The cognitive neuroscience of visual short-term memory
Bradley R Postle

Our understanding of the neural bases of visual short-term memory (STM), the ability to mentally retain information over short periods of time, is being reshaped by two important developments: the application of methods from statistical machine learning, often a variant of multivariate pattern analysis (MVPA), to functional magnetic resonance imaging (fMRI) and electroencephalographic (EEG) data sets; and advances in our understanding of the physiology and functions of neuronal oscillations. One consequence is that many commonly observed physiological ‘signatures’ that have previously been interpreted as directly related to the retention of information in visual STM may require reinterpretation as more general, state-related changes that can accompany cognitive-task performance. Another is important refinements of theoretical models of visual STM.

Addresses
University of Wisconsin-Madison, Psychology and Psychiatry, 1202 W. Johnson St., Madison, WI 53706, United States

Corresponding author: Postle, Bradley R (postle@wisc.edu)

Signal intensity-based versus multivariate analyses of fMRI data
Reconsidering the link between delay-period activity and ‘storage’

For decades, a governing assumption in STM research has been that the short-term retention of visual information is supported by regions that show elevated levels of activity during the delay period of STM tasks. Thus, for example, debates over the role of the prefrontal cortex (PFC) in STM and the related construct of working memory were framed in terms of whether or not its delay-period activity showed load-sensitivity — systematic variation of signal intensity as a function of memory set size [1–4]. Similarly, patterns of load-sensitive variation of activity in the intraparietal sulcus have been used to test and refine theoretical models about mechanisms underlying capacity limits in visual STM e.g., 5,6]. With the advent of MVPA, however, this signal-intensity assumption has been called into question.

A fundamental difference between MVPA and univariate signal intensity-based analyses is that the former does not entail thresholding the dataset before analysis, but, rather, analyzes the pattern produced by all elements in the sampled space. The analytic advantages to this approach are marked gains in sensitivity and specificity e.g., 7]. In the domain of visual STM, this was first demonstrated with the successful decoding of delay-period stimulus identity from early visual cortex, including V1, despite the absence of above-baseline delay-period activity [8,9]. Subsequently, it was demonstrated that although the short-term retention of specific directions of motion was decodable from medial and lateral occipital regions (despite the absence of elevated delay-period activity), this information was not decodable from regions of intraparietal sulcus and frontal cortex (including PFC) that nonetheless evinced robust elevated delay-period activity [10*]. Further, in these posterior areas the strength of MVPA decoding, a proxy for the fidelity of neural representation, declined with increasing memory load. Importantly, these changes in MVPA decoding predicted load-related declines in behavioral estimates of the precision of visual STM [11**] (Figure 1). Relatedly, an fMRI study using a forward encoding-model approach [12*] has demonstrated that interindividual differences in the dispersion (i.e., ‘sharpness’) of multivariate channel tuning functions in areas V1 and V2 predicts recall precision of STM for orientations [13**]. Thus, studies [11**] and [13**] indicate an important link between the fidelity of the distributed neural representation and the fidelity of the mental representation that it is assumed to support.

The localization of visual STM, and insight into mechanism

It is not the case that intraparietal sulcus and frontal cortex are inherently ‘indecodable’ (see Box 1), nor that they are never recruited for the short-term retention of information. A determinant of whether a network will be engaged in the short-term retention of a particular kind of information is whether it is engaged in the perception or other processing of that information in situations that do not explicitly require STM. Thus, for example, when the short-term retention of abstract visuospatial patterns [23] or dynamically morphing flow-field stimuli [24] is tested, MVPA reveals delay-period stimulus representation in intraparietal sulcus, in addition to occipital regions; the same is true for face, house, and human-body stimuli in ventral occipitotemporal regions (e.g., [20**]). When the to-be-remembered stimulus affords oculomotor planning, its identity can also be decoded from oculomotor-control regions of intraparietal sulcus and of frontal cortex [25**].
Figure 1

(a) Visual short-term memory

(b) Estimated Memory Precision (Concentration [K])

(c) Peak classifier direction sensitivity

Sample Delay
Indeed, [25**] demonstrated that an MVPA classifier trained on only one condition — attention to a location, planning a saccade to a location, or STM for a location — can decode the other two. This could only be possible if similar patterns of neural activity, implying similar mechanisms, underlie the behaviors that have traditionally been categorized as ‘attention’ versus ‘intention’ versus ‘retention’.

