The Neural Underpinnings of Short-term Memory for Visual Motion

By

Adam C. Riggall

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Psychology)

at the

UNIVERSITY OF WISCONSIN–MADISON

2014

Date of final oral examination: 5/15/2014

The dissertation is approved by the following members of the Final Oral Committee:

Bradley R. Postle, Professor, Psychology
Timothy T. Rogers, Associate Professor, Psychology
Bas Rokers, Assistant Professor, Psychology
C. Shawn Green, Assistant Professor, Psychology
Xin Huang, Assistant Professor, Neuroscience

Acknowledgments

Thank you, Brad, for fighting for me from the very beginning to the very end. Without your unwavering intellectual, emotional, and financial support, none of this would have been possible. When I left Madison after my visiting weekend, I knew it would be the perfect place for me to learn and grow, but I had no idea how amazing the whole experience would be. You constantly pushed me to be a better scientist and a better person, and for that I will always be greatful.

I would also like to that the members of my committee for their time and energy during this exciting time. Tim Rogers, Bas Rokers, Shawn Green and Xin Huang, you have all helped insure I'm the very best researcher I can be. Thanks as well to Keith Kluender, Xiaojin (Jerry) Zhu, and Tatiana Pasternak, all of whom provided valuable direction at various points along the journey

A special thanks as well to the those who allowed to me get involved in brain science in the first place. Thank you Mike Gazzaniga for reading my out-of-the-blue email and allowing me to start my journey in your lab. Thanks to Jack Van Horn for finding a truly unique way for me to get involved with the fMRI Data Center and promote neuroimaging more broadly. Finally, thanks to Clay Curtis for bringing me to NYU as a lab manager and system administrator, allowing me to get fully involved in research, and then supporting me when I left for Wisconsin.

To all the PostLab member over the years, I would not have made it without you: Jeff Johnson, Jarrod Lewis-Peacock, Alex Shackman, Andrew Drysdale, Ellen Guller, Bornali Kundu, Dave Sutterer, Stephen Emrich, Josh LaRocque, Mike Starrett, Nate Rose, Olivia Gosseries, Drew Sheldon, and Jason Samaha. It's impossible to overestimate the influence that the daily interactions I've had with each of you have had on my learning and thinking.

Thank you to all the participants who volunteered to be research subjects. I know it wasn't always the most entertaining or pleasant of tasks, but, quite literally, this work would be nothing without your brainpower!

I would like to thank the staff at the Lane Neuroimaging Laboratory for their assistance during fMRI data collection as well as the staff at the High-Throughput Computing Center (CHTC) for

providing an amazing resource for large computational problems. Thanks as well to our funding sources from the NIH. A big thank you to Kevin Belt for taking care of numerous, little, last-minute issues with administrative paperwork, requirements, and payroll.

A huge thank you to all of the developers who open source their software and share it with the world. Without your often thankless work on the following tools, next to none of this work would be possible: AFNI, SPM, PyMVPA, the Princeton MVPA package, Psychtoolbox, D30 Inverse Solver, R, IPython, Vim. Also thanks to Cynthia Brewer and Mark Harrower for the incredibly useful ColorBrewer palettes.

And finally, endless thanks to those people, though not directly involved in the research, who have made this journey possible. Firstly, to my parents, Ginna and Fred Riggall. You have always valued learning and pushed me to better myself. I am who am I am today because of you. And lastly, to Josie. You have been nothing but supportive and have gone out of your way countless times to help me in any way you could. I needed it often, and you were always there. I only hope I can always be there to return the favor. Thank you!

Contents

Abstract

The visual world is complex and ever changing. To cope in such an environment, the visual system must be able to filter and maintain information, even when that information is no longer visually present. This ability, termed visual short-term memory, is critical for action planning and other higher-order cognitive functions. In the series of studies detailed in this dissertation, we have sought to understand the neural processes supporting this short-term storage, focusing on the retention of visual motion information. We began by analyzing patterns of brain activity, as measured with functional magnetic resonance imaging (fMRI), during a delayed-recognition task for single dot motion stimuli. Using a multi-voxel pattern analysis (MVPA) approach, we demonstrated the successful decoding of individual directions of remembered motion (stimulus-specific decoding) throughout the memory delay of the trial. Critically, this decoding was only possible using data from posterior visual regions (those regions traditionally associated with the initial processing of sensory information), and not from frontoparietal regions that demonstrated robust, elevated activity during the memory delay. We followed this up with two additional fMRI studies looking at memory for multiple visual motion directions, one a delayed-recall task for sequentially presented motion directions, the other a delayed-discrimination task for simultaneous, transparently presented motion direction. Across these two studies we observed similar results, with no evidence for stimulus representation in the frontal and parietal areas that showed a load-dependent effect during the delay, but rather strong evidence for representation within posterior visual areas. We were also able to demonstrate a relationship between the patterns of activity observed during different memory loads and the precision with which participants maintained items, as well as evidence for a recoding of the representation throughout the delay. We also performed a detailed analysis of the specific patterns of blood-oxygen level-dependent (BOLD) activity underlying classification performance during our single item task, investing the spatial distribution and temporal stability of these representations. Having demonstrated that frontal and parietal regions are not involved in the storage of short-term memory information, we sought to investigate what other

functions these regions might be performing during short-term memory by looking at the connectivity between these elevated delay-period regions and the posterior visual regions identified as performing the actual storage. In a final study we collected fMRI data while subjects performed a delayed-recall task for visual motion or color. This data was used to identify areas that were reliably active across the direction and colors trials. These areas were then targeted with transcranial magnetic stimulation (TMS) during a subsequent electroencephalography (EEG) session, during which participants performed the same delayed-recall task for visual motion and color. Results from connectivity analyses with the fMRI data and analyses of the effects of task and memory load on the TMS-evoked response in the EEG suggested a role for these frontoparietal regions in controlling top-down attention in the support of the memory representations in posterior regions. Taken together, the results from this work provide strong support for two closely related ideas. Firstly, that visual short-term memory is accomplished via the recruitment of sensory regions for storage (the sensory recruitment hypothesis), and that this recruitment is supported by the top-down guiding of internal attention by frontoparietal regions (the internal attention account).

1

Introduction

The visual world is complex, dynamic, and cluttered. Yet we have no problems living and interacting in this environment. Information, both relevant and irrelevant to our current behavioral goals, continually enters and exits the visual scene. To be successful in such a world, the visual system must retain and represent information that is no longer present in the visual stream. This ability, termed *visual short-term memory*, is critical for planning and executing actions and for integrating information over time. Once information has entered the visual system for initial processing, it must be maintained and protected so long as it might be relevant to the task at hand.

The underlying neural basis for this ability has been of interest to researchers for decades. How is it that information can be quickly sorted through and maintained within the brain for short periods of time? What computational mechanisms allow for such processing by the visual system? The current work addresses these important questions, focusing on the neural mechanisms supporting the short-term retention of visual motion information.

One popular idea for how this type of information maintenance might occur is via *persistent neural activity* that reverberates even in the absence of the original source of the activity. This idea was first proposed by Lorente de Nó (1933) and then expanded upon by Hebb (1949), who suggested that transient memory traces could be maintained by reverberations within assemblies of cells. This persistent reverberatory activity could thus bridge the gap between the initial sensory encoding of information and more permanent long-term storage that requires structural changes at the synapse level.

Evidence for this type persistent activity was first reported by Kubota and Niki (1971) in recordings done in neurons in dorsolateral prefrontal cortex (dlPFC) while monkeys performed a motor

delayed-alternation task. They identified a subset of neurons with elevated firing rates during the delay period of the task, between responses. At the same time, Fuster and Alexander (1971) recorded from neurons in the monkey prefrontal cortex (PFC) and in the medial dorsal thalamus, finding cells that changed firing frequency during the delay period of a delayed-response task. No evidence, however, was found to indicate that any of these cells where sensitive to the specific details of the food reward used to elicit the behavior. Niki (1974a,c,b) demonstrated the first conclusive evidence for selective, persistent delay-period activity in monkey dlPFC during both delayed-alternation and delayed-response tasks.

Such persistent activity has since been taken to be a hallmark of temporary storage. In the ensuing decades following the original discoveries, persistent delay-period activity has been found in numerous brain regions in the monkey across a wide variety of visual tasks: in inferior temporal (IT) cortex during serial recognition memory (Baylis and Rolls, 1987; Miller et al., 1993), in anterior ventral temporal cortex during object delayed-match to sample (Miyashita and Chang, 1988), in posterior parietal cortex (PPC; Gnadt and Andersen, 1988) and PFC (Funahashi et al., 1989, 1993b) during oculomotor delayed-response. While most studies of primate persistent activity require training on the task, it has also been observed in ventral premotor cortex without training, with cells firing persistently in response to a visual target even when the lights are turned off and the target is no longer visible (Graziano et al., 1997).

Persistent activity has also been observed in the human with a variety of different imaging methods in an impressively large number of brain areas across a wide variety of tasks (reviewed by Curtis and D'Esposito, 2003). The widespread existence of persistent activity throughout the animal kingdom and throughout the brain supports the original assertion by Hebb (1949) that reverberatory activity could be a broadly used mechanism for the temporary maintenance of information.

Given the strong empirical evidence for persistent neural activity, an equally strong computational modelling literature has sprung up to model this phenomenon. A wide range of models have been developed, at many different levels of abstraction, to attempt to explain how persistent

activity might be achieved within the brain (e.g., Goldman-Rakic, 1995; Durstewitz, 2009; Fransén et al., 2006; Wei et al., 2012).

Recent work, however, has begun to call into question the assumption that persistent activity is indicative of storage and to explore other ways that information may be maintained for short periods of time. One promising proposal embeds storage within the phase of ongoing neural oscillations (Lisman and Idiart, 1995; Jensen, 2006; Axmacher et al., 2006; Uhlhaas et al., 2009). Another promising avenue includes work exploring whether synaptic mechanisms might be involved, such as the residual calcium levels at the presynaptic terminal (Mongillo et al., 2008) or through GluR1-dependent short-term potentiation (Erickson et al., 2010). Such work is in the very earliest stages, so it will be some time before each theory can be properly assessed. The extent to which these models complement or contrast with spiking explanations will help to explain recent empirical findings suggesting information storage in brain regions that do not appear to show persistent activity (Harrison and Tong, 2009; Serences et al., 2009; Crowe et al., 2010; Riggall and Postle, 2012; Christophel et al., 2012)

Closely related to the question of *how* information is retained during short-term memory is the question of *where* in the brain this storage takes place. Considerable focus has been placed on the PFC as the locus of storage, following the proposal by Goldman-Rakic (1987, 1990) that the persistent activity observed in PFC represents the storage of information in a specialized buffer, as famously proposed by Baddeley and Hitch in their multiple component model of working memory (Baddeley and Hitch, 1974; Baddeley, 1987).

An alternative to the specialized storage buffer account is the *sensory recruitment hypothesis*. This theory posits that short-term storage is accomplished using the same neural systems used for sensory processing, rather than specialized storage buffers (Awh and Jonides, 2001; Pasternak and Greenlee, 2005; Jonides et al., 2005; Postle, 2006; D'Esposito, 2007; Serences et al., 2009; Sreenivasan et al., 2014a). Recent empirical evidence for this has idea has emerged, showing storage of orientation (Harrison and Tong, 2009; Serences et al., 2009; Ester et al., 2009; Albers et al., 2013), more complex stimuli (Christophel et al., 2012; Christophel and Haynes, 2014; Vicente-

Grabovetsky et al., 2014), and visual objects (Linden et al., 2012) within posterior visual regions specialized for the sensory processing of these features.

Understanding the processes supporting the short-term storage of information are further complicated by the fact that we are able to store not just one, but many items simultaneously. Thus any accurate description of the mechanisms supporting short-term retention must explain multi-item memory. From a descriptive standpoint, based on behavioral and neuroimaging results, there are currently two popular models for how visual short-term memory storage might occur. One model posits that short-term memory is best described as slot-based, with individuals having a specific number of 'slots' with which to hold memory items (Pashler, 1988; Luck and Vogel, 1997; Zhang and Luck, 2008). Once the number of slots is exhausted, no additional items can be retained in memory. The slots are envisioned to be able to hold items of roughly any complexity. Under this model, visual short-term storage is limited by the number of items in memory, not the specific features of the items themselves.

The second model suggests that rather than a specific number of slots, short-term memory is better thought of as being supported by a shared resource (Wilken and Ma, 2004; Bays and Husain, 2008; Bays et al., 2009). Under this framework, each individual has a finite amount of "memory resources" that can be used for storage. When only storing one item, all the resources can be used for that item. As more and more items are added to memory, each receives fewer and fewer resources (as resources are split among items in memory) and the resolution or detail with which they can be represented declines.

The aim of the current work is to address these two questions concerning *how* and *where* the brain stores visual information short-term. A series of studies was conducted, using short-term memory for visual motion as a representative system, exploring these questions in humans with functional magnetic resonance imaging (fMRI), transcranial magnetic stimulation (TMS), and electroencephalography (EEG). Visual motion is a particularly interesting feature to use to explore short-term memory, as motion is ubiquitous in the real world, it is relatively complex, and it's initial sensory processing is one of the best understood of all the visual subsystems (Newsome et al., 1990;

Born and Bradley, 2005).

The first study, discussed in Chapter 2, directly tackles the question of where in the brain information is retained during short-term memory. We scanned participants (fMRI) while they performed a delayed-recognition task for simple dot motion stimuli. Our goal was to identify, with multi-voxel pattern analysis (MVPA) methods, patterns of brain activity during the memory delay that corresponded to specific motion directions in memory. We hypothesized that if elevated, persistent activity is a marker of information storage, we would be able to decode stimulus-specific remembered motion information from these regions. We found no evidence for stimulus information within these regions during the delay period. Instead, we found this information in posterior visual regions which were active during the initial processing of the motion information, but showed no elevated activity during the memory delay.

Chapter 3 covers a follow-up fMRI study which required participants to remember one, two, or three directions of motion (presented sequentially) per trial. It also changed the task requirements from recognition to recall, allowing us to look at the effects of memory load on the precision of memory representations. We again used MVPA to probe for patterns of activity related to specific directions of motion in memory. We tested the hypothesis that areas showing a parametric increase in activity during the delay with increased memory load (rather than overall elevated activity, per se), would contain stimulus representations. As in the first study, we found no evidence for stimulus representation in the hypothesized areas, instead finding such evidence again in posterior visual regions. We also investigated the relationship between the precision of the recalled behavioral responses and the neural representations in posterior areas at different memory loads, demonstrating that the evidence for specific patterns in memory at different loads correlated with the precision with which individual subjects were able to recall those motion directions.

In a second follow-up (Chapter 4), we increased the difficulty of the task by switching to transparent motion, lengthening the memory delay period, and requiring participants to compare remembered directions with a new probe direction and report the direction that the probe had been rotated relative to the remembered item. The switch to transparent motion, where multiple di-

rections of motion are presented completely overlapping within a single aperture simultaneously, removed the individuation cues (of temporal sequence or spatial position) often used to individuate items in multiple item memory tasks. We hypothesized that individual directions of motion would be recoverable (using MVPA) from patterns of activity in posterior visual regions, which turned out to be true for memory loads of one and two. We also investigated the evolution of the patterns of activity supporting representation during the trial, finding evidence suggesting a recoding process happening, within posterior visual regions, after the initial sample presentation during the delay.

These first three studies all used a similar methodology (MVPA with fMRI data) to answer important theoretical questions about the short-term storage of visual motion information. Chapter 5 looks in detail at the specifics of exactly what it is that the pattern classifiers are using to identify and separate specific directions of motion in memory. By looking directly at the weights learned by the classier during training, we can infer information about the underlying representations supporting memory. This allowed us to answer questions about the temporal stability and spatial distribution of these representations.

The final study (Chapter 6) sought to answer the question of what functions, if not storage, the frontal and parietal areas (which show reliable elevated activity during the delay-period) are performing during short-term memory. We hypothesized that these regions are involved in the top-down direction of attentional resources to coordinate the recruitment of sensory resources for storage and in protecting remembered information from disruption. That is, these regions are functionally connected to the posterior visual regions performing the actual storage. To test this hypothesis, we had participants perform a delayed-recall task for direction of motion or colors over two sessions, an fMRI session and a TMS/EEG session. Participants performed the same task in both conditions, alternating between blocks of remember direction trials and blocks of remember color trials. Data from the fMRI session were used to define functionally active targets for TMS. These data were also used to look for psychophysiological interactions (PPI) between posterior visual regions and frontal and parietal sites. During the TMS/EEG session, TMS was applied to the

middle temporal area (MT), intraparietal sulcus (IPS), or lateral prefrontal cortex (lPFC) during the delay period. An analysis of the changes in the TMS-evoked response across task conditions and memory loads was used to assess effective connectivity changes between frontoparietal and occipital regions.

Together these studies provide strong support for the sensory recruitment hypothesis: short-term retention of visual motion information is mediated by the same cortical regions used for sensory processing. Though not directly involved in storage, load sensitive areas such as PFC and IPS play a critical role in controlling attention to maintain these representations.

2

The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging

Does the sustained, elevated neural activity observed during working memory (WM) tasks reflect the short-term retention of information? Functional magnetic resonance imaging (fMRI) data of delayed-recognition of visual motion in human participants were analyzed with two methods, a general linear model (GLM) and MVPA. Although the GLM identified sustained, elevated delay-period activity in superior and lateral frontal cortex and in IPS, pattern classifiers were unable to recover trial-specific stimulus information from these delay-active regions. The converse—no sustained, elevated delay-period activity, but successful classification of trial-specific stimulus information—was true of posterior visual regions, including MT+ (which contains both middle temporal area and medial superior temporal area) and calcarine and pericalcarine cortex. In contrast to stimulus information, pattern classifiers were able to extract trial-specific task instruction-related information from frontal and parietal areas showing elevated delay-period activity. Thus, the elevated delay-period activity that is measured with fMRI may reflect processes other than the storage, per se, of trial-specific stimulus information. It may be that the short-term storage of stimulus information is represented in patterns of (statistically) "subthreshold" activity distributed across regions of low-level sensory cortex that univariate methods cannot detect.

Adam C. Riggall & Bradley R. Postle

Data published in 2012: *J Neurosci* 32(38), 12990-12998.

2.1 Introduction

For decades, a prevailing view has been that working memory (WM) storage is accomplished via sustained, elevated neural activity. Such activity, first identified with extracellular recordings in the non-human primate (Fuster and Alexander, 1971; Niki, 1974a; Funahashi et al., 1989), has been observed in numerous areas of the human brain with functional magnetic resonance imaging (fMRI; Curtis and D'Esposito, 2003). The idea of a role for delay-period activity in storage is reinforced by its sensitivity to manipulation of memory-influencing factors, such as persistence across varying delay lengths, and variation of signal magnitude with memory load (Vogel and Machizawa, 2004; Xu and Chun, 2006; Postle, 2006).

There are, however, complications with the mnemonic interpretation of delay-period activity. One is cases of activity that appear mnemonic, but can be shown to support other functions, such as attention or response preparation (Fuster, 2002; Lebedev et al., 2004). Relatedly, lesion-induced deficits originally interpreted as mnemonic (Jacobson, 1936; Funahashi et al., 1993a) have subsequently been reinterpreted as reflecting factors other than memory per se (Malmo, 1942; Tsujimoto and Postle, 2012).

A second complication is that delay-period activity can fail to show properties thought necessary for a mnemonic signal. In one such case, although fMRI activity at several sites was elevated throughout a long delay period (24 seconds), none showed load sensitivity, leaving uncertain whether these regions contribute to storage (Jha and McCarthy, 2000). In another, monkeys showed excellent short-term memory (STM) for direction of motion despite the absence of directionally tuned neurons in either the middle temporal area (MT) or the prefrontal cortex (PFC) that sustained elevated activity across the delay (Bisley et al., 2004; Zaksas and Pasternak, 2006; Hussar and Pasternak, 2012).

A third complication relates to assumptions of homogeneity of function, often only implied or tacit, but nonetheless inherent, in massively univariate analyses of neuroimaging data. When activity is identified in a large volume of "activated" tissue, the extraction of a spatially averaged signal from contiguous voxels necessarily assumes that all are "doing the same thing". Further, its

interpretation entails assuming that this locally homogeneous activity can be construed as supporting a mental function independent of other brain areas. These assumptions, however, are difficult to reconcile with the increasingly common recognition that neural representations are high-dimensional and supported by anatomically distributed, dynamic computations (Norman et al., 2006; Kriegeskorte et al., 2006; Bullmore and Sporns, 2009; Cohen, 2011).

Given these complications, we sought to test core assumptions about elevated delay-period fMRI activity using an information-based analysis. Multivariate pattern classification was used to test two hypotheses about STM for visual motion information: 1) that elevated delay-period activity carries trial-specific stimulus information, and 2) that trial-specific stimulus information can be encoded in subthreshold patterns of activity. The first tests an assumption that has underlain most neuroimaging research on WM. Confirmation of the second would extend findings for visual STM for oriented gratings (Serences et al., 2009; Harrison and Tong, 2009) and offer insight about the physiological basis of STM for visual motion (Hussar and Pasternak, 2012).

2.2 Materials and Methods

To test our hypotheses we scanned subjects (fMRI) while they performed a delayed-recognition task for visual motion (Fig. 2.1). We then trained pattern classifiers to discriminate the direction of motion from individual time points in the trial and tested to see how these classifiers labeled the data from all the other time points during the trial (creating a decoding time series). A similar approach has been shown to be sensitive to the dynamics of memory content, capturing changes in the memory trace on a continuous basis throughout the trial (Polyn et al., 2005; Lewis-Peacock and Postle, 2008, 2012).

