), when a visual stimulus was the AMI during Delay 2 (“Cued visual information”) and when it was the UMI during Delay 2 (“Uncued visual information”). Classifier sensitivity to the visual category, determined by calculating the area under the ROC curve (AUC), is plotted for each k-fold cross validation analysis time window, averaged across subjects. The width of the ribbon corresponds to the standard error of the AUC, calculated across subjects. Width of brackets surrounding significance markers indicates extent of delay period used for statistical analysis. Time is represented on the horizontal axis, with stimulus presentation (circles) from -2 to 0 s, the first cue (triangle) at 5 s, and the first probe (square) at 10.5 s. *$p<0.05$, **$p<0.005$.

Reprinted with permission from LaRocque, J.J., Lewis-Peacock, J.A., Drysdale, A.T., Oberauer, K. & Postle, B.R., 'Decoding Attended Information in Short-Term Memory: An EEG Study', Journal of Cognitive Neuroscience, 25:1 (January, 2013), pp. 127-142. © 2013 by the Massachusetts Institute of Technology.


Cued visual information
Area under curve (AUC)
.8
.7
.6
.5
.4
*
*
0 sec
4
8
12


Uncued visual information
Area under curve (AUC)
.8
.7
.6
.5
.4
**
0 sec
4
8