Patterns of localization can also reflect how the brain supports the strategic recoding of information from the format presented at study into one best suited for the impending memory-guided action. One study first presented subjects with a sample object, then, early in the delay, indicated whether memory for fine-grained perceptual details or for category membership would be tested. For the former, MVPA found evidence for delay-period stimulus representation in inferior occipito-temporal cortex, but not PFC; for the latter, the converse was true [19**]. Combining MVPA with univariate and functional connectivity analyses has revealed a role for frontal cortex and intraparietal sulcus in implementing such strategic shifts of mental coding in visual STM [20**]. MVPA can also track the evolution of mental coding in the absence of instructions, demonstrating, for example, that the verbal recoding of visually presented information also entails the recruitment of a semantic code [26].

Neural data also provide important constraints on models of capacity limitations of visual STM [27**]. One influential model holds inferior intraparietal sulcus to be important for individuating objects that are to be encoded into visual STM, whereas superior intraparietal sulcus and an area of lateral occipital cortex are responsible for identifying these objects [6]. Recently, however, although the univariate analyses of data from a follow-up experiment [29**] did reproduce many of the findings from the earlier study, MVPA of the same data failed to support a model of segregated circuits performing these two operations. Instead, the study of Naughtin et al. [29*] produced two novel findings. First, the contrasts intended to operationalize individuation versus identification recruited primarily overlapping regions, thereby calling into question the dissociability of these two hypothesized mechanisms. Second, many regions outside of the intraparietal sulcus regions emphasized by [6] were also sensitive to these contrasts, suggesting that broadly distributed systems underlie the control of visual STM (Box 2).

**Signal intensity-based versus multivariate analyses of EEG data**

**Event-related potential (ERP) correlates of STM**

Another neural effect that has influenced models of visual STM capacity limitation is the contralateral delay activity (CDA), an ERP component that scales monotonically with STM load, but asymptotes at the psychophysically estimated capacity of an individual [34]. The

(Figure 1 Legend) Dissociating elevated delay-period signal from the short-term retention of information. Summary of results from [11**], in which subjects were scanned with fMRI while viewing one, two, or three sample displays of moving dots, then probed to recall the direction of one. (a) Univariate statistical maps indicating regions showing load sensitivity during sample presentation, the delay period, or both. (b) Time series data from 'sample-only' voxels (panel on left) or 'delay-only' voxels. Teal waveform illustrates decoding performance of a classifier trained at the time point with the maximal stimulus-evoked response (indicated with dot) then swept across the remainder of the trial. Maroon and solid gray waveforms are the analogous performance of classifiers trained at a time point late in the delay period, or 2 sec before sample onset, respectively. Asterisks indicate better-than-chance decoding at $p < .05(*)$ and $p < .01**(*)$. Superimposed is the trial-averaged BOLD activity, depicted in the dotted waveform and aligned with the vertical axis on the right-hand side of the plot. C. Plots of neural precision against behavioral precision. Each color corresponds to an individual subject and each digit (3, 2, or 1) to that individual’s neural and behavioral precision at the corresponding memory load. Lines are the fit indicated by ANCOVA ($r^2 = .35$).
CDA is widely interpreted as an index of the short-term retention of information (e.g., [35]), such that, for example, the presence of a CDA during visual search has been taken as evidence for ‘memory in search’ [36,37], and the diminution of the CDA across consecutive trials requiring search for the same target as evidence for the ‘handoff’ of the mnemonic representation of the search template from STM to LTM [38].

Not unlike with univariate analyses of fMRI data, however, there can be problems with equating a 1-D, signal intensity-based measure like the CDA with a single psychological construct (in this case, the short-term retention of information). For example, empirically, the CDA can be observed during tasks for which it is unclear that the short-term retention of information is required, such as during multiple object tracking [39], or during change detection ‘even when the observers know that the objects will not disappear from the visual field’ [40] (p. 8257). Further, the CDA during STM and during visual search is markedly reduced after intensive visual working memory training, despite the fact that STM capacity is increased and search performance improves with training [41*]. Under these conditions, a physiological marker specific to the short-term retention of information would be expected to increase in intensity. An additional challenge to the idea that the CDA is specific to the short-term retention of information comes from the proposal that it may, in fact, be the consequence of averaging across trials containing asymmetric amplitude modulation of alpha-band oscillations [42]. From this perspective, because the CDA is linked to alpha-band oscillations (and, hence, to a general aspect of neurophysiological state, such as cortical excitability or inhibitory tone), the CDA may not index a memory storage mechanism per se, but rather a ‘general mechanism for allocation of resources’ [43] (p. 903). Perhaps relatedly, multivariate analyses of alpha-band dynamics have provided important new insights into the neural bases of the short-term retention of visual information.