In contrast to analysis methods that assess the magnitude of activity (whether the firing rate of a neuron or the strength of the blood-oxygen level-dependent (BOLD) signal of one or a group of voxels), information-based analyses, such as pattern classification, focus on obtaining quantitative measures of the information content within a given area (Kriegeskorte et al., 2006; Kriegeskorte,

2011). This approach, often termed multi-voxel pattern analysis (MVPA), uses machine learning methods to identify patterns of activity that are reliably associated with different stimuli or categories of stimuli. The extent to which novel patterns of activity can then be correctly categorized provides a measure of the information available in the underlying voxels (Norman et al., 2006; Haynes, 2011; Jimura and Poldrack, 2012).

Participants

Ten volunteers (5 females) between 21 and 28 years of age (mean 23.8 years) were recruited from the undergraduate and graduate student community of the University of Wisconsin-Madison and paid for their participation. All subjects had normal or corrected-to-normal vision, no reported history of neurological disease, and no other contraindications for MRI. All subjects gave written informed consent according to the procedures approved by the Health Sciences Institutional Review Board at the University of Wisconsin-Madison. Three subjects (1 female) were excluded from the *Results* due to failure to perform the task to criterion level (see *Results* for more details).

Behavioral Paradigm

Participants were scanned while performing seven runs of a delayed-recognition task for visual motion. A schematic representation of the experiment is illustrated in Fig. 2.1. Each trial began with a fixation cross changing color to white, indicating to subjects that they needed to fixate the cross and prepare for the start of the trial. After 1.5 s a patch of coherently moving dots was presented (sample, 1 s). Participants were instructed to remember both the direction and the speed of this motion while maintaining fixation for the duration of the trial. The sample was followed by a 15 s delay-period. Seven seconds into the delay period the fixation cross changed color to indicate the dimension on which subjects would be making a match/non-match comparison judgement between the remembered motion and a new set of moving dots; blue indicated they should make the judgement based on the direction of the moving dots, ignoring speed, while magenta indicated that they should make the judgement based on speed, ignoring direction. Following the delay, a

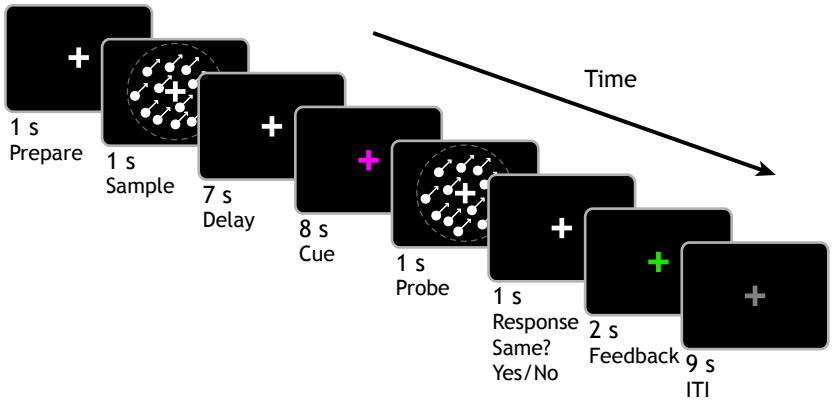

Figure 2.1: Behavioral task. Participants maintained the direction and speed of a sample motion stimulus over a long delay-period. Midway through this delay period they were cued as to the dimension on which they would be making an upcoming comparison, either direction or speed. At the end of the delay period they were presented with a probe motion stimulus and had to indicate with a button press whether it matched or did not match the sample stimulus on the cued dimension. The dashed gray circle indicates the shape of the aperture the dots appeared in and was not present within the stimulus display.

second patch of coherently moving dots (probe, 1 s) was presented and subjects were required to indicate as quickly as possible with one of two buttons if the probe motion matched or did not match the sample on the cued dimension. After a 1 s response period the fixation cross changed color to green if the subjects were correct or red if they were incorrect (feedback, 2 s). A 10 s intertrial interval (ITI) followed, during which the fixation cross changed color to gray and subjects were instructed they could break fixation and relax their eyes.

Sample and probe stimuli consisted of circular patches (15° diameter) of coherently moving dots. Sample motion could be in one of four directions (42°, 132°, 222°, 312°) and at one of three speeds (4°/s, 8°/s, 12°/s). Directions were chosen to be off the cardinal axes to reduce potential verbalizations. Probe stimuli on match trials had the same value in the cued dimension, and a different value from the non-cued dimension (e.g., for a match trial where direction was cued the sample might have moved toward 132° at 8°/s, while the matching probe might have moved toward 132° at 4°/s). On direction-cued non-match trials the probe stimulus was rotated (randomly clockwise or counter-clockwise) by a degree threshold value from the direction of sample

and the speed was randomly drawn from one of the two values not used in the sample. On speed-cued non-match trials the probe stimulus speed was changed (randomly increased or decreased) by a proportional threshold value from the speed of the sample and the direction was randomly drawn from one of the three values not used in the sample. The threshold values for non-match direction and speed were updated using separate adaptive staircases (Levitt, 1971) to keep performance around 75% correct. The logic of this approach was that by holding task difficulty constant throughout an experimental session, we could assume comparable fidelity of representation on trials for which the response ended up being correct versus incorrect. This would maximize sensitivity by allowing inclusion of all trials in the analyses. Before scanning, subjects practiced a block of 24 trials to familiarize themselves with the experimental procedure and to determine starting threshold values for non-match trials.

Participants completed 168 trials over the course of 7 runs while in the scanner. Sample stimuli included all possible pairwise combinations of directions and speeds (12 different combinations, each seen 14 times). Direction was cued on 96 trials, with the remaining 72 cued for speed. This disparity balanced the overall number of individual examples at each dimension value across the cued dimensions. Match and non-match trials were equally likely. The experimental stimuli were controlled by E-Prime 2.0 (Psychology Software Tools) and viewed through fiber-optic goggles mounted on the head coil (SV-7012; Avotec). Participants responded via two buttons on a fiber-optic button box (Psychology Software Tools).

Subjects also performed one block of an MT+ localalizer task, similar to that used by Huk et al. (2002). In summary, participants viewed alternating 18 second blocks of stationary and moving dot patterns (alternating from expanding and contracting every second) within a circular aperture (15°) while maintaining fixation, for a total of 8 blocks of each.

Data Acquisition and Preprocessing

Whole brain images were acquired with the 3T scanner (Discovery MR750; GE Healthcare) at the Lane Neuroimaging Laboratory at the University of Wisconsin-Madison. For all subjects a high-

resolution T_1-weighted image was acquired with a fast spoiled gradient-recalled-echo (FSPGR) sequence (8.132 ms TR, 3.18 ms TE, 12° flip angle,156 axial slices, 256 x 256 in-plane, 1.0 mm isotropic). A gradient-echo, echo-planar sequence (2 s TR, 25 ms TE) was used to acquire data sensitive to the BOLD signal within a 64 x 64 matrix (39 sagittal slices, 3.5 mm isotropic). Seven runs of the delayed-recognition task were obtained for each subject, each lasting 12 min, 8 sec (364 volumes). All task runs were preceded by 10 sec of dummy pulses to achieve a steady state of tissue magnetization. One run of the MT+ localizer was obtained for each subject, lasting 4.8 minutes (144 volumes).

The functional data were preprocessed using the Analysis of Functional NeuroImages (AFNI) software package (Cox, 1996). All volumes were spatially aligned to the final volume of the final run using a rigid-body realignment and corrected for slice time acquisition. Linear, quadratic and cubic trends were removed from each run to reduce the influence of scanner drift. For univariate analyses, data were spatially smoothed with a 6 mm FWHM Gaussian kernel and transformed into Talairach space (Talairach and Tournoux, 1988). For classification analyses, data were z-scored separately within run for each voxel. Data were not smoothed and were left in their native space.

Univariate Analyses

Each within-trial event of the delayed-recognition task (i.e., sample, pre-cue delay, cue, post-cue delay, probe, response) was modeled separately for direction- and speed- cued trials. Sample and probe were modeled as 1 s boxcars, each delay as a boxcar of appropriate duration, and the cue and response as impulses. All were convolved with a canonical hemodynamic response function (HRF). Each of these independent regressors was entered into a modified general linear model (GLM) for analysis using AFNI. For the present purposes, a more generous boxcar-shaped covariate was used to model the delay-periods (rather than a more conservative mid-delay delta function (Zarahn et al., 1999; Postle et al., 2000)), to ensure that we would not miss any delay-active voxels, although at the expense of likely also being sensitive to some variance that is attributable to the sample presentation. In this way, we implemented a generous feature selection step that

included as many "delay active" voxels as possible for MVPA, thereby being careful not to exclude any such voxels that may potentially carry stimulus-specific information. The localizer was modeled with boxcars for both stationary and moving dot patterns. This localizer was used to insure that the regions of interest used for the MVPA analysis included MT+.

Pattern Classification Analyses

Classification was performed using the Princeton Multi-Voxel Pattern Analysis (http://www.pni.princeton.edu/mvpa) toolbox and custom routines in MATLAB (The MathWorks, Natick, MA, USA). Preprocessed fMRI data from individual trial time points were used to train separate classifiers to classify the direction of motion (4 possible directions) or the speed of motion (3 possible speeds) in the sample and by inference, the direction/speed of motion in memory (stimulus-specific classification), or to classify whether the subject had been cued that direction or speed was the relevant dimension on a given trial (trial-dimension classification).

Classification was accomplished using L2-regularized logistic regression (LR), a linear classification approach used widely in the machine learning community and well suited for application to fMRI data as it tends to generalize well after learning in high-dimensional feature spaces with limited training examples (Pereira et al., 2009). The lambda penalty term, which reduces the contribution of less informative voxels to classification and thus improves generalization, was determined ($\lambda = 25$) by repeating the wholebrain testing procedure described below for penalty terms at powers of 10 from -4 to 4, then at a finer grained resolution within the best interval. The penalty term was chosen to maximize the mean decoding performance across all subjects. During decoding, a trial was considered correctly classified if the correct direction/speed had the highest likelihood estimate (winner-take-all classification). Overall classification accuracy was determined using leave-one-trial-out cross-validation, in which the classifier was repeatedly trained on data from all but one trial, and then tested on the left out trial, rotating through all the trials as the left out testing trial.

Stimulus-specific classifiers were always trained with data from direction-cued trials or speed-

cued trials, never both. Testing of trained classifiers was done on trials of both type. This allowed us to compare how representations changed when subjects were cued that they would be judging speed, and thus direction was no longer relevant, and vice versa. Trial-dimension classifiers were trained and tested on all data.

In order to examine the dynamics of the memory trace, each classifier was trained using data from only a single time point in the trial (e.g., the first volume acquisition following the target) and then tested on all time points in the left-out trial (i.e., including both time points preceding and following the training time point). The result of this procedure is a time course of decoding accuracy for the entire trial. By doing the initial training of the classifier using different time points in the trial (e.g., a time point just after the sample, a time point in the later part of the delay, etc.), it was possible to estimate the stability of a given representation throughout the duration of the trial.

Classification was initially performed on whole brain data that had undergone a basic feature-selection step whereby only those voxels that showed a main effect for task ($t > 2$) in the univariate GLM were included. This step was included to reduce the chances of overfitting during training. Subsequent region of interest (ROI)-based analyses used only those voxels within individual ROIs, created from the intersection of anatomically defined ROIs and voxels that showed either significant sustained delay-period activity or no delay-period activity, depending on the specific hypothesis. Four anatomically defined ROIs were hand drawn for each subject by tracing gray matter on the high-resolution anatomical scans: frontal, parietal, lateral occipital and temporal, and medial occipital. The frontal region included the entire precentral sulcus (PCS) and the posterior portion of the inferior frontal sulcus (IFS). For finer-grained analyses, this was subdivided into three frontal ROIs that showed robust delay-period activity: the superior rostral bank of the PCS bounded superiorly by the intersection of the superior frontal sulcus (SFS); a more inferior portion of the rostral bank of the PCS bounded ventrally by the intersection with the IFS; and the caudal third of the IFS. The parietal region included the entire intraparietal sulcus (IPS) and the superior parietal lobule (SPL). For finer-grained analyses, it was also subdivided into three ROIs:

medial-caudal IPS comprising the descending segment and the caudal half of the horizontal segment, dorsolateral IPS comprising the rostral half of the horizontal segment and the ascending segment, and SPL. The lateral occipital and temporal region included all of the lateral occipital gyrus, the fusiform gyrus, the posterior portion of the middle and inferior temporal gyri and the posterior portion of the inferior temporal sulcus. The medial occipital region covered the medial portion of the occipital lobe from the lingual sulcus to the occipitoparietal sulcus, including all of the calcarine sulcus.

The significance of classifier performance was determined using a random permutation test (Golland and Fischl, 2003) to determine the likelihood of observing a specific accuracy under the null hypothesis that there is no relationship between the data and the specific class labels used to train the classifier (directions/speeds of motion). A null distribution was generated by rerunning the entire classification cross-validation procedure 1000 times, randomly shuffling the class labels each time. A p-value was then computed by determining the proportion of permuted accuracies that were higher than the observed accuracy. This procedure was repeated for all classification results.

2.3 Results

Behavioral Results

Task difficulty was equated across subjects by feeding real-time, trial-by-trial performance information to a staircasing algorithm that dynamically adjusted the difficulty of non-matching probe stimuli (direction and speed independently) so as to maintain performance at a target level of 75% correct. Overall mean behavioral performance ($n = 7$) was 80.38% correct for direction trials and 80.32% correct for speed trials, both slightly better than the target performance level because several subjects reached a predefined minimum threshold value, at which point the staircase procedure could no longer reduce the threshold to further increase task difficulty. The average non-match direction threshold was 10.9° (SD 5.33°). The average non-match speed threshold was a

38% change (SD 15.3%). Three subjects were dropped from the study because their inordinately high thresholds indicated that they were responding at random.

Univariate Results

To test the first hypothesis, that elevated delay-period activity carries trial-specific stimulus information and can thus be interpreted as a neural correlate of memory storage, we first identified areas showing elevated, sustained delay-period activity by solving a massively univariate GLM for each subject. The resultant individual thresholded statistical maps were then used, on a subject-by-subject basis, to select the voxels used for training the classifiers. Group-level statistical maps for the sample-evoked (Fig. 2.2A) and delay-period (Fig. 2.2B) activity illustrate several important characteristics. Activity evoked by the sample was widespread, located bilaterally in posterior visual areas, lateral occipitotemporal regions including the inferior temporal sulcus (including the putative MT+ complex identified with the localizer), IPS, posterior IFS and SFS, and PCS. Delay-period activity was more circumscribed, with clusters of significant activity ($p < 0.05$, false discovery rate (FDR) corrected) bilaterally in the IPS, IFS, and PCS. This pattern of elevated activity agrees with many other findings from studies of visuo-spatial WM (Curtis and D'Esposito, 2003).

Whole Brain Pattern Classification

Before directly testing our hypotheses, it was necessary to demonstrate that motion information could be decoded on a time point-by-time point basis. All analyses were performed on single-subject data, with statistical reliability subsequently assessed across the sample. For each subject's data, a set of four classifiers was trained to discriminate the direction of the sample motion stimulus (and by inference, the remembered motion direction; 4 possible directions), each using feature-selected whole-brain data from direction-cued trials and restricted to a single time point within the trial. The first classifier (*Sample*) was trained on data from only the time point 4 seconds after the sample onset, which corresponds to the peak of the sample-evoked response. The second classifier

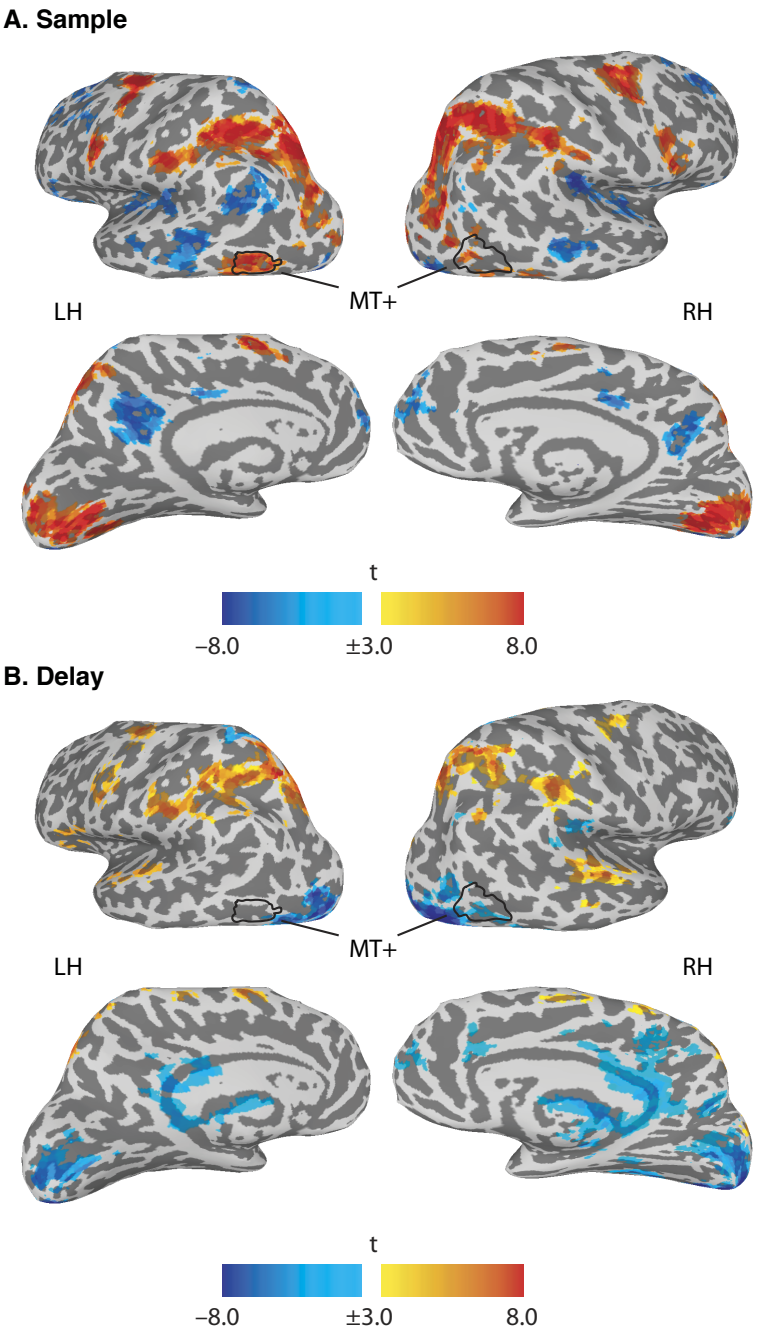

Figure 2.2: Univariate GLM results. (*A*) Sample-evoked, and (*B*) Delay-related activity, as estimated from a group-level GLM, thresholded at $p < 0.05$, FDR corrected, and displayed on a representative subject's inflated surface. Note that images are for illustrative purposes only, as all analyses were performed on single-subject data. Superimposed is an outline of MT+ as defined by the localizer. Note for this region that it is robustly activated by the sample, but that its activity does not differ from the baseline during the delay period. A qualitatively similar pattern is observed in calcarine and pericalcarine cortex. LH, Left hemisphere; RH, right hemisphere.

(*Late Delay*) was trained on data from the time point 16 seconds after sample onset, corresponding to the post-cue portion of the delay period, prior to the onset of the probe stimulus. The third classifier (*Probe*) was trained on data from the time point 20 seconds after sample onset (4 seconds after probe onset), corresponding to the peak of the probe-evoked response. The final classifier (*ITI*) was trained on data from the time point 26 seconds after sample onset, corresponding to the middle of the intertrial interval. This classifier was included as a control, as we would not expect there to be any stimulus-specific information retained once the trial had been completed.

Each trained classifier was then used to construct a decoding time course of direction representation on every trial time point from the held-out direction-cued trials. This approach allowed us to detect changes in the neural representation of information across the trial. For example, by training a classifier on data from very early in the trial (i.e., *Sample*), we would expect to capture early, sensory-based representations. By testing such a classifier on every time point in the trial, we could determine if such a representation remains stable throughout the trial, or if it deteriorates over time. Similarly, training a classifier with data from a time point late in the delay-period (*Late Delay*) might capture a recoded representation (e.g. verbal or numeric or clock-face), which would be expected to be absent at the beginning of the trial and to strengthen over the course of the delay-period when we look at the time point-by-time point decoding performance of that classifier. By limiting our training data to a single time point during the trial, we hoped to maximize our ability to resolve time point-by-time point changes, at the expense of lower signal-to-noise for the classifier inputs.

Results from this analysis can be seen in Fig. 2.3A. The mean decoding accuracy for the *Sample* classifier was significantly above chance (25 %, $p < 0.05$, permutation test) throughout the entire delay period. Similar results were obtained with the *Late Delay* and *Probe* classifiers. These results suggest that the memory representation remains relatively stable and unchanging across the delay period. Decoding accuracies for the *ITI* classifier were always near chance, which was expected given the absence of any stimulus-related information for the classifier to learn (i.e., signal from 8 sec after the offset of the memory probe would not be expected to carry information about the

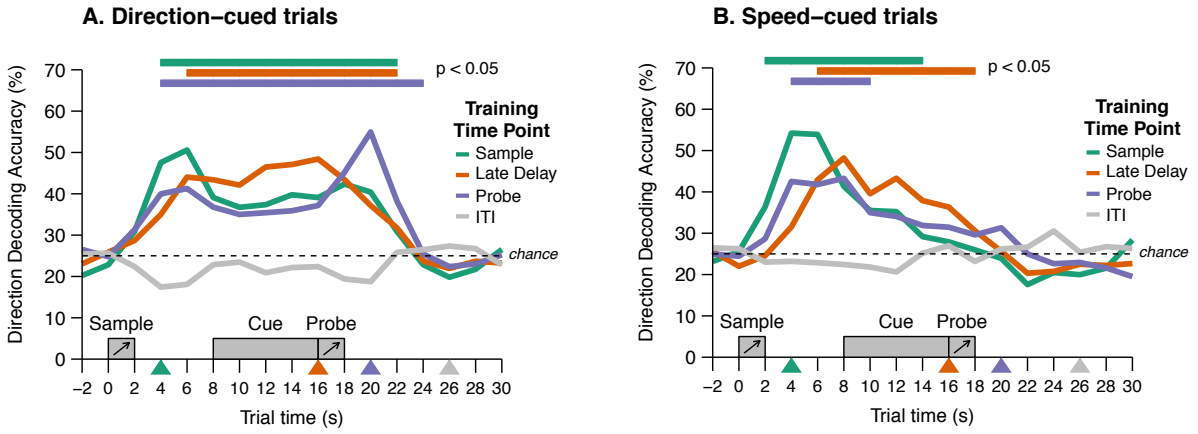

Figure 2.3: Whole brain direction-decoding results. Decoding time courses after training on whole-brain data from direction-cued trials. (*A*) Decoding of direction information from direction-cued trials. (*B*) Decoding of direction information from speed-cued trials. Each waveform represents the mean 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.

stimulus from the preceding trial). As with the univariate analyses, all trials were included in the analyses. Follow-up analyses using only correct trials produced qualitatively similar results.

To assess the specificity of the classification, the set of classifiers trained to discriminate direction with data from direction-cued trials were also used to decode direction information from speed-cued trials. (Fig. 2.3B). Classification with the *Sample*, *Late Delay* and *Probe* classifiers was above chance for time points before the cue, when subjects needed to hold both speed and direction information in memory, but fell to chance levels following the cue, suggesting that subjects discarded direction information when it was no longer relevant to the current trial.

Two features in the data confirm that successful *Late Delay* decoding of direction-cued trials represents the sustained retention of stimulus information and does not reflect an artifact of the slow recovery of the hemodynamic response, nor effects of motion adaptation. First, if the results were purely driven by the residual hemodynamic response to the sample stimulus, or to adaptation,

we would expect to see similar decoding performance on both direction-cued and speed-cued trials, because the sample stimuli are identical across these two conditions. Second, such "residual" effects would also result in successful decoding by the *ITI* classifier, which was trained with data 10 seconds following the probe stimulus (compared to 16 seconds following the sample for the *Late Delay*), which it clearly does not.

The whole-brain classification procedure was also applied to the stimulus dimension of speed, using data from only speed-cued trials. Unlike direction, however, classifier performance never exceeded chance for any time point during the trial. Because subjects performed these trials at the same level of proficiency as direction-cued trials, we interpret this null result to mean that the representation of speed, at least across the range used in this experiment, may be carried out on too fine a spatial scale neurally to be discriminated with our fMRI procedure. It is also possible that by collapsing across different directions when classifying the speed trials we added too much noise to the signal to classify, given the close relationship between speed and direction (Born and Bradley, 2005). Additionally, the use of only three speeds may have encouraged a coding strategy that varied over time (e.g., verbal labels that changed as the stimuli became more familiar to subjects). The remainder of the *Results* and *Discussion* will focus on decoding direction information.

ROI-Constrained Classification

To test our first hypothesis—that elevated delay-period activity carries stimulus-specific information—we repeated the classification procedure that we used with the whole brain data, but applied it to only those voxels in the frontal or parietal cortices that showed elevated delay-period activity as identified with the univariate analysis. As can be seen in the average BOLD time series in Fig. 2.4A and 2.4B, each of these areas showed elevated activity that was sustained throughout the delay period. However, decoding performance for motion direction never differed from chance in either area, regardless of the time point used to train the classifier (Fig. 2.4E and 2.4F). Therefore, we failed to find evidence that the sustained, delay-period BOLD activity in these regions carried stimulus-related information; the first hypothesis was not supported. To rule out the possibility

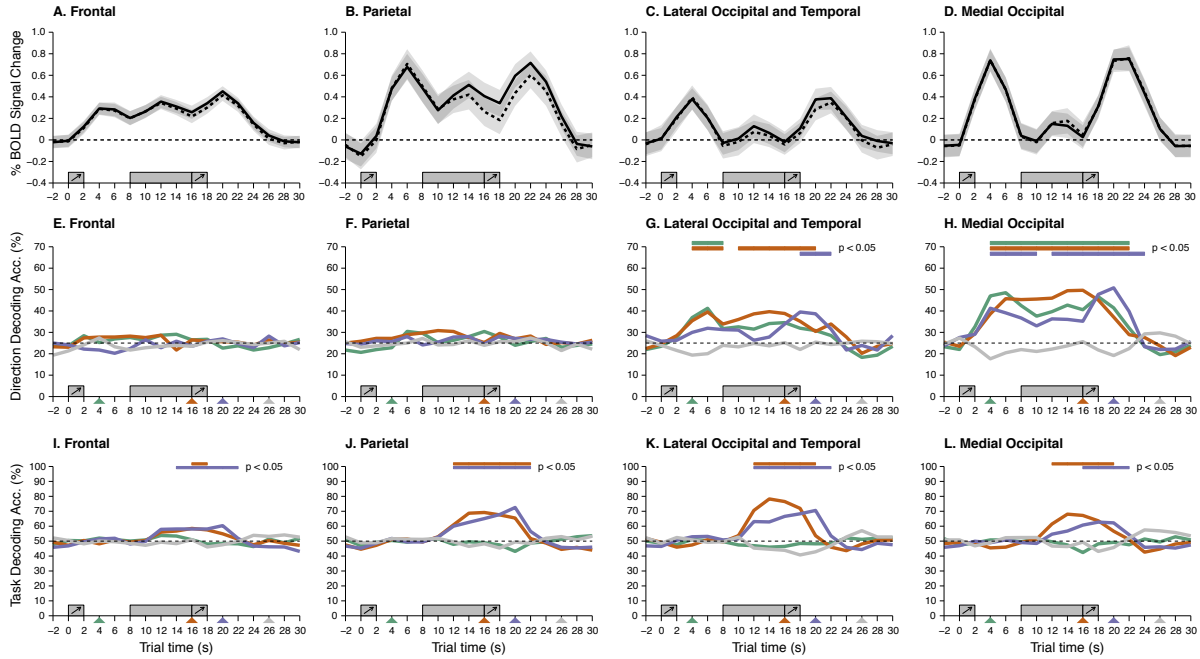

Figure 2.4: ROI BOLD and decoding time courses. (*A-D*) Average ROI BOLD activity. Data from direction-cued trials use solid lines and speed-cued trials use dashed-lines, bands cover average standard error across subjects. (*E-H*) ROI stimulus-direction decoding results and (*I-L*) ROI trial-dimension decoding results. Graphical conventions same as Fig. 2.3. All averaged across individual data from 7 subjects.

that using such large ROIs may have obscured the presence of stimulus-representation in smaller regions, we repeated the ROI-classification with the smaller, more specific ROIs described in the *Methods* section. Results in all smaller ROIs mirrored those of the larger regions: no evidence for stimulus-related information was found in any of the regions.

To test our second hypothesis—that stimulus-specific information can be recovered from sub-threshold patterns of activity (i.e., activity whose signal intensity does not surpass a statistical threshold in conventional univariate analysis)—we repeated the classification procedure as above, limiting the training data to only voxels in medial occipital or lateral occipital and temporal cortex that showed no evidence of elevated activity during the delay period in the GLM. As can be seen in the average BOLD time series in Fig. 2.4C and 2.4D, these regions showed large sample- and probe-evoked responses, as well as smaller cue-evoked responses, but no sustained, elevated

delay-period activity. Decoding performance from these regions, however, was significantly above chance throughout the delay-period (Fig. 2.4G and 2.4H).

Although these posterior regions did not show sustained delay-period activity at the group level (Fig. 2.2B.), there were voxels in the individual-subject data of each subject that did. The results of pattern classification did not change appreciably when these voxels were included or excluded from the analyses. Additionally, when the classifiers were trained only on these posterior delay-active voxels, in no subject was decoding performance sustained at above-chance levels across the delay period. Overall, these findings were consistent with the second hypothesis, that brain regions can carry stimulus-specific information in a sustained manner despite the absence of sustained, above-baseline levels of activity.

These results form a double dissociation, with frontal and parietal regions showing elevated, sustained delay-period activity but no delay-period stimulus representation, and posterior regions the converse. One possible concern about applying these results to our understanding of WM storage, however, is that doing so requires the acceptance of the null MVPA findings in frontal and parietal cortex. Might it be the case, for example, that these regions are simply less amenable to MVPA (perhaps, e.g., because they represent information at a finer grain of spatial detail than is measurable with our fMRI methods)? To address this possibility, we trained a new set of classifiers on a different discrimination – cue identity (i.e., whether the relevant stimulus dimension was direction or speed) – and repeated the procedure with each of the four ROIs. As shown in Fig. 2.4I-L, the relevant trial-dimension was decodable from each of the four ROIs. Importantly, for each this was only true for the *Late Delay* and *Probe* time-point trained classifiers, and only for time points following cue presentation. These results indicate that frontal and parietal regions are not inherently "undecodeable" with our methods, and, thus, lend more credence to the possibility that they did not represent stimulus-specific information during the delay period of our WM task.

2.4 Discussion

The aim of the present study was to test long-standing views about the relationship between the short-term retention of information and sustained delay-period activity. Using an information-based analysis approach with fMRI data collected during a delayed-recognition task for visual motion, we tested two hypotheses: first, that sustained, elevated delay-period activity carries stimulus-specific information; and second, that stimulus information can be encoded in distributed patterns of subthreshold activity. To test the first hypothesis we trained pattern classifiers with BOLD signal from frontal and parietal areas that showed sustained, elevated delay-period activity. We failed to find evidence that these voxels carried stimulus-specific information during the delay period. To test the second hypothesis, we applied the same procedure to BOLD data from posterior regions that showed robust responses to visual stimuli, but no elevated delay-period activity. The classifiers were successfully able to decode the remembered direction throughout the delay period, providing strong evidence in support of this hypothesis.

The first finding can be seen as a failure to support an enduring assumption in cognitive neuroscience, albeit one that is increasingly being called into question (Curtis and D'Esposito, 2003; Lebedev et al., 2004; Curtis and Lee, 2010; Lewis-Peacock and Postle, 2012). Although on its own it might be qualified as a null result, there are several factors that must influence its interpretation. Most saliently, it is paired with a positive result using the same method, and derived from statistically "subthreshold" voxels located in areas that are active during the perception of the to-be-remembered information. Empirical evidence thus shows that this method is sensitive. Indeed, although there remains some controversy about the physiological and representational factors that underlie the patterns of activity that correspond to different brain states (Freeman et al., 2011; Thompson et al., 2011), we are not familiar with any suggestion that there may exist brain states to which MVPA is less sensitive than traditional analysis of activation levels of individual voxels or groups of voxels. To the contrary, the near-consensus view is that MVPA methods are more sensitive than traditional activation-based analyses (Kriegeskorte et al., 2006; Norman et al., 2006; Haynes, 2011; Lewis-Peacock and Postle, 2012; Jimura and Poldrack, 2012).

Further, although we cannot rule out the possibility that stimulus information might be represented in frontoparietal cortex at either a spatial scale that is too fine to be detected with our fMRI methods, or perhaps via a signal to which BOLD is relatively insensitive (e.g., low-frequency oscillations in local field potentials), we did demonstrate that this is not a limitation for the decoding of trial-specific task *instruction*-related information. From this perspective our results are consistent with, for example, the finding from monkeys that PFC and posterior parietal cortex represent the *category* to which a stimulus belongs (Freedman and Assad, 2006; Swaminathan and Freedman, 2012). It is also worthy of note that although MVPA has been applied successfully to sensory processing in topographically organized cortex (e.g., the decoding of orientation (Harrison and Tong, 2009; Serences et al., 2009)), it has also been successfully applied to "higher level" processing in polymodal cortex. Thus, for example, MVPA has demonstrated contextual reinstatement during episodic memory retrieval (Polyn et al., 2005), the recognition of individual faces (Kriegeskorte et al., 2007), and neural correlates of free choice (Soon et al., 2008), all entailing the decoding of information from polymodal temporal, parietal, and/or frontal cortex.

Consistent with our preferred interpretation of the null findings in frontal cortex are several factors. First, there are the results from extracellular recording in monkeys performing a similar task with similar stimuli, where no evidence for direction-selective persistent activity was found in the PFC throughout the delay period (Zaksas and Pasternak, 2006; Hussar and Pasternak, 2012). Second, a similar pattern to the MVPA results that we describe here has been reported for STM for four categories of visual objects (Linden et al., 2012) and for complex artificial visual stimuli (Christophel et al., 2012). Third, the fact that STM can be intact despite lesions of PFC (D'Esposito and Postle, 1999; Tsujimoto and Postle, 2012) is consistent with the failure to find physiological evidence for STM representations in this region.

The frontoparietal network that has been a focus of this study is known to support the endogenous control of attention (Corbetta and Shulman, 2002; Beck and Kastner, 2009; Noudoost et al., 2010). Interestingly, one account of WM storage is that it is supported by this same top-down mechanism (e.g., Armstrong et al., 2009; Curtis and D'Esposito, 2003; Postle, 2011). From

this perspective, the sustained delay-period activity observed in this study may correspond to a control signal that does not vary with stimulus identity. Future work would need to reconcile this possibility with the finding that multivariate patterns of frontoparietal activity do discriminate between directions of motion during a sustained attention task (Liu et al., 2011). In addition to specifically memory-related functions, many other functions might be supported by sustained delay-period activity of frontal and parietal regions. Because the frontal and parietal activity observed in the present study (Fig. 2.4) resembles activity that has been reported in countless prior neuroimaging studies (Curtis and D'Esposito, 2003), it may well be that it does not correspond to a stimulus-specific or even task-specific function. More general demands that many cognitive tasks (including working memory tasks) have in common include decision making (Curtis and Lee, 2010), prioritizing certain task-relevant representations and/or processes over others (Miller and Cohen, 2001), monitoring the environment to control the processing of potentially interfering exogenous events (Chao and Knight, 1998; Postle, 2005), actively representing a "behavioral set" (Woolgar et al., 2011), and monitoring behavior so as to prevent prepotent responses (Knight and D'Esposito, 2003), including perseverative responses (Milner, 1963; Tsujimoto and Postle, 2012). (Note that although the behavioral set account might be consistent with the successful decoding of cue identity in frontal and parietal regions, this explanation does not generalize to the first portion of the delay period.) This is, of course, an incomplete list.

One important question for future study is the nature of the mental codes with which subjects represent motion information across the delay period. In the monkey, a psychophysical study using backward masking provided evidence that the initial memory trace is perceptually based, retaining a high-fidelity representation of the sample (including such trial-irrelevant information as the local velocity of individual dots in the random-dot motion stimulus). This representation only endured few hundred milliseconds into the delay period, however, perhaps because, in this study, the animals could predict the major features of the impending memory probe (Zaksas et al., 2001). Although the BOLD signal did not afford high temporal resolution in the present study, results with classifiers trained on different portions of the trial suggested that the mnemonic repre-

sentation is relatively stable. We cannot know with certainty, however, whether this representation was primarily perceptual, motoric, or categorical in nature, or perhaps some combination of these. Our working assumption is that the mnemonic representation of direction was perceptually based, because it is from visual regions that we were able to recover stimulus direction information. Had subjects used, for example, a covert eye-movement strategy, we would have expected to have been able to decode stimulus information from frontal and parietal regions (Ikkai and Curtis, 2011). The same reasoning makes us skeptical that subjects depended on a verbal strategy for remembering either direction or speed. We did not, however, monitor eye movements, nor take steps to discourage covert speech.

The results presented here highlight the differing conclusions that can be drawn from activation- vs. information-based analyses of the same data set. In so doing, they raise questions about the longstanding belief that information retained during working memory is stored via sustained delay-period activity, preferentially in frontal and parietal cortex. Instead, the memory trace may be represented in patterns of "subthreshold" levels of activity distributed across regions of low-level sensory cortex.

3

Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory

Traditionally, load sensitivity of sustained, elevated activity has been taken as an index of storage for a limited number of items in visual short-term memory (VSTM). Recently, studies have demonstrated that the contents of a single item held in VSTM can be decoded from early visual cortex, despite the fact that these areas do not exhibit elevated, sustained activity. It is unknown, however, whether the patterns of neural activity decoded from sensory cortex change as a function of load, as one would expect from a region storing multiple representations. Here, we use multivoxel pattern analysis to examine the neural representations of VSTM in humans across multiple memory loads. In an important extension of previous findings, our results demonstrate that the contents of VSTM can be decoded from areas that exhibit a transient response to visual stimuli, but not from regions that exhibit elevated, sustained load-sensitive delay-period activity. Moreover, the neural information present in these transiently activated areas decreases significantly with increasing load, indicating load sensitivity of the patterns of activity that support VSTM maintenance. Importantly, the decrease in classification performance as a function of load is correlated with within-subject changes in mnemonic resolution. These findings indicate that distributed patterns of neural activity in putatively sensory visual cortex support the representation and precision of information in VSTM.

Stephen M. Emrich*, Adam C. Riggall*, Joshua J. LaRocque & Bradley R. Postle

* These authors contributed equally

Data published in 2013: J Neurosci 33(15), 6516-6523.

3.1 Introduction

Although performance on many everyday cognitive tasks depends on the ability to maintain and manipulate multiple items in visual short-term memory (VSTM), it remains unclear precisely how multiple visual items are represented simultaneously in the brain. A widely accepted hallmark of VSTM storage is the presence of sustained, elevated delay-period activity (Fuster and Alexander, 1971; Funahashi et al., 1989), which is thought to underlie the "active" mechanisms of short-term maintenance. In human neuroimaging, the sensitivity of this activity to load is considered even stronger evidence for VSTM storage. For example, functional magnetic resonance imaging (fMRI) signal in intraparietal sulcus (IPS) increases with increasing memory load and asymptotes at apparent capacity limitations (Linden et al., 2003; Xu and Chun, 2006).

More recently, studies using multi-voxel pattern analysis (MVPA) have demonstrated that the identity of a single item held in VSTM can be decoded during the delay period from early visual regions (e.g., V1-V4, MT), even in the absence of sustained delay-period activity (Harrison and Tong, 2009; Serences et al., 2009; Riggall and Postle, 2012). Although these more recent findings suggest that VSTM representations are coded in distributed patterns of activation in early visual cortex, it remains unclear whether such presumably sensory-based representations respond to changes in VSTM load and, critically, how these changes may reflect load-dependent changes in VSTM performance.

The present study investigated the sensitivity of sensory-cortex-based short-term memory representations to memory load by acquiring fMRI while subjects performed delayed recall of one, two, or three presented directions of motion (Fig. 3.1). Pattern classifiers were trained to classify the remembered direction of motion and we examined how direction-specific classifier evidence changed as a function of VSTM load. The manipulation of load also enabled us to investigate the neural basis of precision in VSTM. Recent evidence suggests that there is a measureable decline in the level of detail or "precision" of VSTM representations as information load increases (Bays and Husain, 2008; Zhang and Luck, 2008; van den Berg et al., 2012). According to these models, this decrease in mnemonic resolution reflects a decrease in the proportion of neural resources

dedicated to each representation. Accordingly, we predicted that this loss of mnemonic resolution as a function of load should be reflected in classification performance, reasoning that declines in both classification performance and behavioral precision may result from noisier neural representations. Relating classification performance to behavioral precision would be strongly consistent with the idea that patterns of activity in sensory cortex reflect the neural representation of information held in VSTM, even in the absence of sustained, elevated delay-period activity in these regions.

3.2 Matrials and Methods

Participants

Ten right-handed volunteers, three female, ages 23-31 years (median = 25.5 years), from the University of Wisconsin-Madison community participated in the study for a small remuneration ($15/h). All subjects provided informed consent according to the procedures approved by the Health Sciences Institutional Review Board at the University of Wisconsin-Madison. Subjects had normal or corrected-to-normal vision, no contraindications for MRI, and no reported history of neurological disease.

Stimuli and procedure

Participants underwent fMRI while performing six runs of a delayed recall task for visual motion (Fig. 3.1). A gray fixation cross (~0.22°) was presented at the beginning of each trial on a black background and was changed to white 2 s before the onset of the memory sample. The memory sample lasted a total of 2 s (1 TR), during which 3 patterns of dots were presented. Dot patterns were circular patches (~11° diameter) of dots (~0.13°) with a density of ~0.7 dots-per-square degree. Dot patterns were centered around the fixation and uniquely and randomly presented in red, green, or blue without repeating colors within a trial. Each dot pattern was presented for 500 ms and stimuli were separated by an interstimulus interval of 250 ms. On any given trial, one, two, or

three of the dot patterns could be moving with 100% coherence at 2.75°/s while the remaining dots (loads 1 and 2) remained static. On 90% of trials, one of the presented directions was sampled from one of three directions (7°, 127°, 247°). These were the directions used as labels for the patterns used to train the classifier and were also selected as the probed (target) directions on those trials. Because we wanted behavioral estimates of precision to correspond directly to the neural data, the probed item was always the stimulus with one of these three directions of motion. The remaining moving patterns on these trials were sampled randomly from the 360° space. No restrictions were placed on the selected directions of these remaining items. On the remaining 10% of trials, all moving patterns were randomly selected, including the probed direction, to prevent participants from learning the target directions. The sample display was followed by an 8 s delay period, over which participants were instructed to remember the direction of motion of all moving patterns. After the delay, a probe display appeared for 5 s. The probe consisted of a colored circle (~11° in diameter) with a line (~4.5°) extended between the center of the screen and random location on the circumference of the circle. Participants were instructed to rotate the line so that it matched as closely as possible the direction of motion of the color-matching dot pattern. Participants rotated the line by pressing buttons that rotated the line 10° clockwise, 10° counterclockwise, or flipped it 180°. The starting position of the line segment was randomly selected, with the constraint that it was presented within a multiple of 10° of the probed sample direction. On those trials in which one of the critical items was presented, the probe color always matched the color of the critical stimulus. On the remaining 10% of trials, the probed stimulus direction was randomly selected. The temporal position of the target stimulus was also counterbalanced within each block. A 7 s intertrial interval followed the probe, during which the fixation cross changed to gray.