**Multivariate analysis of EEG in STM**

Using a multivariate forward-encoding-model approach similar to [13**], Anderson et al. [44**] constructed channel tuning functions for two narrowly filtered components of the EEG: alpha-band oscillations that were *evoked* by memory-sample onset; and alpha-band oscillations whose amplitude, but not phase, was modulated by sample onset (i.e., *induced*). Their results indicated that spatially distributed patterns in induced — but not evoked — delay period-spanning alpha-band activity predicted both inter-subject and intra-subject variation in precision of STM for line orientation. Note that these results do not necessarily implicate induced alpha-band oscillations in the delay-period representation, *per se*, of stimuli. Alternatively, they may reflect distributed patterns of local inhibition and/or the long-range synchronization of localized representations of features, either of which would nonetheless be unique to each stimulus (cf [17**]). Although several oscillatory phenomena have been associated with the short-term retention of information (including, e.g., local field potential oscillations at different frequencies, local and distal cross-frequency coupling, phase-amplitude coupling, and long-distance spike-field coherence (reviewed, e.g., in [45*]), their investigation with multivariate methods (e.g., [46]) will be an important step in determining their specificity for stimulus representation versus their possible contributions to other processes engaged by STM tasks.

**Do distributed patterns of activity reflect STM or attention?**

The multivariate methods reviewed here draw on two longstanding assumptions about STM. First, that stimulus representation is accomplished by anatomically distributed networks. Second, that the short-term retention of these representations is accomplished via elevated activity in these networks. Most often, however, STM tasks confound the focus of attention with the short-term retention, *per se*, of information. Recent studies have addressed this by first presenting two sample items, then indicating with a delay-period retcove which of the two will be relevant for the impending memory probe. (Thus, the cue designates an ‘attended memory item.’) Because the first memory probe will be followed by a second delay period, a second retcove, and a second probe, the item that was not cued during the initial delay (the ‘unattended memory item’) must be retained in STM, because it may be cued as relevant for the second probe. Intriguingly, MVPA of fMRI [47**] and EEG [48] variants of this task fail to find evidence for an active neural representation of the unattended memory item, even though its active neural representation is reinstated if it is selected by the second retcove (Figure 2).¹ These findings provide empirical support for the possibility that elevated activity may correspond more directly to the focus of attention than to the short-term retention of information, *per se*. The short-term retention of information, by this account, may depend on the establishment of representations encoded in distributed patterns of transiently modified synaptic weights, a code that would not be detectible by activity-based measurements. This phenomenon has been observed directly in the PFC of monkeys performing a visual working-memory task [15**], and has been simulated in many computational implementations [49*]. It has also been inferred to support the short-term retention of visual information in inferotemporal cortex [50], and so need not be assumed to be a PFC-specific phenomenon. An

---

¹ Note that, although [47**,48] decoded delay-period activity at the category level, and may therefore have lacked the sensitivity to detect the active representation of a single item, this finding has been replicated with item-level MVPA for STM for specific directions of motion, thereby reducing concerns that poor sensitivity may explain failure to find evidence for an active representation of UMI [LaRocque, Riggall, Emrich, and Postle, unpublished data].
important focus of current study is whether there are differences between the neural representation of unattended memory items, which are presumed to passively 'slip out of' the focus of attention versus of items that are intentionally removed from STM [20**,35].

**Conclusion**

High-level cognition, including STM, emerges from dynamic, distributed neural interactions that unfold on multiple time scales. The adoption of methods that more closely align with these principles of brain function is leading to discoveries with important implications for cognitive models of STM and working memory (e.g., [51,52]), and is informing ongoing research into such questions as the factors that underlie capacity limitations of visual STM [27*,28*], and the relation between STM and attention (e.g., [53,54]).

**Conflict of interest statement**

I declare that I have no conflict of interest.

**Acknowledgments**

I thank Nathan Rose for helpful comments on this manuscript, and Adam Riggall for help with figures. The author was supported by National Institutes of Health grants MH064498 and MH095984.

**References and recommended reading**

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


The first demonstration with MVPa that elevated delay-period activity may not correspond to stimulus representation per se.