Participants performed 180 trials over the six runs in the scanner. Each block contained 27 critical trials, three of each load X direction X temporal position pairing, as well as three non-critical trials, one of each load. The color of the sample was always randomly selected. Experimental stimuli were controlled by the Psychophysics Toolbox (http://psychtoolbox.org; Brainard, 1997) running in MATLAB (MathWorks), presented using a 60 Hz projector (Silent Vision 6011;

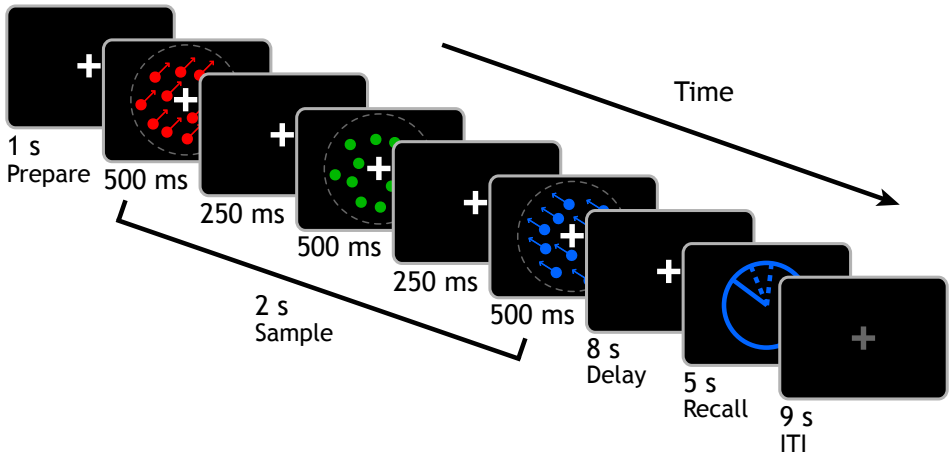

Figure 3.1: Behavioral task. Schematic of the delayed-recall task performed by participants while undergoing fMRI. On every trial, participants were presented with 3 patches of dots displayed within a single 2s TR and instructed to remember the direction of motion of the moving dots. Memory load was manipulated by varying the number of dot patterns that contained 100% coherent motion. After the 8 s delay, participants indicated the direction of the probed (color-matching) direction by rotating the angle of the line segment. The dashed gray circle indicates the shape of the aperture the dots appeared in and was not present within the stimulus display.

Avotec), and viewed through a coil-mounted mirror.

Data Acquisition and Preprocessing

Whole brain images were acquired with the 3T scanner (Discovery MR750; GE Healthcare) at the Lane Neuroimaging Laboratory at the University of Wisconsin-Madison. High-resolution T1-weighted images were acquired for all subjects with a fast spoiled gradient-recalled-echo (FSPGR) sequence (8.132 ms TR, 3.18 ms TE, 12° flip angle, 156 axial slices, 256 x 256 in-plane, 1.0 mm isotropic). blood-oxygen level-dependent (BOLD)-sensitive data were acquired using a gradient-echo, echoplanar sequence (2 s TR, 25 ms TE) within a 64 x 64 matrix (39 sagittal slices, 3.5 mm isotropic). Six runs, each lasting 13 min (390 volumes), were obtained for each subject. The first three TRs from each run were removed before analysis to achieve a steady state of tissue magnetization.

Behavioral Analysis

Behavioral performance was assessed using the mixture-model developed by Zhang and Luck (2008) and extended by Bays et al. (Bays et al., 2009; Zokaei et al., 2011). This model breaks down the distribution of error (the angular deviation between the sample direction and the indicated response direction) into estimates of target accuracy [proportion of responses toward the target direction (P_T)], error [nontarget responses (P_{NT}) and guesses (P_G)], and measure of precision (κ, or concentration; the variability of recall of the target and nontarget responses). That is, responses are modeled to be a mixture of a Von Mises distribution (a circular analog of the Gaussian distribution) with a concentration of κ for all target and nontarget responses plus a uniform distribution of random responses (guesses). The concentration parameter, κ, corresponds to the variability of the target (and nontarget) responses, with larger values indicating less variability. The model is defined as follows:

$$p(\hat{\theta}) = P_T \phi_\kappa(\hat{\theta} - \theta) + P_{NT} \frac{1}{m} \sum_{i}^{m} \phi_\kappa(\hat{\theta} - \theta_i^*) + P_G \frac{1}{2\pi}$$

where θ is the target motion direction, $\hat{\theta}$ is the response direction on a given trial, P_T is the probability of reporting the target direction, and P_{NT} is the probability of reporting a nontarget direction (of m nontarget directions). P_G is defined as $1 - P_T - P_{NT}$, and represents the probability of responding at random (Bays et al., 2009; Zokaei et al., 2011, each contain figures depicting the relationship between the response distribution and the model parameters).

Parameter estimates for P_T, P_{NT}, P_G, and κ were obtained using maximum-likelihood estimation (expectation maximization) using MATLAB routines (available at http://www.bayslab.com). Responses were entered for each trial, and separate estimates were obtained for each participant and condition.

Imaging Analyses

Functional data were preprocessed using the Analysis of Functional NeuroImages (AFNI) software package (http://afni.nimh.nih.gov; Cox, 1996). All volumes were spatially aligned to the final volume of the final run using a rigid-body realignment and corrected for slice-time acquisition. Linear, quadratic, and cubic trends were removed from each run to reduce the influence of scanner drift. For univariate analyses, data were spatially smoothed with a 6 mm FWHM Gaussian. For classification analyses, data were z-scored separately within run for each voxel. Data were not smoothed and were left in their native space.

Univariate region-of-interest analysis.

The goal of these analyses was to identify regions of interest (ROIs) in which the activity, either in response to stimulus presentation or during the delay period, varied monotonically with load. We did this by solving a modified general linear model (implemented in AFNI) in which regressors modeling the sample presentation and delay period were generated as 2 and 8 s boxcars, respectively, and convolved with the canonical hemodynamic response function supplied with AFNI. This approach is commonly used to discriminate transient, "stimulus-evoked" activity that returns to baseline upon the offset of the sample from activity that persists at an elevated level across the delay period despite the absence of stimulus input (Postle et al., 2000). The model also included three levels of stimulus load (1, 2, and 3).

Figure 3.3 illustrates regions in which "sample" and/or "delay" activity varied significantly with load. To construct ROIs for hypothesis testing MVPA analyses, the parameter estimates used to produce the map shown in Figure 3.3 were used to identify regions for which load sensitivity of the sample epoch was statistically greater than load sensitivity of the delay epoch ("Sample ROI") and regions for which the converse was true ("Delay ROI").

Anatomically defined ROIs.

In addition to the univariate ROIs, MVPA was also performed on anatomical ROIs created with the automatic parcellation routines (Fischl et al., 2008) implemented by FreeSurfer (http://surfer.nmr.mgh.harvard.edu/). These ROIs included: V1 and V2, middle temporal area (MT), intraoccipital sulcus (including the transverse occipital and superior occipital sulci) and intraparietal sulcus (including the transverse parietal sulcus). All voxels within the anatomical ROIs were included in the MVPA analyses.

Note that we did not include frontal areas among anatomical ROIs for two reasons. First, many previous studies using fMRI, fMRI-guided repetitive transcranial magnetic stimulation, and patients with frontal lesions have either failed to find load sensitive delay-period activity in frontal cortex (Postle et al., 1999; Feredoes and Postle, 2007) or failed to find that disruption of frontal cortex affects the retention of multiple items (D'Esposito and Postle, 1999; Postle et al., 2006; Feredoes et al., 2007; Koenigs et al., 2009). Second, two recent studies using MVPA methods similar to those used here failed to find evidence for delay-period retention of stimulus-related information in frontal cortex (Christophel et al., 2012; Riggall and Postle, 2012).

Pattern classification analyses.

To examine the neural representations associated with VSTM maintenance, we trained pattern classifiers to classify each of the three critical directions of motion and examined classifier sensitivity for each direction using a leave-one-trial-out approach. Classification was performed using the Princeton Multi-Voxel Pattern Analysis toolbox (http://www.pni.princeton.edu/mvpa/) and custom routines in MATLAB. Preprocessed fMRI data from individual trial time points were used to train separate classifiers to classify the direction of motion (three possible directions) or the memory load (three possible loads). Classification was accomplished using L2-regularized logistic regression with a lambda penalty term of $\lambda = 25$. Training for direction was performed collapsed across all three loads; similarly, training for load was performed collapsed across all three directions. Note that all MVPA analyses were performed on data that were neither smoothed

nor time-shifted in any other way. Classification was performed on the top 2000 voxels from the sample- and delay-load-sensitive general-linear-model-derived ROIs and on all voxels within the anatomically defined ROIs.

For each 2 s TR of fMRI data, the trained classifier produced a probability estimate (from 0-1) of the extent to which the observed pattern on the tested trial matched the trained pattern for each of the trained items (e.g., the three directions of motion). Classification performance was determined using leave-one-trial-out cross-validation, in which the classifier was trained on data from all but one trial and then tested on the remaining trial, rotating through all possible permutations. Training was performed using all possible trials (collapsing across the untrained dimension of load or direction). For each of the three motion directions, we computed a recieve-operating characteristic (ROC) based on the values obtained across all tested trials, and then averaged across the dimensions of interest. Therefore, this method evaluated whether, across trials, evidence for one direction of motion was higher when that direction was in fact presented on that trial relative to when it was not presented. The area under the ROC curve, averaged across the three directions of motion, provides a measure of overall classifier sensitivity, with chance performance at 0.5. Analogous methods of classification and ROC-based sensitivity analysis were used to classify memory load, collapsing across direction of motion.

To examine the dynamics of the memory representation, each classifier was trained on data from only a single time point in the trial (e.g., TR 5) and then tested on all points in the left-out trial. This resulted in a time course of decoding sensitivity for the entire trial that was determined by the weights of the single training point and provides a measure of how stable the representations were over time.

The significance of classifier performance was determined using one-tailed, one-sample *t tests*, testing against chance performance of 0.5. Tests performed across individual TRs were corrected for multiple comparisons using the method of Larzelere and Mulaik (1977). Within-subject correlations were computed using the method of Bland and Altman (1995) by calculating ANCOVA with the factors of classifier sensitivity and subjects as the covariates and behavioral performance

as the dependent measure, and fitting the model to parallel lines.

Within-subject correlation.

Although classification performance was significantly above chance at the group level, there was significant variance between subjects in the ability to classify direction of motion. Consequently, we attempted to examine the relationship between classification performance and behavioral precision while accounting for these individual differences in classification performance. A number of studies of VSTM have typically normalized between-subject differences in neural signal by calculating pairwise differences between loads (Vogel and Machizawa, 2004; Vogel et al., 2005; Emrich et al., 2009). Here we instead attempted to examine whether changes in the relative amount of neural information across all three tested memory loads predicted the change in the quality of representations maintained in VSTM; that is, we took advantage of the variation in dependent measures across all three loads by using a within-subject correlation approach (Bland and Altman, 1995) to examine the change in both classifier performance and behavioral precision across loads. Specifically, we used ANCOVA to remove between-subject differences in classification performance and measured the remaining variation in classification performance explained by variation in behavioral precision.

3.3 Results

Behavioral Results

Behavior was analyzed according to the method of Bays et al. (Bays et al., 2009; Zokaei et al., 2011). This method uses a mixture-model of the behavioral response error (the angular distance between the reported direction and the actual direction of the probed memory stimulus) to obtain maximum-likelihood estimates for the proportion of correctly recalled targets, the precision of those responses (the distribution of error around the target), and the proportion of errors (both nontarget responses and guesses; see Materials and Methods).

Examining the concentration parameter (κ, the estimate of the variability of memory precision) of this model using repeated-measures ANOVA revealed a significant effect of load ($F_{(2,18)} = 8.55$, mean squared error $[\mathrm{MSE}] = 500.4$, $p = 0.0025$; Fig. 3.2A). A similar effect was observed for the proportion of target responses ($\mathrm{P_T}$; $F_{(2,18)} = 7.97$, $\mathrm{MSE} = 0.011$, $p = 0.0033$). Therefore, consistent with previous studies examining memory for direction of motion (Zokaei et al., 2011), both the number of correctly recalled targets and the precision of those responses decreased as the amount of to-be-remembered information increased. We also observed a significant number of nontarget (i.e., "binding") errors that increased at larger set sizes ($\mathrm{P_{NT}}$; $F_{(2,18)} = 7.15$, $\mathrm{MSE} = 0.01$, $p = 0.0052$). Because we wanted behavioral responses to match as closely as possible the decoded patterns of brain activity, participants were probed on the three critical directions on a majority of trials. This procedural limitation could have allowed participants to learn which directions were most likely to be probed, thereby affecting their behavioral responses. To determine whether such learning occurred, we obtained parameter estimates for the first and second half of the experiments, collapsed across load. This analysis revealed no significant changes in either the concentration of target responses ($\mathrm{t}(9) = -1.31$, $\mathrm{SD} = 12.33$, $p = 0.22$) or the likelihood of reporting the target ($\mathrm{t}(9) = -1.12$, $\mathrm{SD} = 0.043$, $p = 0.29$), which is consistent with previous studies that failed to show learning for individual items in tasks of VSTM (Olson and Jiang, 2004; Beck et al., 2008). Moreover, the estimates obtained here are very similar to those obtained in a task that used a similar design, but that selected probes at random (Zokaei et al., 2011), suggesting that participants' responses were unaffected by non-randomzied target probes.

Classification Results

ROI analysis

Before examining whether classification performance changes as a function of VSTM load, we first tested whether it was possible to successfully decode the direction of motion of one of the memory stimuli while multiple directions were held in VSTM. A previous study by our group (Riggall and Postle, 2012) reported that decoding the direction of a single item is possible within those ar-

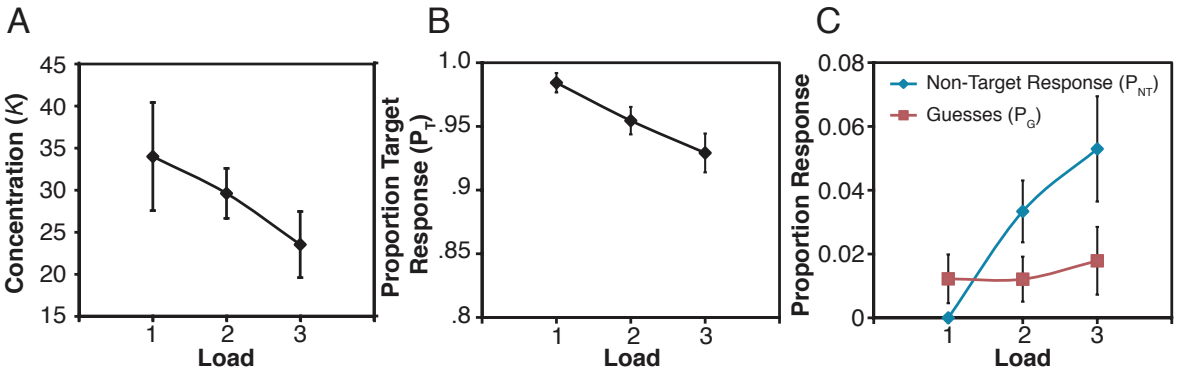

Figure 3.2: Behavioral results derived from maximum likelihood estimates of the response error. (*A*) The concentration parameter, κ, decreased as a function of memory load, indicating a decrease in mnemonic precision. (*B*) The P_T similarly decreased with increasing VSTM load. (*C*) P_{NT} and P_G. Nontarget responses account for the majority of the errors and increased with VSTM load.

eas that support motion perception (i.e., calcarine and extra-calcarine occipital cortex, including MT+) but not in areas that showed sustained activity during the delay (IPS, frontal areas). Accordingly, we focused our initial analysis on nonoverlapping ROIs that demonstrated load-sensitive BOLD signal intensity either in a phasic response to the memory sample ("Sample ROI") or in a response that was sustained throughout the delay period ("Delay ROI"). In this way, we sought to determine whether information about the remembered direction of motion was coded in cortical areas sensitive to visual stimulation independently from those that exhibited load-sensitive, sustained, delay-period BOLD signal.

Examining the patterns of activity in the Sample ROI, classification of the direction of motion resulted in significantly greater-than-chance sensitivity, averaged across all three loads (Fig. 3.4C). This was true both when the classifier was trained on the time point that captured the peak of the sample-evoked response and when it was trained on the final time point of the delay period, a time point that would be dominated by delay-period activity. In contrast, those areas that exhibited sustained, elevated delay-period BOLD signal (i.e., the Delay ROI) showed no evidence of direction sensitivity during any point in the trial (Fig. 3.4D), replicating the results of Riggall and Postle (2012). As a control analysis, training the classifier on intertrial fixation activity resulted in

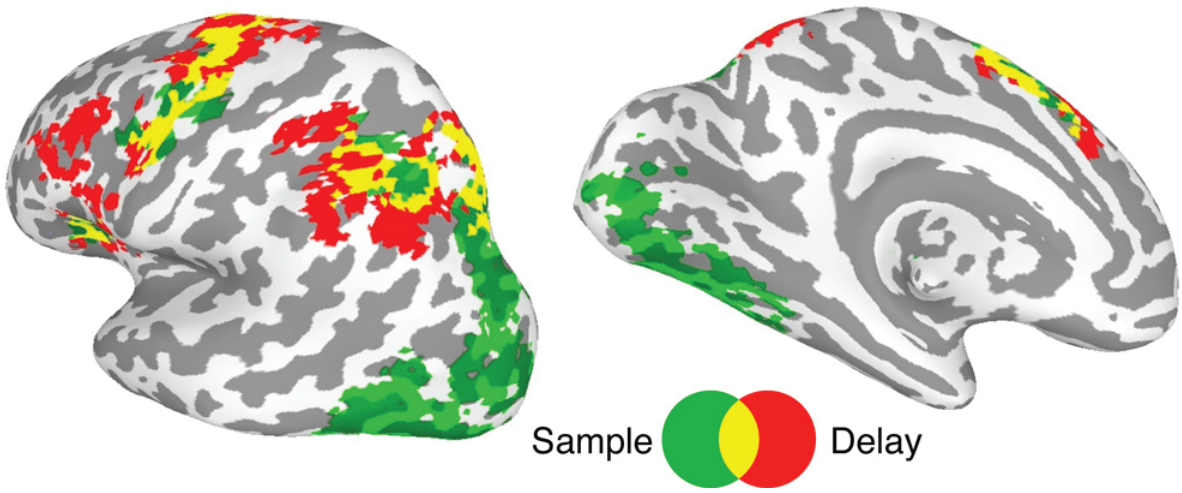

Figure 3.3: ROIs used for classification analysis. Initial analysis was restricted to those areas that showed parametricaly increasing BOLD activity as a function of load that were either transiently evoked by the initial sample of sustained throughout the delay. Subtractions were performed to isolate the Sample (green) and Delay (red) ROIs. Areas showing load sensitivity during both periods (yellow) and those not showing load sensitivity were also included in the subsequent analysis of anatomically defined ROIs. Note that this figure demonstrates group-averaged data, whereas all analyses were performed on single-subject ROIs.

chance-level classification performance for both ROIs.

It is possible that by focusing on areas in which the load-sensitive response was specific to either the sample or delay periods, this analysis overlooked important voxels that are critical to the maintenance of information in VSTM (e.g., those that show elevated levels of BOLD signal intensity to both encoding and maintenance-those that appear yellow in Fig. 3.3). However, further analysis restricted to several anatomically defined areas confirmed that the majority of information about direction of motion during the delay period was exhibited in calcarine and extra-calcarine occipital cortex (Fig. 3.4E). Classification performance was not significantly better than chance in the IPS. Therefore, although elevated delay-period activity was absent from the observed BOLD signal intensity in areas associated with visual perception (Fig. 3.4A), the ability to maintain the representation of multiple items in VSTM appears to be supported by distributed patterns of activity in these areas rather than in those areas that exhibit elevated delay-period activity.

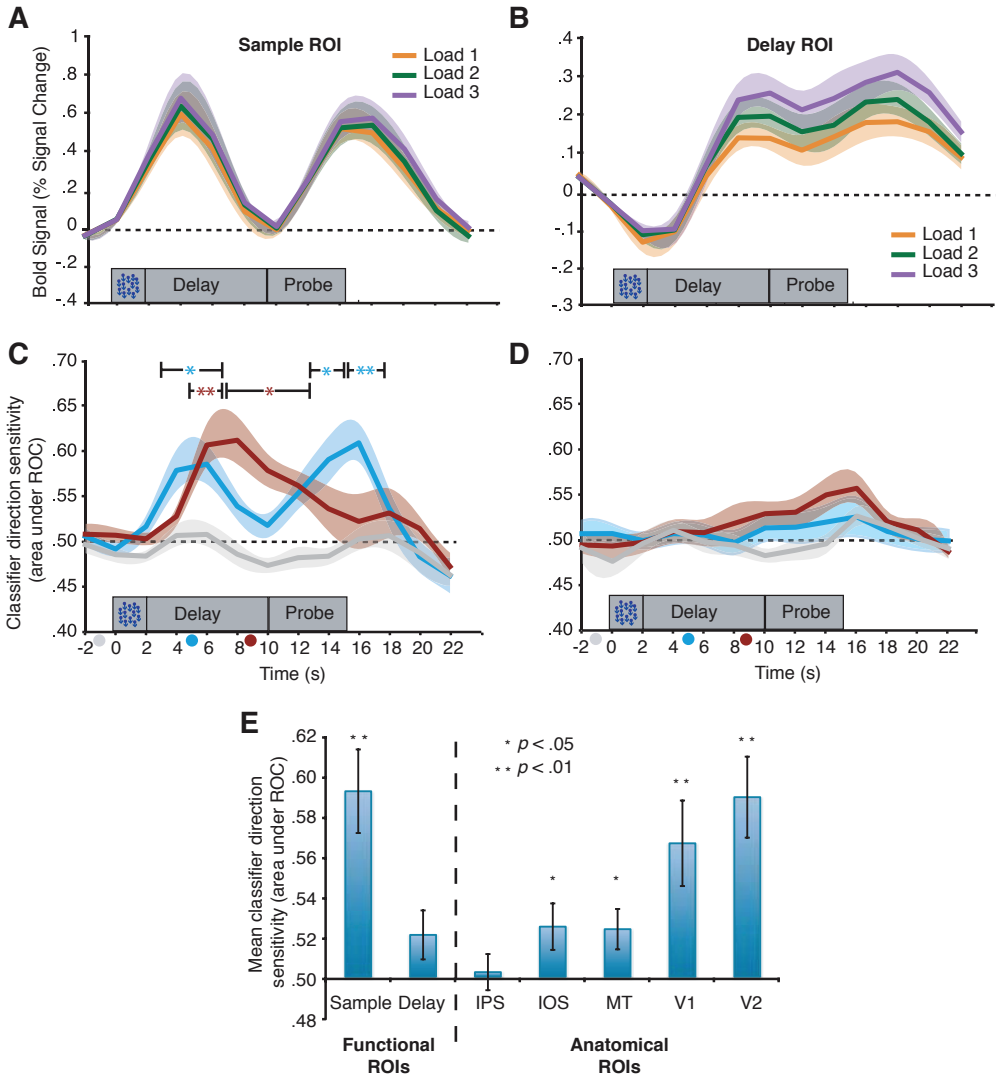

Figure 3.4: BOLD signal intensity and classification performance in functional and anatomical ROIs. (*A*) Diagram showing load-sensitive stimulus-evoked activity, but no sustained delay activity, in the functionally defined, Sample ROI. (*B*) Diagram showing elevated, sustained load-sensitive activity in the functionally defined, Delay ROI. (*C-D*) Classifiers were trained on data using only the TR at -2 s (gray circle), at 4 s (teal circle), or at 8 s (red circle), and then tested across the entire trial, with performance of each indicated by the waveform with the same color. TRs at which a classifier's performance exceeded chance (*p < 0.05, **p < 0.01, one-tailed, corrected) are indicated and color-coded by classification training point. Classification sensitivity is presented for both the Sample (*C*) and Delay (*D*) ROIs. (*E*) Mean classification across the delay period (2-12 s) for functionally defined (Sample; Delay) and anatomically defined (intrococciptial sulxus; V1, V2; MT) ROIs. Significance (one tailed, corrected) for each ROI is indicated. Shaded areas (*A-D*) and error bars (*E*) denote SEM.

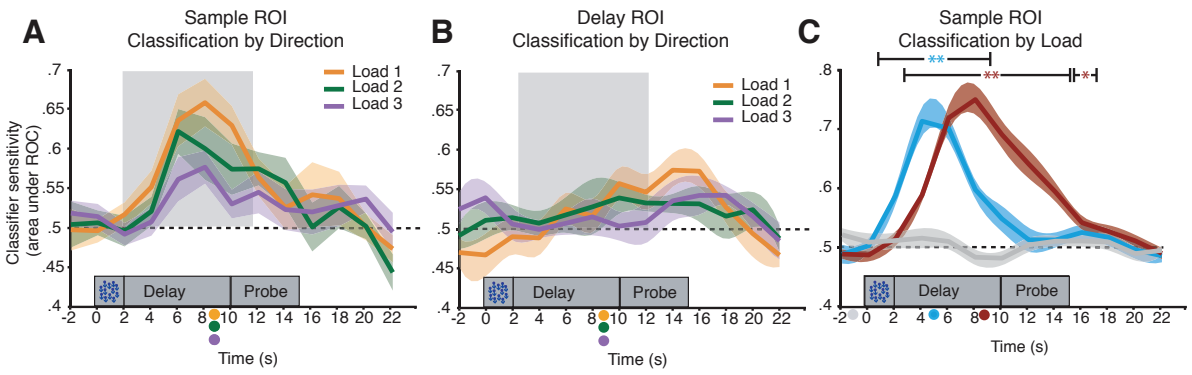

Figure 3.5: Load sensitivity of patterns of neural activity. Classifiers were trained on a single TR (denoted by the colored dots) and tested across the entire trial. (*A*-*B*) Classification sensitivity for direction of motion as a function of load in the functionally defined Sample (*A*) and Delay (*B*) ROIs. Averaging across the entire delay (denoted with gray shading), classification performance for all three loads is significantly greater than chance ($p < 0.01$, one-tailed corrected) in the Sample ROI (*A*), but not in the Delay ROI (*B*). (*C*) Classification sensitivity for load (number of moving dot patterns) in the Sample ROI, collapsed across direction. Significance is tested at individual TRs (*$p < 0.05$, **$p < 0.01$, one-tailed, corrected). Shaded areas denote SEM for each TR.

Load Sensitivity

A critical aim of the present study was to determine whether classification performance would decrease as a function of memory load in a manner that reflected the decline in behavioral precision. To investigate this, we trained a classifier on the final time point of the delay period for both of the Sample and Delay ROIs. Next, for each ROI, we decoded classification performance at each time point from each of the three load conditions, resulting in a classification time course that reflected the extent to which stimulus-specific neural representations were present across the entirety of the trial.

Examining the entire delay period (4-12 s) in the Sample ROI, mean classification sensitivity was significantly greater than chance for all three loads ($p = 0.01$ adjusted; Fig. 3.5A). Therefore, even in the presence of competing within-category representations, memory representations can be identified in the patterns of activity in sensory cortex. Direction sensitivity significantly decreased during the delay period at higher loads ($F_{(2,18)} = 5.63$, MSE $= 0.009$, $p = 0013$), which is consistent with the decline in behavioral performance at higher set sizes. One concern with this

result might be that presenting multiple stimuli in a single location might lead to a loss in classifier sensitivity at higher loads independently of VSTM; that is, by presenting multiple items at one location, the same neurons (within the same voxels) in visual cortex would be required to process the perceptual information of each of the directions of motion, resulting in a loss in sensitivity for any given direction. If this alternative account were correct, one would expect to see the same load-related decline in classification during the stimulus presentation epoch of the trial. However, no such decrease as a function of load was observed when this procedure was repeated with classifiers trained on the time point that captured the peak of the sample-evoked response (i.e., 4-6 s) and testing classification sensitivity during the sample-evoked period (i.e., 4-8 s; $F_{(2,18)} = 2.5$, $\mathrm{MSE} = 0.006$, $p = 0.11$). Therefore, our results are consistent with the interpretation that the load-dependent decrease in classification sensitivity during the delay period was due to a reduction in the proportion of neural resources dedicated to representing that information in VSTM.

In addition, even though delay-period BOLD signal intensity in the Sample ROI was not significantly affected by load (Fig. 3.3; $F_{(2,18)} = 2.64$, $\mathrm{MSE} = 0.0052$, $p = 0.0991$), VSTM load could be successfully decoded from sensory regions (Fig. 3.5C). This finding provides further evidence that the patterns of delay-period activity in areas involved in visual perception reflect the representation of multiple items in VSTM independently of changes in BOLD signal intensity identified in univariate signal intensity. In contrast, despite the presence of load-sensitive BOLD signal in the Delay ROI (Fig. 3.3; $F_{(2,18)} = 5.11$, $\mathrm{MSE} = 0.01386$, $p = 0.0174$), classification performance of the contents of VSTM (i.e., direction of motion) was not significantly above chance at any load ($p > 0.05$; Fig. 3.5B). Therefore, although BOLD activity in these regions (including IPS) is sensitive to VSTM load, we failed to find evidence that these areas are involved in the representation of the contents of VSTM per se. In the Discussion section, we consider other aspects of task performance to which this activity may correspond.

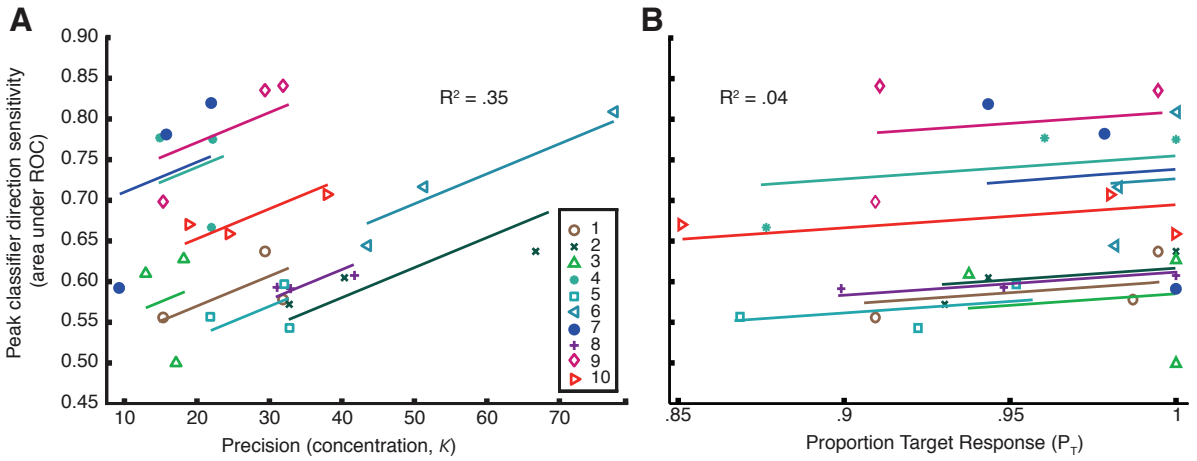

Figure 3.6: The relationship between Sample ROI classifier sensitivity and behavioral performance. (*A-B*) Within subjects, changes in classifier sensitivity across load are significantly correlated with changes in behavioral precision (*A*), but not with changes in the proportion of target responses (*B*). Data are modeled for each subject and fit with parallel lines with ANCOVA using the method of Bland and Altman (1995).

Relating classification performance to response precision

Does the decrease in classifier performance as a function of memory load reflect the associated decline in mnemonic precision? To examine this question, we specifically focused on within-subject changes across load (see Materials and Methods), because there may be sources of error that affect between-subject classification performance that are unrelated to differences in behavior (e.g., differences in encoding strategies, Linke et al., 2011a; Vicente-Grabovetsky et al., 2014, or individual differences in signal-to-noise ration of BOLD signal, Tong et al.,2012) This analysis revealed that, across loads, the observed decrease in peak classifier sensitivity was significantly correlated with an individual's change in mnemonic precision ($r = 0.58$, $p = 0.006$; Fig. 3.6A). This was also true when the analysis was restricted to the individual TRs corresponding most closely to the encoding (TR 4; $r = 0.45$, $p = 0.04$) or delay activity (TR 6; $r = 0.53$, $p = 0.01$).

It is possible, however, that the decrease in classifier performance reflects the decreased probability that the probed (target) direction was stored in memory at higher loads; that is, given that both the proportion of correctly recalled targets, and the precision of those responses, decreased

as a function of load, the decline in classifier sensitivity could reflect the proportionally fewer trials on which the target direction was successfully encoded and maintained. Contrary to this alternative account, however, P_T was uncorrelated with peak classifier sensitivity ($r = 0.21$, $p = 0.35$; Fig. 3.6B). In other words, the change in classifier sensitivity across memory loads cannot be predicted by a decrease in the percentage of trials on which the critical direction of motion was stored in VSTM. Therefore, the decrease in classification performance appears to reflect the decline in mnemonic resolution, rather than decreases in other aspects of behavior or random decreases that are independent of the internal representations that guide behavior (i.e., noise).

3.4 Discussion

In the present study, we sought to investigate how the neural patterns associated with the contents of VSTM change as a function of load. Consistent with several recent studies, we observed that stimulus-specific properties of information maintained in VSTM could be decoded from the patterns of activity in sensory cortex (Ester et al., 2009; Harrison and Tong, 2009; Serences et al., 2009; Riggall and Postle, 2012). We have extended these previous findings by demonstrating that the successful classification of the contents of VSTM is not an all-or-none phenomenon; instead, classification sensitivity varies as a function of VSTM load. Therefore, these findings reveal the presence of load sensitivity in the patterns of activity in sensory cortex, providing strong evidence that these patterns may represent a neural correlate of VSTM maintenance.

We also observed a significant within-subject correlation between classifier sensitivity and mnemonic precision. This finding provides some of the first evidence that variation in mnemonic precision can be extracted from the neural patterns associated with the storage of information in VSTM. A number of recent psychophysical models have established that the quality of VSTM representations decreases as a function of memory load (Palmer, 1990; Wilken and Ma, 2004; Bays and Husain, 2008; Zhang and Luck, 2008; Bays et al., 2009; van den Berg et al., 2012) Although the models that explain these data differ in their stance on whether VSTM performance is limited

primarily by a finite pool of resources or if it also contains a fixed capacity of a limited number of slots, these models all emphasize that the precision of maintained representations varies with load, particularly at low set sizes (i.e., when the number of items is below any putative capacity limits). Therefore, "slots + resource" (Alvarez and Cavanagh, 2004; Buschman et al., 2011) and "slots + averaging" (Zhang and Luck, 2008, 2011; Anderson et al., 2011) hybrid models and pure resource models (Wilken and Ma, 2004; Bays and Husain, 2008; van den Berg et al., 2012) can accommodate the findings reported here, in that all hold that the precision of maintained representations will vary with load at small set sizes. Specifically, our results are consistent with the prediction that, as the amount of to-be-remembered motion information increases, fewer resources will be dedicated toward representing a given item, resulting in less signal present in the patterns of neural activity that reflect this representation. Critically, this relationship between VSTM precision and classification performance provides evidence consistent with the idea that sensory visual cortex may play a critical role in supporting the representation of information in VSTM (Postle, 2006). These findings take the important step of linking the neural patterns associated with content-specific VSTM encoding and maintenance with the internal representations associated with behavior.

One possible concern with these results is that a decrease in decoding accuracy with increasing load, such as we present here, could also result from decoding noise that is unrelated to the mnemonic representation. For example, the neural response to multiple objects has been shown to be similar to the weighted sum (MacEvoy and Epstein, 2009) or average (Reddy et al., 2009) of individual responses. This alternative seems unlikely, however, for a number of reasons. First, each condition contained the same number of stimuli, but varied only in how many of the dot patterns contained motion. Therefore, this control reduces the stimulus differences between conditions that might affect the weighted averaging or summation of neural signal. Second, classification training was performed over all three tested loads. Accordingly, classification performance was based on patterns of activity that included the increased noise present on load 2 and 3 trials. Third, classification performance was unaffected by load when training and testing to the patterns of activity transiently evoked by the sample stimuli, indicating that classification performance did

not decrease obligatorily due to a decrease in the signal-to-noise ratio at higher set sizes. Finally, a recent study examining the response patterns of single neurons in monkeys has shown that the neural information present about an item encoded and stored in VSTM decreases when multiple items are presented within the same hemifield (Buschman et al., 2011), providing evidence that this effect is not idiosyncratic to the classification technique used here. Together, these facts support our conclusion that classification performance can be accounted for by a decrease in the relative amount of information about the target stimulus present in the patterns of activity in sensory cortex.

Although our findings are consistent with studies that have failed to observe evidence for the contents of VSTM in the IPS (Linden et al., 2012; Riggall and Postle, 2012), one previous study did observe above-chance classification in posterior parietal cortex (Christophel et al., 2012). A possible explanation for these discrepant findings may be that the study by Christophel et al. (2012) used artificial stimuli that required significant training to differentiate. Therefore, decoding accuracy in posterior parietal cortex in this task may reflect processes that are not specific to VSTM maintenance per se. In addition, although the use of the data-driven MVPA approach in the current study has its advantages, there are limitations to this methodology. For example, evidence indicates there is a coarse-scale topographic map of orientation preference in human V1 that is strongly correlated with the retinotopic map (Freeman et al., 2011). Therefore, it is possible that multivariate classification methods could exploit these biases, thereby driving up classification sensitivity only in those areas that have strong topographic organization (e.g., V1 and V2). However, other studies have demonstrated that classification performance of orientation-selective activity is predicted by the average BOLD amplitude in a given region (Tong et al., 2012). Therefore, given that we observed greater BOLD amplitude during the delay period in the IPS than in sensory visual cortex, it is unlikely that our findings can be explained by differences in classification sensitivity across different regions. Nevertheless, future studies should examine load-dependent VSTM classification sensitivity on a range of stimuli (e.g., colors, faces) to determine the reliability of this effect across different features and cortical regions.

If the maintenance and precision of VSTM representations is mediated by patterns of activity in sensory cortex, what is the role of sustained, elevated, delay-period activity? Previous studies have provided evidence that delay-period activity may reflect general cognitive or attentional demands of the task (Mitchell and Cusack, 2008; Magen et al., 2009), as well as goal- or action-related information (Curtis and Lee, 2010; Riggall and Postle, 2012), rather than reflecting VSTM maintenance. One possibility is that whereas VSTM representations may be supported by patterns of activity in sensory cortex, sustained, elevated, delay-period activity (particularly in IPS) may reflect the directed attention toward a limited number of representations (Anderson et al., 2011). This mechanism could potentially account for both apparent capacity limits (i.e., those within the focus of attention), as well as a larger capacity for lower-resolution representations (Sligte et al., 2008, 2009, 2010).

The finding that the precision of VSTM representations is predicted by areas that mediate perceptual encoding is largely consistent with the evidence that VSTM performance is likely limited by encoding processes (Buschman et al., 2011; Linke et al., 2011a; Emrich and Ferber, 2012; Mazyar et al., 2012; van den Berg et al., 2012); that is, the continued recruitment of the same populations of neurons that mediate encoding and perception for VSTM will limit the precision of memory representations to that of encoding. In addition, when attention is allocated across two different memory categories and then retroactively cued to a single item, classifier evidence for the uncued item dropped to chance, suggesting that the allocation of resources in VSTM may be dynamic (Lewis-Peacock and Postle, 2012; LaRocque et al., 2013). Therefore, while it is possible that the fidelity of VSTM may be further affected by subsequent processes, such as by fluctuations in attention (van den Berg et al., 2012) or stochastic degradation (Fougnie et al., 2012), the correlation between classifier performance and VSTM precision during early stages of encoding rule out the possibility that all variation in precision reflects limitations in later stages of processing (Awh et al., 2007; Barton et al., 2009, e.g., during recall or comparison).

At the neural level, the successful classification of multiple items in VSTM from sensory cortex, as well as its relationship with mnemonic precision, further calls into question the widely held

assumption that VWM maintenance of multiple items occurs in a network of frontoparietal regions (Courtney et al., 1997; Linden et al., 2003; Todd and Marois, 2004; Vogel and Machizawa, 2004; Xu and Chun, 2006; Reinhart et al., 2012). In particular, it calls into question the role of elevated, sustained, load-sensitive delay-period activity in IPS as a hallmark of VSTM maintenance (but see Christophel et al., 2012). Instead, our findings support the view that VSTM maintenance is accomplished by actively maintaining representations in "lower level" sensory cortex (Postle, 2006; Ester et al., 2009; Harrison and Tong, 2009; Serences et al., 2009; Linke et al., 2011b; Riggall and Postle, 2012).

4

The representation of transparent motion during visual short-term memory

A number of recent results have begun to question long-standing beliefs in how information is represented during short-term memory (STM) and working memory (WM). Using MVPA these studies show evidence for stimulus representation within posterior visual regions during visual STM, providing strong support for the sensory recruitment hypothesis wherein the same neural systems used for sensory processing are reused for storage. These same studies typically fail to evidence for delay-period stimulus representation in regions whose BOLD activity is nonetheless elevated during the delay period (including the IPS and dorsolateral prefrontal cortex (dlPFC)). In the present study we extended this work by examining the representation and precision of multiple items in memory for transparent motion stimuli. We tested the ability of participants to remember the direction of motion for multiple (1, 2, or 3) overlapping, coherently-moving random dot surfaces. After a delay, one of the moving surfaces was redisplayed and participants reported in which direction it had been rotated relative to its initial direction. Participants showed decreased precision with increased memory load. In line with recent findings, stimulus-specific information was only decodable from posterior visual regions, not frontal and parietal regions that nonetheless showed elevated delay-period BOLD signal. This information was only recoverable from load 1 and 2 trials, suggestive of an encoding limitation. Finally, cross-temporal decoding showed that the representations were not stable over the delay period, in line with a recoding process during the trial.

Adam C. Riggall & Bradley R. Postle

4.1 Introduction

The short-term storage of information no longer present in the environment is critical for many higher-level cognitive functions. Consequently, understanding the neural bases of such storage is of paramount importance in better understanding complex behavior. Recent work by several labs applying multi-voxel pattern analysis (MVPA) to human functional magnetic resonance imaging (fMRI) data has shown evidence for the short-term retention of orientation (Harrison and Tong, 2009; Serences et al., 2009; Ester et al., 2009; Albers et al., 2013), motion (Riggall and Postle, 2012; Emrich, Riggall et al., 2013), more complex stimuli (Christophel et al., 2012; Christophel and Haynes, 2014; Vicente-Grabovetsky et al., 2014), and visual objects (Linden et al., 2012) within posterior visual regions. Taken together, these data provide strong support for the *sensory recruitment hypothesis*, which posits that short-term storage of information is accomplished using the same neural systems used for sensory processing, rather than specialized storage buffers (Awh and Jonides, 2001; Pasternak and Greenlee, 2005; Jonides et al., 2005; Postle, 2006; D'Esposito, 2007; Serences et al., 2009; Sreenivasan et al., 2014a).