This study both failed to find MVPa evidence for stimulus representation in frontal and parietal regions showing load-sensitive delay-period activity, and demonstrated that the fidelity of the neural representation in extrastriate cortex predicts the behavioral precision of STM.


A cogent, accessible tutorial introduction of principles underlying multivariate encoding models, and their potential for understanding brain-behavior links.


Provides strong evidence for a sensory recruitment model of visual STM by confirming, with multivariate encoding models, the hypothesis that ‘the relative ‘quality’ of [multivariate] patterns [of activity in visual cortex] should determine the clarity of an individual’s memory.’ (p. 754).


Extensive review of evidence from human neuroimaging and neuropsychology, and from monkey electrophysiology, in support of theory that regions of PFC and intraparietal sulcus are key nodes in a ‘multiple demand’ network that underlies many aspects of cognitive control.


Population-level MVPa reveals that PFC transitions through several high-dimensional states during a working memory trial, across which an initial stimulus representation is superceded by representations of trial context and behavioral choice. This high-dimensional trajectory may be supported by ‘hidden states’ of patterned change in networks of synapses, rather than as states that are ‘explicitly’ manifest in firing rates.


With analyses and conclusions broadly consistent with [15*], these authors also emphasize the dynamic nature of the distributed patterns of activity in PFC, noting, for example, that ‘task-relevant information in several neurons was present for only short periods of time relative to the duration of the ... delay period ... [and so] 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’ (p. 4652).


Multivariate analyses of population-level activity in PFC reveal the delay-period representation of factors that, although not explicitly required by the task, are presumably nonetheless experienced by the monkey. Additionally, findings from local field potentials (LFPs) may help relate intracranial electrophysiology to extracranial measures like the CDA [55].


Neuronal recordings show population-level delay-period stimulus representation (detection of motion) in spiking patterns in motion-sensitive visual region MST and in PFC, and in a broad-band range of the local field potential (LFP) in area MT, plus significant spike-field coherence between PFC and the beta-band range of the LFP in MT. This highlights important questions about interregional dynamics, and interregional differences in stimulus coding that must also be addressed in future studies of human STM.


Demonstrated that the neural systems that represent a stimulus in STM vary depending on the informational format required by the task. Thus, in the human, as with the monkey [33], category membership is represented in different neural systems than is perceptual detail.


Systematic MVPa of stimulus representation in occipitotemporal regions when item is being perceived, versus when it is in STM with varying levels of attentional status. Additionally, the supplementation with univariate and functional-connectivity analyses illustrates how PFC and parietal regions interact with distributed occipitotemporal stimulus representations.


Together with [10], the first MVPa study to report the failure to identify delay-period stimulus representation in PFC.


By demonstrating that an MVPa classifier trained on only one condition — attention to, saccade planning for, or STM for, a location — can decode the other two, this study provided definitive evidence for sensorimotor recruitment models of visual STM.

26. Lewis-Peacock JA, Drysdale A, Postle BR: Neural evidence for the flexible control of mental representations. Cerebral Cortex, in press.


Recent articulation of the ‘slots’ model, whereby visual STM capacity limits arise from a structural limit to the number of discrete visual object representations that can be actively maintained simultaneously.


Summary of recent evidence in support of ‘shared resource’ models that, in contrast to slots models, account for visual STM capacity limitations as resulting from the depletion of a single resource ‘shared’ by multiple items for their active representation.
29. Naughtin CK, Mattingley JB, Dux PE: Distributed and overlapping neural bases for object individuation and identification. Cerebral Cortex, in press. Clear illustration, via head-to-head comparison of univariate versus MVPA analyses of an fMRI dataset, of how the former are inherently biased toward producing results that support localizationist models, whereas the latter can reveal that functions in question are supported by more broadly distributed, and likely less ‘functionally specific’, networks.


44. Anderson DE, Serences JT, Vogel EK, Awh E: Induced alpha rhythms track the content and quality of visual working memory representations with high temporal precision. J Neurosci 2014, 34:7587-7599. This application of multivariate encoding models to EEG data provides strong, specific evidence that the short-term retention of visual information is supported by changes in the physiological state of circuits that were active before stimulus presentation, and that contribute to the perception of this information, as well as to its STM.


47. Lewis-Peacock JA, Drysdale A, Oberauer K, Postle BR: Neural evidence for a distinction between short-term memory and the focus of attention. J Cogn Neurosci 2012, 24:61-79. First report, with MVPA, of a failure to find evidence for the active neural representation of information that is in STM, but presumably outside the focus of attention.