Within this framework, a critical question remains: how are multiple items retained in short-term memory (STM)? Previous work using univariate methods to isolate regions that showed an increase in activity with memory load has suggested intraparietal sulcus (IPS) as a probable location for such storage (Linden et al., 2003; Todd and Marois, 2004; Xu and Chun, 2006; Kawasaki et al., 2008; Cusack et al., 2010). Recent work by our group has challenged this conclusion, at least for STM for the direction of motion, showing the successful decoding from posterior occipital and temporal visual regions of individual motion directions during short-term memory for multiple sequentially presented directions and not from parietal or frontal regions (Riggall and Postle, 2012; Emrich, Riggall et al., 2013).

The present study follows up on these results by examining the neural representation of multiple, simultaneously presented motion directions using transparent motion. Transparent motion provides an extreme test for the storage system, as spatial and temporal individuation cues are eliminated. Behavioral evidence suggests that, at least for small loads, individual transparent mo-

tion directions are maintained as separate representations (Braddick et al., 2002; Zokaei et al., 2011). Accordingly, we sought to find neural evidence for the memory representation of individual motion directions from transparent motion stimuli using MVPA.

We also aimed to follow up on another intriguing disparity seen in our previous work. With single item decoding of motion directions, using temporal cross-generalization (King and Dehaene, 2014), we observed stable decoding performance from classifiers trained at different time points during the trial, suggesting a stable memory representation throughout the trial (Riggall and Postle, 2012). In our multiple item version, the results were suggestive of a temporally dynamic representation which transformed over the course of the delay period from the sample stimulus-evoked code to a completely different code by the time of the memory probe onset. However, a relatively short delay period in the Emrich, Riggall et al. (2013) study made this challenging to interpret. Such a result would fit well with recent work showing temporally dynamic representations (Stokes et al., 2013; Sreenivasan et al., 2014b) during short-term memory. To that end, we increased the delay length of our task considerably to allow us to examine the temporal dynamics of the representations throughout the trial.

4.2 Materials and Methods

Participants

Seven volunteers (3 females) between 18 and 30 years of age (mean 24.6 years) were recruited from the University of Wisconsin-Madison community and paid for their participation. All subjects had normal or corrected-to-normal vision, no reported history of neurological disease, and no other contraindications for MRI. All subjects gave written informed consent according to the procedures approved by the Health Sciences Institutional Review Board at the University of Wisconsin-Madison.

Behavioral Paradigm

Participants performed nine runs of a delayed-recognition task for directions of transparent visual motion (Fig. 4.1) while being scanned (fMRI). Each trial began with a fixation cross changing color from gray to white, indicating to subjects that they needed to fixate the cross and prepare for the start of the trial. After 2 s a circular aperture containing moving dots centered on the fixation point was presented (sample, 2 s). This aperture contained three sets of dots (each a different color: orange, purple, and blue). On a given trial, one, two, or all three of these sets of dots could be moving, while the remaining sets were stationary. When more than one set of dots moved, this created the perception of transparent surfaces moving in different directions on top of each other. Participants were instructed to remember the direction of motion for any moving dot sets while maintaining fixation. The sample was followed by a 12 s delay-period. Following the delay, one of the sets of coherently moving dots (probe, 1 s) was redisplayed and subjects were required to indicate as quickly as possible with one of two buttons if the direction of motion in the probe was rotated clockwise or counterclockwise relative the direction of motion of the same colored dots in the sample. Following the probe period the fixation cross changed color (feedback, 1 s) to indicate if the response was correct (green) or incorrect (red). A variable intertrial interval (ITI) followed (4, 6, or 8 s, equally likely), during which the fixation cross changed color to gray and participants were instructed to wait for the next trial. On ~15% of trials, immediately after the sample presentation the fixation cross turned pink (catch, 14s), indicating to subjects that they did not need to remember the sample items, the trial was over and they would never be probed, and they should simply wait for the next trial while maintaining fixation.

Before scanning, subjects were trained on the clockwise/counterclockwise response, introduced to transparent motion stimuli containing two and three sets of coherently moving dots, and practiced three blocks of 21 trials to familiarize themselves with the experimental procedure

All stimuli were presented on a black background. A fixation cross (~0.13°) was present on the center of the screen at all times. Sample and probe stimuli consisted of circular patches (11° diameter) of dots (~0.24°) centered on the screen. The sample was composed of three independent sets

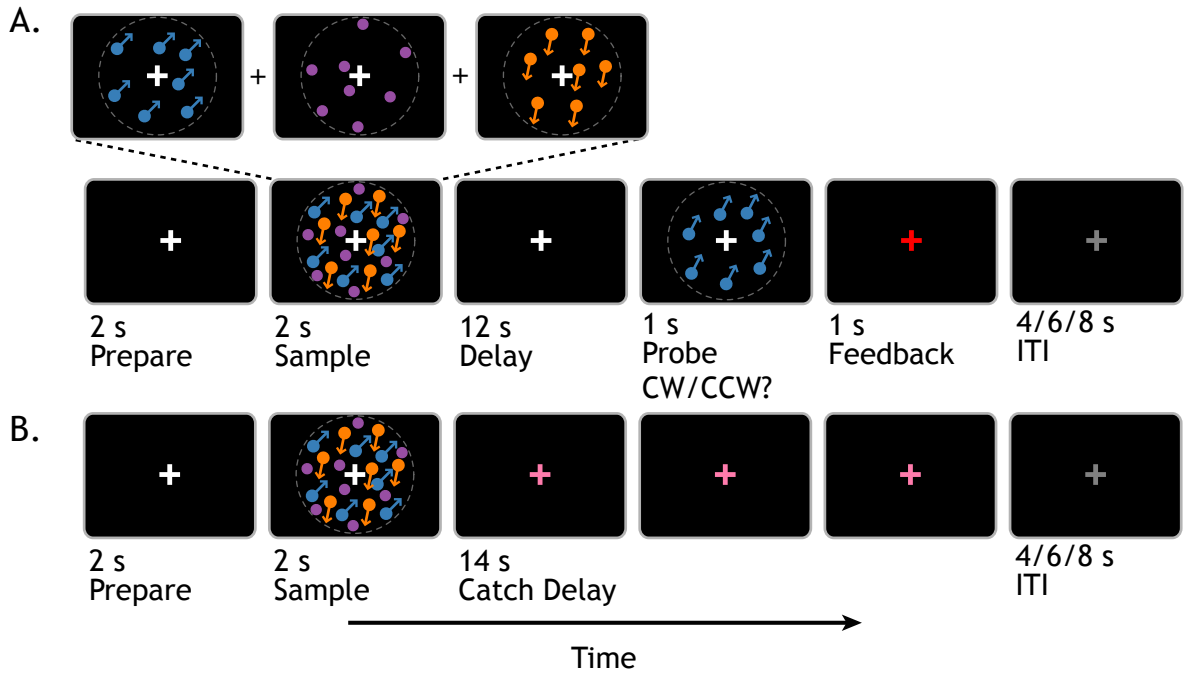

Figure 4.1: Behavioral task. (*A*) Participants maintained the direction of motion of 1, 2, or 3 sets of coherently moving dot stimuli presented transparently over a long delay period. At the end of the delay period they were presented with a probe motion stimulus and had to indicate with a button press whether the direction of motion was rotated clockwise or counterclockwise from the remember motion direction of the same color. (*B*) On catch trials participants were instructed to maintain fixation while waiting for the next trial.

of dots (individuated by color: orange, purple, or blue) which could each be stationary or moving coherently at 8°/s, with an overall density of ~0.7 dots-per-square degree. The probe consisted of one of the colored sets of dots (which had been moving during the probe) being represented, with the motion direction rotated clockwise or counterclockwise at one of three specific magnitudes of change (7°, 22°, or 45°).

To facilitate a decoding analysis, ~85% of trials contained one of three specific target directions (71°, 191°, 311°). On load 2 and 3 trials, the other directions were randomly selected, constrained so that each would be at least 50° and no more than 165° from the other directions in a trial, to insure the perception of transparent surfaces (Braddick et al., 2002; Felisberti and Zanker, 2005) and to prevent directions from being directly opposite one and other. On the remaining ~15% of

trials all directions were randomly sampled, constrained as before to not be too close or directly opposite. The probed item could be any of the presented directions, leading to each of the critical target directions being probed on ~28% of load 1 trials, on ~14% of load 2 trials, and on ~9% of load 3 trials. This distribution was designed to lessen the likelihood of participants learning the specific target directions, as the majority of probed items were not of the specific target directions.

Participants completed 189 trials over the course of 9 runs while in the scanner: 78 load 1 trials (24 each containing one of the specific target directions, 6 random), 42 load 2 trials (12 each of the target directions, 6 random), 42 load 3 trials (12 each of the target directions, 6 random), and 9 catch trials of each load. Each of the possible probe change magnitudes (-45°, -22°, -7°, 7°, 22°, or 45°) occurred 13 times on load 1 trials, 7 time on each on load 2 and 3 trials. The color of the moving dot sets and probed item were pseudorandomly selected, counterbalanced across all stimulus presentations.

Experimental stimuli were controlled by the Psychophysics Toolbox (http://psychtoolbox.org; Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) running in MATLAB (MathWorks), rear projected onto a screen in the scanner bore near the participants head, at 60 Hz (Silent Vision 6011; Avotec), and viewed through a coil-mounted mirror.

Data Acquisition and Preprocessing

Whole brain images were acquired with the 3T scanner (Discovery MR750; GE Healthcare) at the Lane Neuroimaging Laboratory at the University of Wisconsin-Madison. A gradient-echo, echo-planar sequence (2 s TR, 25 ms TE) was used to acquire data sensitive to the blood-oxygen level-dependent (BOLD) signal within a 64 x 64 matrix (39 sagittal slices, 3.5 mm isotropic). Nine runs of the delayed-recognition task were obtained for each subject, each lasting 8 min, 44 sec (262 volumes). The first 6 seconds of all task runs were discarded to insure a steady state of tissue magnetization. Additionally, a high-resolution T_1-weighted image was acquired for all subjects with a fast spoiled gradient-recalled-echo (FSPGR) sequence (8.132 ms TR, 3.18 ms TE, 12° flip angle,156 axial slices, 256 x 256 in-plane, 1.0 mm isotropic).

The functional data were preprocessed using the Analysis of Functional NeuroImages (AFNI) software package (Cox, 1996). All volumes were spatially aligned to the final volume of the final run using a rigid-body realignment and corrected for slice time acquisition. Temporal trends were removed from each run using 0-5th order Legendre polynomials to reduce the influence of scanner drift. For univariate analyses, data were spatially smoothed with a 6 mm FWHM Gaussian kernel, converted to percent signal change and transformed into Talairach space (Talairach and Tournoux, 1988). For classification analyses, data were z-scored separately within run for each voxel. Data were not smoothed and were left in their native space.

Behavioral Analysis

For each memory load a probit regression model was used to estimate the parameters of the cumulative Gaussian distribution that best fit the relationship between the participants responses and the magnitude of the direction change in the probe. The mean (μ) of the fitted Gaussian provides a measure of response bias, while the reciprocal of the standard deviation ($1/\sigma$) provides a measure of the precision of the memory representation.

Univariate Analyses

While our main analyses focused on multivariate classification, we used a univariate general linear model (GLM) to identify load-sensitive delay-period activity. The model included regressors for each within-trial event of the task (i.e., sample, delay, probe, response, and feedback). Sample and probe were modeled as load-dependent 2 s and 1 s parametric boxcars, respectively. The delay was modeled with a 1-sec load-dependent parametric boxcar located in the middle of the delay period (Zarahn et al., 1999; Postle et al., 2000) to minimize contamination with the sample and probe. The response and feedback were modelled as impulses. The model also contained separate parametric sample and delay regressors for catch trials. All were convolved with a canonical hemodynamic response function (HRF). Each of these independent regressors was entered into a modified GLM for analysis using AFNI.

To aid in visualizing the BOLD response within our regions of interest (ROIs), trial timeseries for each load (1, 2, and 3) and trial type (full, catch) were deconvolved using a finite impulse response (FIR) model (Glover, 1999; Serences, 2004). Each trial type and load was modeled with a separate set of finite impulse response functions covering the entire duration of the trial (14 TRs). Additional regressors for head motion and scanner drift were also included in the model.

Pattern Classification Analyses

Classification was performed using the Princeton Multi-Voxel Pattern Analysis toolbox (http://www.pni.princeton.edu/mvpa) and custom routines in MATLAB (MathWorks). Preprocessed fMRI data from individual trial time points were used to train separate classifiers to classify the direction of motion (3 possible directions) in the sample and by inference in memory (stimulus-specific classification), or to classify the memory load of the trial (trial-dimension classification). For the direction-specific classification the classifier was initially trained on data from only load 1 trials, to insure a clean representation of the specific pattern of activity underlying the representations of each of the specific target motion directions. The trained classifier was then tested on load 2 and 3 trials to assess performance on trials which contained one of the three specific target directions. The load-specific classifier was trained on all data from full trials, collapsing across motion directions.

Classification was accomplished using L2-regularized logistic regression with a λ penalty term of 25. During decoding, a trial was considered correctly classified if the correct direction had the highest likelihood estimate (winner-take-all classification). Overall classification accuracy was determined using leave-one-trial-out cross-validation, in which the classifier was repeatedly trained on data from all but one trial, and then tested on the left out trial, rotating through all the trials as the left out testing trial.

To examine the dynamics of the memory representation, each classifier was trained using data from only a single time point in the trial (e.g., the first volume acquisition following the sample presentation) and then tested on all time points in the left-out trial (i.e., including both time points

preceding and following the training time point). The result of this temporal cross-generalization procedure is a time course of decoding accuracy for the entire trial. By doing the initial training of the classifier using different time points in the trial (e.g., a time point just after the sample, a time point in the later part of the delay, etc.), it was possible to examine the stability of a given representation throughout the duration of the trial King and Dehaene (2014).

Given our hypotheses and previous results (Riggall and Postle, 2012; Emrich, Riggall et al., 2013), we again focused the decoding analyses on ROIs that showed either a parametric effect for load during the sample presentation or the delay period. ROIs were created from the intersection of anatomically defined ROIs and voxels that showed a parametric effect with load in the univariate GLM. Four broad anatomically defined ROIs (lateral prefrontal, parietal, lateral occipital and medial occipital) were created by extracting region masks from the Tailarach atlas template included with AFNI (as originally labeled by Lancaster et al., 2000) and warping them to the subjects native space. The lateral prefrontal ROI consisted of the inferior frontal gyrus mask, the middle frontal gyrus mask, and the precentral gyrus mask. The parietal region consisted of the posterior cingulate, inferior parietal lobule, precuneus, and superior parietal lobule masks. The lateral occipital region consisted of masks for the fusiform gyrus, inferior occipital gyrus, inferior temporal gyrus, middle occipital gyrus, and superior occipital gyrus. The medial occipital ROI consisted of the lingual gyrus and cuneus masks. To create the final ROI used for analysis, those voxels that showed a parametric effect during either the sample or the delay period were identified within each region. This approach was used to insure that we did not artificially exclude any voxels that might contribute to decoding performance.

The significance of classifier performance was determined using a random permutation test (Golland and Fischl, 2003) to determine the likelihood of observing a specific accuracy under the null hypothesis that there is no relationship between the data and the specific class labels used to train the classifier. A null distribution was generated by repeating the entire classification cross-validation procedure 1000 times, randomly shuffling the class labels each time. A p-value was then computed by determining the proportion of permuted accuracies that were higher than the

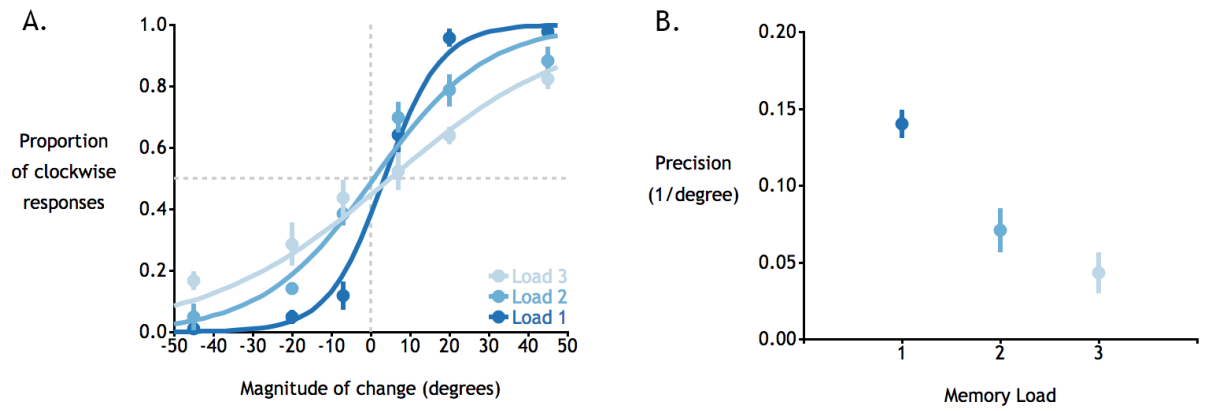

Figure 4.2: Behavioral results. (*A*) Mean proportion of probes judged clockwise compared to the remembered item, as a function of the actual probe rotation. Clockwise probe rotations are shown as positive magnitude changes. Error bars indicate standard errors across participants. Curves show the cumulative Gaussian fits to the group data. (*B*) Precision of the responses at each load, as determined by the reciprocal of the standard deviation of the fitted cumulative Gaussian for each participant. Error bars indicate ± 1 standard error.

observed accuracy. This procedure was repeated for all classification results.

4.3 Results

Behavioral Results

With a single motion direction, participants were quite accurate at indicating whether the probe stimulus was rotated clockwise or counterclockwise relative to the remembered direction, even for small changes in direction (Fig. 4.2A.) Performance decreased as additional transparent motion directions were added. Behavioral responses at each of the three loads were fit with a cumulative Gaussian function. The mean of the fits for all three loads did not differ from 0, indicating no systematic bias in participants' responses across loads (load 1: $t = -1.13$, $p = 0.30$, load 2: $t = 0.65$, $p = 0.54$, load 3: $t = -1.61$, $p = 0.16$). The precision of responses (Fig. 4.2B), as measured by the reciprocal of the standard deviation of the fits, dropped as the load increased. Reaction time also increased with load and with smaller direction changes during the probe, indicating that

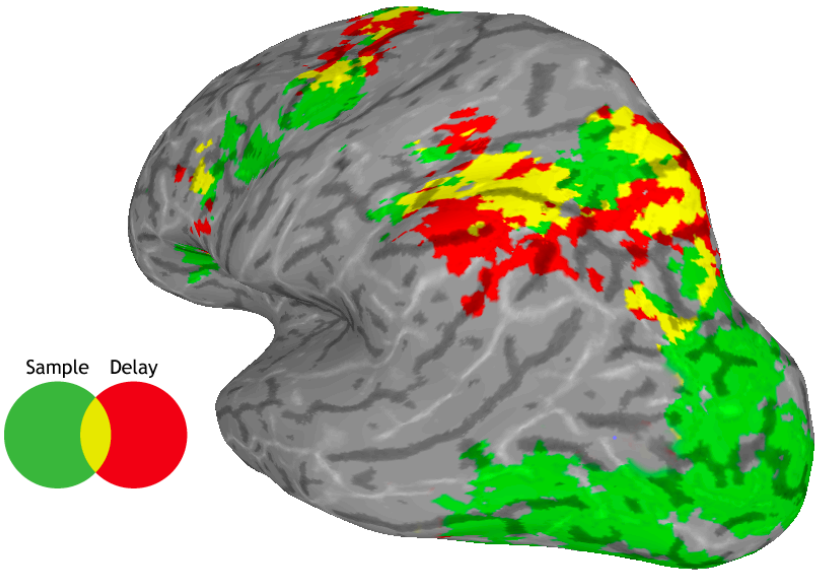

Figure 4.3: ROIs derived from univariate results. Areas in green are regions that showed an increase in BOLD response according to load during the sample period. Areas in red are those that showed a parametric load effect during the delay period. Areas in yellow scaled with load for both the sample and the delay. Note that data are shown as a group-map, but all analyses were conducted at the individual subject level.

participants did not trade speed for accuracy on higher load and smaller change trials.

Univariate Results

A univariate GLM was used to identify regions of the brain that showed a parametric increase in BOLD activity either during the sample period or during the delay (Fig. 4.3). Load sensitive sample regions included medial and lateral occipital cortex, IPS and portions of lateral frontal cortex. Delay-period load-sensitive regions were limited to IPS and prefrontal cortex (PFC), where they overlapped considerably with load-sensitive sample regions.

ROIs were created by intersecting anatomical masks with the sample parametric and delay-period parametric maps. The BOLD timeseries for these regions can be seen in Figure 4.4. There is a strong sample and probe evoked response in the medial and lateral occipital regions (Fig. 4.4A-B), but no load-sensitivity (or elevated activity) during the delay-period. In both the parietal and

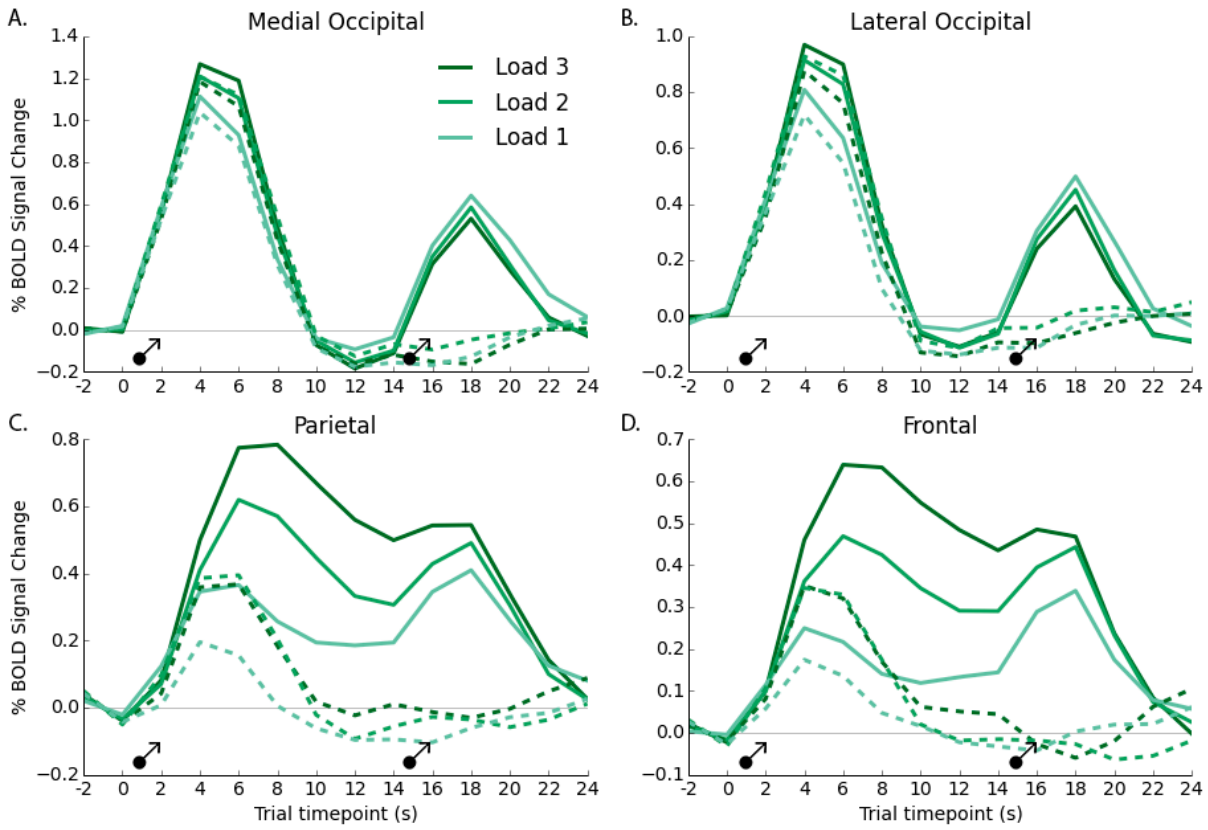

Figure 4.4: Trial-averaged time series data derived from a deconvolution analysis. (*A*) Medial occipital, (*B*) lateral occipital, (*C*) parietal, and (*D*) prefrontal ROI created from the intersection of anatomical ROIs and those regions showing load-sensitive sample or delay activity in the univariate GLM. Solid lines show data from the full trials, dashed lines are from catch trials.

frontal ROIs (Fig. 4.4C-D) there is strong load-sensitive activity throughout the trial. In all regions a similar evoked response is seen during the sample on catch trials, but this quickly returns to baseline.

ROI-Constrained Classification

Given the results from our previous work decoding remembered motion directions (Riggall and Postle, 2012; Emrich, Riggall et al., 2013), we again focused on ROIs that captured initial load-dependent processing during the sample or during the delay-period. In contrast to our previous work in which we chose regions that showed only sample-load-sensitivity or delay-load-sensitivity,

we included regions that showed load-sensitivity during either time-period, to ensure no voxels that might be involved in stimulus representation were excluded.

We began by investigating trial-dimension decoding by testing how well we could decode the memory load on a trial by trial basis. Classifiers were trained to label the specific memory load (1, 2, or 3) of specific trials based on the data from single timepoints during the trial. These classifiers were then tested on all timepoints in the trial to produce a temporal-cross-generalization decoding timeseries. In all four ROIs, classifiers trained on data from shortly after the presentation of transparent motion were successfully able to decode the load on a trial-by-trial basis at an above chance level (chance=33%, $p < 0.05$, permutation test, Fig. 4.5, greenish lines). This decoding performance quickly dropped off in all regions, falling back to chance within the medial occipital and lateral occipital regions within 6 seconds. Classifiers trained on data from later in the trial, during the delay period, also showed reliable above chance performance, but with a different temporal evolution (Fig. 4.5, reddish lines). Decoding only started to be above chance 4-6 later than the peak of the earlier-timepoint-trained classifiers, but then remained stably above chance throughout the delay-period, dropping back to chance levels abruptly with the onset of the probe. Importanly this was true in all ROIs, including the medial and lateral occipital regions, which showed no evidence for delay-period load-sensitive activity in the univariate GLM. Finally, testing the trained classifiers on data from catch trials showed the expected pattern: the early timepoint classifiers were able to decode motion direction in the sample, but quickly dropped to chance; the later timepoint classifiers were at chance for all timepoints (Fig. 4.5, dashed-lines).

We next examined the evidence for stimulus-specific information by training classifiers to decode specific directions of motion (one of three directions). As with the load-specific classification, we again used a temporal cross-generalization procedure, training classifiers on individual timepoints in the trial and testing on all other timepoints. We trained the classifier using data from only load 1 trials to allow the classifier to learn optimal representations for the directions individually (Fig. 4.6, solid lines). Decoding performance in the medial and lateral occipital ROIs showed a similar pattern to the load-decoding, with classifiers trained early in the delay showing

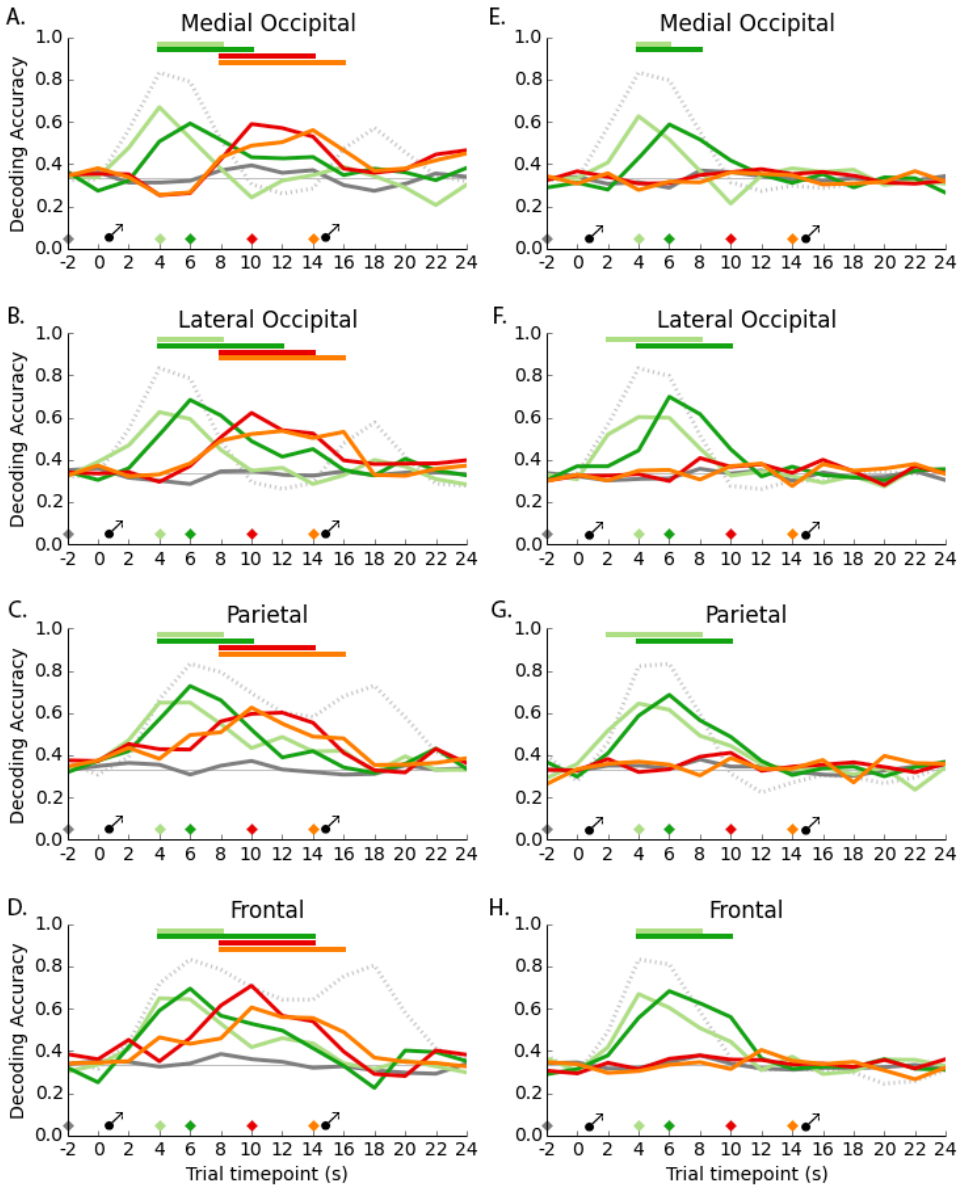

Figure 4.5: Trial-dimension decoding. Temporal cross-generalization timeseries for classifiers trained to decode trial load from data at various timepoints during the trial for different ROI on (*A-D*) full trials and (*E-H*) catch trials. Each waveform represents the mean load-decoding accuracy ($n = 7$) from a classifier trained at a single timepoint in the trial, indicated by the diamond of the identical color along the horizontal axis. Horizontal bars along the tip indicate points at which the decoding accuracy for the corresponding classifier was significantly above chance (0.33, $p < 0.05$, permutation test). Schematic icons of sample and probe presentation are shown at the appropriate times. Data are unshifted in time. Light gray dotted lines show the shape of the average BOLD percent signal change (collapsed across loads) within the ROI, in arbitrary units scaled so that no change occurs at the chance line.

above chance classification for only the sample period, then quickly dropping back to chance levels. This was particularly true for the medial occipital region. Classifiers trained later in the delay again showed a delay in the onset of above chance decoding, relative to the earlier classifiers, but sustained above chance decoding throughout the delay, dropping back to chance at the onset of the probe. Decoding was never different from chance for any of the classifiers in the parietal and frontal ROIs.

Next we applied the classifiers trained on data from load 1 trials to load 2 and 3 trials, which each contained one of the classifier directions and one or two random directions respectively. As can be seen in Figure 4.6 (dashed lines, load 2; dotted lines, load 3), within the medial and lateral ROIs, direction decoding was above chance on load 2 trials, following a very similar temporal evolution to the load 1 trials. Decoding did not differ from chance on load 3 trials.

4.4 Discussion

In the present study we investigated the neural representations underlying the short-term retention of visual motion information. Participants performed a short-term memory task for one, two, or three directions of motion presented transparently in a single circular aperture while in the scanner (fMRI). Using an MVPA decoding approach, we showed the successful decoding of stimulus identity from posterior visual regions, but not from parietal and frontal regions (which nonetheless showed reliable parametric increases in activity with increases in memory load). This is in agreement with the growing body of literature supporting the *sensory recruitment hypothesis*, which posits that short-term storage is accomplished in the same neural systems recruited for sensory processing, rather than by dedicated storage systems (Awh and Jonides, 2001; Pasternak and Greenlee, 2005; Jonides et al., 2005; Postle, 2006; D'Esposito, 2007; Serences et al., 2009; Sreenivasan et al., 2014a).

We have now shown, across three separate studies (Riggall and Postle, 2012; Emrich, Riggall et al., 2013, the current study), the reliable presence of stimulus-specific patterns of delay-period

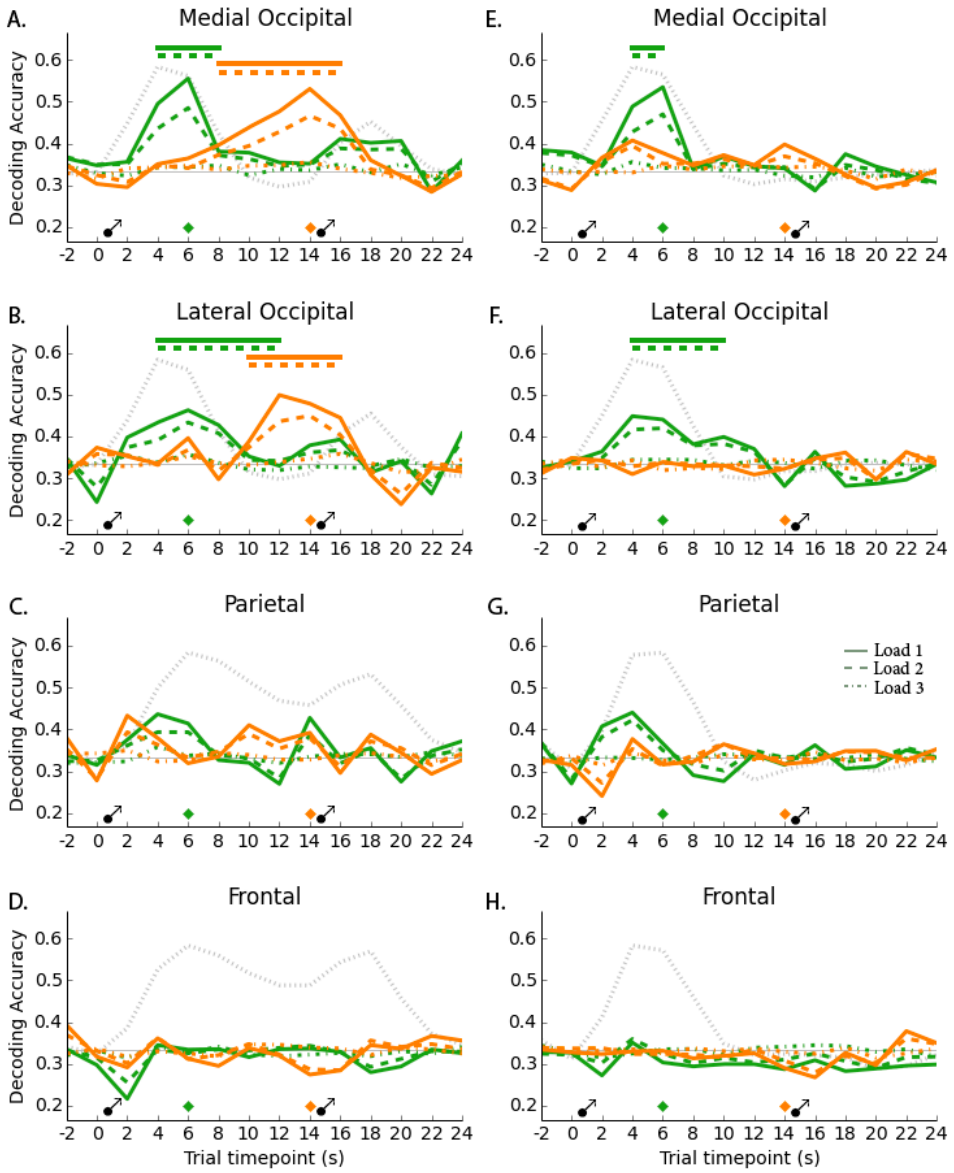

Figure 4.6: Stimulus-dimension decoding. Temporal cross-generalization timeseries for classifiers trained to decode specific directions of motion from data at various timepoints in the trial. Conventions the same as Fig. 4.5, scale on vertical axis changed. Classifiers were trained on data from load 1 trials, then tested on trials of all three loads (load 1: solid line, load 2: dashed line, load 3: dash-dot line; see panel G for legend).

activity in posterior visual regions for individual remembered motion directions, and no evidence for stimulus representation in parietal and frontal regions. These three studies have relied on substantially different task demands: same-different recognition (Riggall and Postle, 2012), memory-to-test displacement (the current study), and direction recall (Emrich, Riggall et al., 2013) as well as different stimulus presentations: single motion direction, multiple directions sequentially, and multiple directions simultaneously, respectively. Despite these differences, the general pattern of results is remarkably similar across the studies. There are, however, important differences as well.

In our original single-item task, we showed stable cross-temporal generalization (Riggall and Postle, 2012). Classifiers trained on data corresponding to the peak of the initial sample presentation were still successful at decoding the remembered stimulus 14 seconds later at the end of the delay period. Likewise, classifiers trained right at the end of delay period were successful at decoding all timepoints throughout the delay. In our first multi-item follow-up (Emrich, Riggall et al., 2013) we observed a slightly different pattern. Classifiers trained on the sample period were successful at decoding during the sample and the probe, but not during the delay period. Classifiers trained during the delay-period were able to decode the remembered direction, but this representation appeared to build more slowly than the initial sensory representation. This was suggestive of a possible recoding of information, but the relatively short delay period (8 s), made it difficult to confirm this. In the present work, with a longer delay period, we see clear evidence for a recoding of information. Classifiers trained during the sample period were able to decode stimulus information, but this ability quickly died out into the delay-period. Classifiers trained on delay-period data, however, showed the opposite pattern, ramping up in decoding performance just as the sample decoding performance was falling. Taken together, these results strongly suggest a recoding of information following the initial sample presentation. Critically, however, this recoded information is still retained in the same regions involved in the initial sensory representation.

Two other differences from our previous work bear mentioning. In both of the previous studies, we observed a relative spike in decoding performance around the presentation of the probe, particularly for classifiers trained on data from the sample period. We do not see anything like

this in the current work. Future work will be necessary to fully investigate the effects of specific task demands on the forms of representation used during responding.

Finally, the complete failure to decode any information, from any timepoints, during load 3 trials is in contrast to our previous finding, where we observed reliable above change decoding at all levels (Emrich, Riggall et al., 2013). One possible explanation could be in the differing methods used to train the classifier. In the previous work, classifiers were trained on data from all three loads simultaneously, then performance on individuals loads was assessed by testing the trained classifiers on individual loads. The current study only ever trained on load 1 trials, then tested for generalization on load 2 and 3 trials. If individual representations of each direction are maintained separately, one would think this method would be more sensitive, because the classifier would be learning the representation without any extra noise. It could be, however, that exposing the classifier to the noise of other directions during training allows for more precise separation of the specific trained directions. We think this is unlikely, given our previously observed correlation between behavioral precision and decoding performance (Emrich, Riggall et al., 2013). Such a correlation would be unlikely if the representation truly was the same across loads. Indeed, Ester et al. (2013) have found a similar correlation between variations in neural representation and precision as assessed with a forward encoding model within a single-item memory task.

An alternative explanation is that, given the complexity of the transparent motion stimulus, encoding difficulties might explain the failure to decode at level three. Three simultaneous transparent motion directions is at or exceeds typically behavioral limits for perception (Felisberti and Zanker, 2005; Suzuki and Watanabe, 2009) and they cannot benefit from attentional reorienting as items that are presented sequentially can be (Ihssen et al., In Press). Thus, it seems quite reasonable that encoding limits, not memory limits per se, were the issue (Linke et al., 2011a).

5

The spatial and temporal distribution of memory representations for visual motion

We recently showed successful timepoint-by-timepoint decoding of remembered directions of motion during a short-term memory task (Riggall & Postle, 2012). Critically, this decoding was only possible within posterior visual ROIs, despite the presence of elevated delay-period signal in frontal and parietal cortex. However, when we applied the same decoding to wholebrain data and inspected the importance maps, we found informative voxels in frontal and parietal regions. Also, despite stable decoding throughout the delay-period, the importance maps suggested that the informativeness of voxels varied greatly throughout the delay. In the present work we investigated these paradoxical observations by examining the spatial distribution and temporal stability of the informativeness of individual voxels. We used two methods to investigate spatial stability: recursive feature eliminination (RFE) and noise-perturbation sensitivity (NPS). Both methods showed the representation is only within posterior visual regions and does not extend to frontal and parietal regions. We used three approaches to examine the temporal stability of voxels' informativeness: comparing the overlap of informative voxels from classifiers trained on individual timepoints, training with multiple timepoints during the delay, and a spatiotemporal approach whereby the entire delay-period of each trial was included as a single exemplar. The results from these analyses showed that while many of voxels were stable throughout the delay period, a large number also changed informativeness throughout the delay, particularly from the first to the second half. Taken together these results support a broadly distributed memory representation across striate and extrastriate cortex that contains both stable and dynamic elements.

Adam C. Riggall & Bradley R. Postle

5.1 Introduction

Since the publication of the original Haxby et al. (2001) paper, multi-voxel pattern analysis (MVPA) has taken the neuroimaging field by storm. The notion that this approach provides an information-based (Kriegeskorte et al., 2006), rather than activation-based, level of analysis for functional magnetic resonance imaging (fMRI) data is intuitively desirable for most research questions. Why settle for making educated-guesses based on reverse inference, with all the pitfalls that accompany it (Poldrack, 2006), when one can directly assess the information content within a given brain area, the thinking goes. So powerful and persuasive has this idea been that no fewer than 8 reviews of the approach have been published within the last decade (Haynes and Rees, 2006; Norman et al., 2006; Mur et al., 2008; Naselaris et al., 2011; Lemm et al., 2011; Tong and Pratte, 2012; Yang et al., 2012; Mahmoudi et al., 2012). This work has lead to a number of interesting new insights into decades old questions that have been revealed through the development of tasks designed to leverage the power of MVPA.

One place where this has been particularly fruitful is in the study of the neural underpinnings of short-term memory (STM). For decades, a prevailing view within this field (or perhaps, even more strongly, outside the direct field) has been that STM storage occurs through sustained, elevated neural activity. Such activity, first identified in the monkey over 40 years ago (Fuster and Alexander, 1971; Niki, 1974a), has been observed in numerous areas of the human brain, most typically frontoparietal regions, with function magnetic resonance imaging (fMRI; Curtis and D'Esposito, 2003). However, the application of MVPA to fMRI during the delay-period of various STM tasks has shown evidence for the short-term retention of orientation (Harrison and Tong, 2009; Serences et al., 2009; Ester et al., 2009; Albers et al., 2013), motion (Riggall and Postle, 2012; Emrich, Riggall et al., 2013), complex visual stimuli (Christophel et al., 2012; Christophel and Haynes, 2014; Vicente-Grabovetsky et al., 2014), and objects (Linden et al., 2012) within posterior visual regions which do not show elevated, sustained activity during the memory delay.

With the success these methods have had in demonstrating the presence of specific task-relevant information during STM, important questions have followed: What information is actually present?

Why is a particular classification approach successful with a given dataset? How can we use the information generated during an MVPA analysis to inform our understanding of the neural system underlying them? Unfortunately, these have not been easy questions to answer. The wide variety of preprocessing steps, the large number of different information-extraction approaches (correlation, classification, regression, etc.), and the varied cognitive domains and brain areas investigated have made it difficult to reach anything resembling a consensus.

The following report details our efforts to pursue these questions further with the data from a previously published study on the short-term retention of visual motion information during delayed-recognition (Riggall and Postle, 2012). In that study we reported finding, with MVPA, evidence for stimulus-specific information (i.e., specific directions of motion) during the delay-period of the task, but only when we used regions of interest (ROIs) covering posterior visual regions. We found no evidence for information in the frontoparietal regions that showed robust elevated delay-period activity during the same task.

However, when we applied the same classification methods to whole brain data that had been feature-selected to include any voxels that changed activity at some point during the trial and looked at the specific voxels that the trained classifier was using to make classification decisions (often visualized as an *importance map* or *sensitivity map*), a number of voxels in frontal and parietal regions turned up. Where these regions part of a broadly distributed representation, or spurious noise voxels included by chance?

We also performed a temporal-cross-generalization analyses (for a recent review of the rationale behind this approach see King and Dehaene, 2014), where we trained our classifier on only the data from a single timepoint in the trial, then tested this classifier on all timepoints in the trial to see how well it would generalize, giving a sense of the stability of the representation. We found broad cross-generalization: classifiers trained anywhere from the post-stimulus-onset timepoint to the timepoint immediately preceding the test probe generalized well to other timepoints. However, when we inspected the importance maps on a timepoint-by-timepoint basis, the specific voxel weights used by the trained classifier suggested a much more dynamic representation with

the contribution of any given voxel to classification waxing and waning over the delay period.

Thus, the aim of the present study was to investigate these two paradoxical observations and assess the *spatial distribution* and *temporal stability* of mental representations of visual motion information as indexed by multivoxel patterns of activity. We present empirical results from a reanalysis of the data from our previous study to investigate these two findings.

5.2 Materials and Methods

This study focused on a reanalysis of data previously published (Riggall and Postle, 2012). For full details about the design and methods, please see the earlier publication. Briefly, participants (7) were scanned (fMRI) while performing a delayed-recognition task for visual motion. Participants were shown a single dot motion stimulus per trial and were instructed to remember both the direction and speed of the motion. Halfway through the 15 s delay period they were cued as to whether they would be making a judgment about the direction or the speed of the remembered stimulus for an upcoming comparison. At the end of the delay, a second dot motion stimulus was presented and participants had to indicate whether it matched the remembered stimulus on the previously cued dimension. Participants completed 168 trials over the course of 7 runs in the scanner, with direction cued 96 times, speed the remaining 72. Critically, the direction of the to-be-remembered motion stimulus was always a combination of one of 4 possible directions (presented off the cardinal axes to reduce verbalization/categorization) and 3 speeds. It was these 4 directions that were used as items to be classified in the MVPA analysis. The following discussion will focus only on data from the direction-probed trials.

Data Acquisition and Preprocessing

Over 7 runs, 2527 volumes were acquired at 3T using a gradient-echo, echo-planar sequence (2 s TR, 25 ms TE, 64x64 matrix, 39 sagittal slices, 3.5 mm isotropic) The functional data were preprocessed using the Analysis of Functional NeuroImages (AFNI) software package (Cox, 1996). All

volumes were spatially aligned to the final volume of the final run using a rigid-body realignment and corrected for slice time acquisition. Linear, quadratic and cubic trends were removed from each run to reduce the influence of scanner drift. Data were z-scored separately within run for each voxel. Data were not smoothed and were left in their native space.

To help in identifying the brain regions contributing to classification performance, we created broad anatomically defined ROIs (lateral prefrontal, parietal, lateral occipital and medial occipital) by extracting region masks from the Tailarach atlas template included with AFNI (as originally labeled by Lancaster et al., 2000) and warping them to the subjects native space. The lateral prefrontal ROI consisted of the inferior frontal gyrus mask, the middle frontal gyrus mask, and the precentral gyrus mask. The parietal region consisted of the posterior cingulate, inferior parietal lobule, precuneus, and superior parietal lobule masks. The lateral occipital region consisted of masks for the fusiform gyrus, inferior occipital gyrus, inferior temporal gyrus, middle occipital gyrus, and superior occipital gyrus. The medial occipital ROI consisted of the lingual gyrus and cuneus masks. Classification analyses were then conducted on all voxels within the union of these four regions that showed any task-related change in activity as determined by a main effect of task contrast in a univariate general linear model (GLM). After classification, the importances maps were then assessed in each of the four ROI to look for contributions to classification performance from that region.

MVPA Procedure

Classification was performed using PyMVPA (Hanke et al., 2009) and custom routines written in python. Preprocessed fMRI data were used to train classifiers to classify the direction of motion in the sample, and by inference, the direction of memory in memory (4 possible directions) from trials on which direction had been cued. Classification was performed using L2-regularized logistic regression, a linear classification approach used widely in the machine learning community and well suited for application to fMRI data given it's ability to generalize well after learning in high-dimensional feature spaces with limited training examples (Pereira et al., 2009). The lambda

penalty term, which reduces the contribution of less informative voxels to classification and thus improves generalization, was set to 25, the same value used in our previous report.

As logistic regression is a binary classification approach and we had four separate classes (individual directions), each classification run consisted of training six different classifiers on the six pairwise combinations of classes (e.g., direction 1 vs. direction 2, direction 1 vs. direction 3, etc.) Classification performance on test samples was computed using a majority voting approach: each pairwise classifier was used to predict the class of the test sample and the class which was predicted as more likely the most (i.e., received the most votes) was assigned as the overall predicted class. In cases of a tie, the winner was randomly chosen from the tying options.

To insure that any results we observed were real and not due to biases introduced during the training phase, we used a cross-validation procedure with balanced random sampling of the training and test sets. Seven validation folds were set with fold boundaries corresponding to scanner run boundaries, so that training samples were from 6 of the runs and the testing samples were from the remaining run. Within each fold, classification was performed on a random, balanced set of 80 samples from the training set (20 of each direction) and tested on 8 samples (2 of each direction) from the test set. This random, balanced sampling was repeated 25 times for each each cross-validation fold, resulting in a total of 175 iterations of training and testing. Overall performance of the classifier was computed as the overall proportion, across the 175 iterations, of correctly labeled test samples. This procedure insured an unbiased estimate of classification performance.

The significance of classifier performance was determined using a random permutation test (Golland and Fischl, 2003) to determine the likelihood of observing a specific accuracy under the null hypothesis that there is no relationship between the data and the specific class labels used to train the classifier A null distribution was generated by rerunning the entire classification cross-validation procedure 1000 times, randomly shuffling the class labels each time. A p-value was then computed by determining the proportion of permuted accuracies that were higher than the observed accuracy. Only classification results where the performance was below $p < 0.05$ on the testing set are reported.

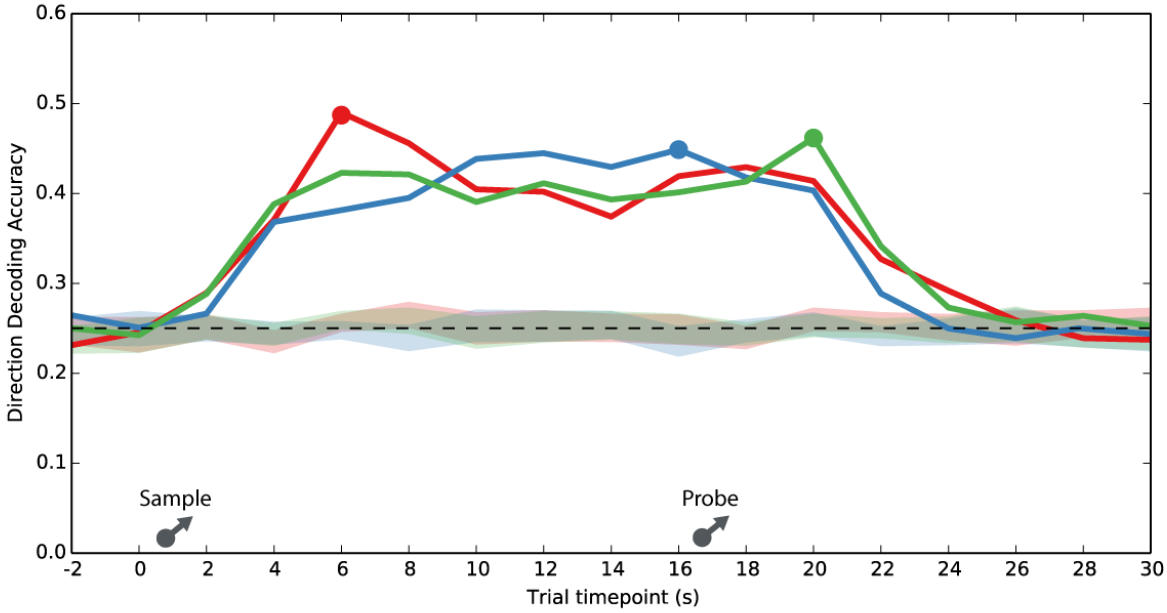

Figure 5.1: Decoding cross-temporal-generalization time courses after training classifiers to decode individual directions of motion (4 possible directions) with data from task-modulated occipital, parietal and frontal regions. Each line represents the mean classification performance across participants ($n = 7$) for a classifier trained on data from a single timepoint. Filled circles indicate the timepoint on which each classifier was trained (e.g., the red line shows the cross-temporal generalization performance for a classifier trained on a timepoint 6 s after the sample presentation and tested on all timepoints on the held-out test set.) Colored bands indicate the distribution of results under the null hypothesis, generated by randomly shuffling the labels and repeating the training and testing procedure many times. Schematic icons of trial events are shown at the appropriate times along horizontal axis. Data are unshifted in time.

High-Throughput Computing

Parts of this research were performed using the computing resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison and the Wisconsin Alumni Research Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

5.3 Results

The aim of the present study was to investigate two aspects of the memory representation of visual motion as measured by stimulus-specific patterns of blood-oxygen level-dependent (BOLD) activity uncovered through the use of MVPA. As can be see in Figure 5.1, classifiers trained on task-modulated data from occipital, parietal and frontal regions using single timepoints during the trial generalized well to other timepoints throughout the trial. We now present results addressing the spatial distribution and temporal stability of the patterns underlying this successful decoding.

Spatial Distribution

In our previous work we demonstrated, using ROI-based decoding, that frontal and parietal regions, on their own, do not appear to contain sufficient information to decode stimulus identity (Riggall and Postle, 2012). However, the possibility remains that these regions could contain information, that when combined with information in posterior regions, provides an improved representation. To address this possibility, we used two closely related approaches, recursive feature eliminination (RFE) and noise-perturbation sensitivity (NPS), to investigate the contribution of individual voxels within our task-modulated brain regions to classification performance.

Recursive Feature Elimination

The first approach we took to identify the spatial distribution of representation was RFE (De Martino et al., 2008). As the name suggests, RFE involves repeatedly eliminating features (i.e., voxels) that are deemed least informative for classification. This procedure is repeated over and over, pruning additional uninformative features each iteration, until further elimination begins to worsen classification performance. Specifically, the classifier is initially trained on the full-featured dataset. The weights from the trained classifier are extracted, and those voxels with the smallest absolute weights (i.e., those contributing least to the classification) are removed. The classifier is then retrained on the new, smaller dataset and the process is repeated until performance of the classifier

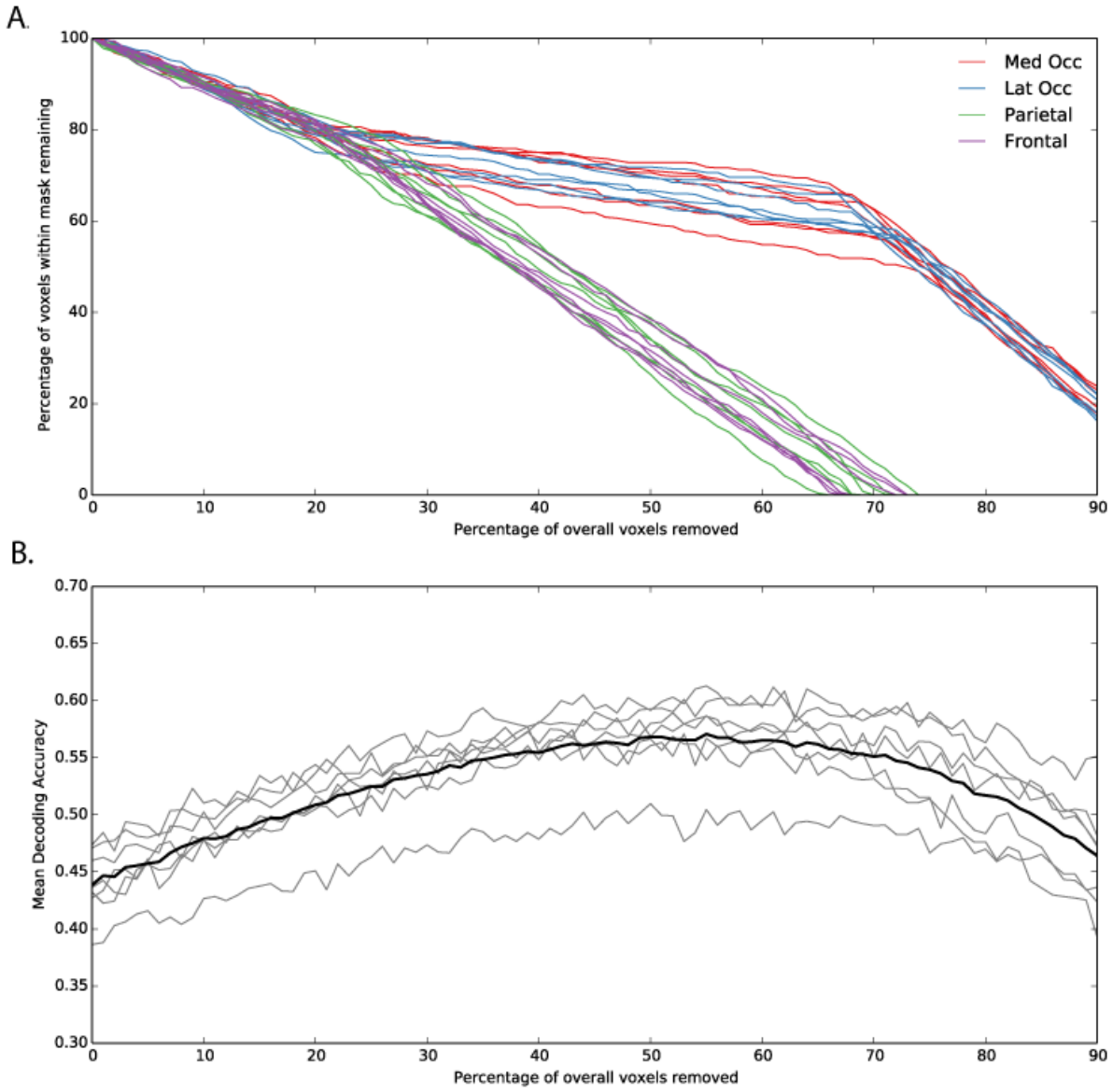

Figure 5.2: Results from RFE applied to data from one timepoint (16 s after the initial presentation of the sample). (*A*) The proportion of voxels remaining in each broad region of interest following each iteration of the RFE procedure. Data from all 7 participants are shown as separate lines, one for each region for each participant. (*B*) Overall mean classification performance after each iteration of the RFE procedure. Each individual subject is shown as a light gray line, the group mean as a black line.

on an independent test test begins to drop.

We ran RFE on our dataset, eliminating the 10 least informative voxels each iteration. To determine which voxels were least informative, we extracted the trained weights from every iteration of our cross-validation training procedure (175 iterations, 6 sets of weights per iteration, one for each pairwise comparison), averaged across iterations, and then assigned the maximum absolute value of each voxel across the six pairwise comparisons as that voxel's informativeness. Voxels were then ranked by this informativeness and the lowest 10 voxels were removed from the set. Ties in rankings at the 10 voxel boundary were broken by randomly selecting the necessary number from all tied voxels.

At each iteration we also recorded which region each voxel being eliminated came from. We then computed the proportion of each region's voxels that remained after each iteration. A plot of the changes with each iteration for one training timepoint can be seen in Figure 5.2A. From this we can see that early on in the elimination procedure an equal number of voxels are eliminated from each region. This pattern soon changes and the vast majority of the voxels removed are from frontal and parietal regions. This pattern continues until all of the voxels in frontal and parietal regions have been removed, at which point the remaining voxels are removed approximately equally from the medial and lateral occipital region. This pattern was remarkably similar across participants: all frontal and parietal voxels were removed from the training set in all participants. These results held true when this process was repeated on all of the individual timepoints in the trial.

Figure 5.2B shows the impact that this feature-elimination process had on classification performance. Performance improved steadily early on with the elimination of voxels, then plateaued as the frontal and parietal voxels were steadily removed, and finally began to drop off as more and more occipital voxels were removed.

Noise Perturbation Sensitivity

The second approach we took for analyzing the spatial distribution of the patterns underlying our decoding results was a NPS approach (Hanson et al., 2004; Hanson and Halchenko, 2008). This

method, similar to the RFE method, computes a measure of informativeness at each voxel. In the case of NPS, that measure is the extent to which the addition of noise at a single voxel impacts overall classification performance. The reasoning behind this approach is that injecting noise into a voxel that is already just noise should have no impact on classification performance. Injecting noise into a voxel carrying signal about the specific classes of interest, however, should negatively impact the overall classification performance as this noise wipes out the signal. The benefit of this method is that is does not directly rely on the specific underlying weights the trained classifier learns, which will be different for varying classification methods, but instead perturbs the data directly. To compute an overall map of sensitivity, this noise addition processes is repeated over-and-over, voxel-by-voxel, to assess the impact of noise at each voxel.

To compute the NPS for our data, we added noise to a single voxel, ran the entire cross validation procedure, and recorded the overall cross-validation accuracy. The noise was uniformly distributed from -2 to 2, corresponding to twice the standard deviation of our z-scored signals. We repeated this procedure 100 times with new random noise each time and took the mean classification accuracy across these 100 iterations as the overall accuracy under noise at that voxel. This process was then repeated for every voxel in the feature-selected dataset. Finally, the accuracy under noise for each voxel was subtracted from the accuracy in the original, noise-free dataset. The absolute value of this difference determined each voxels NPS. To determine an appropriate level at which to deem the impact of the noise meaningful, we added several noise only voxels, consisting of normally distributed noise with mean of 0 and a standard deviation of 1, to the dataset. We then tested the NPS of these voxels, and used the maximum impact they had over all the iterations as a threshold for determining which voxels were significantly impacted by noise, and thus important to the representation being decoded.

The results of this analysis for one subject at one timepoint can be seen in Figure 5.3. The voxels within each region are sorted by their NPS and are shown as vertical lines with heights corresponding to their NPS. As can be readily seen in the figure, no frontal or parietal regions showed a sensitivity beyond the maximum seen in the pure noise voxels. Approximately 30% of

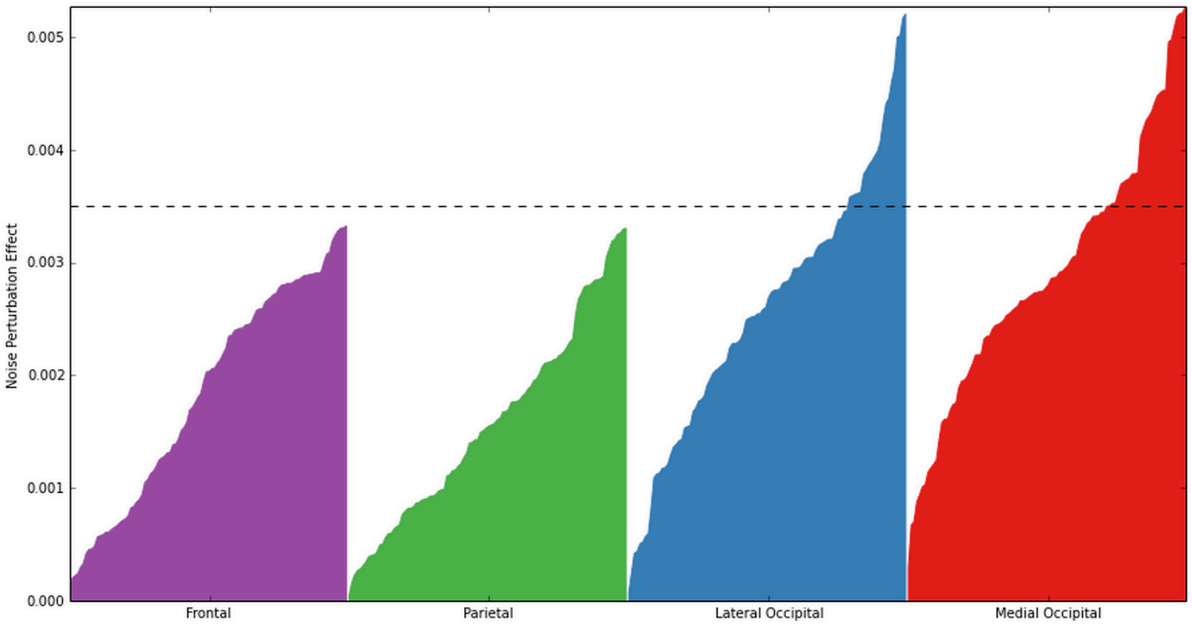

Figure 5.3: Noise perturbation sensitivity (NPS) results for one participant at one training timepoint. Each voxel is presented as a vertical line, colored by the particular broad region on interest which contained it. Within each area the voxels are sorted according to their informativeness. The vertical scale represents the mean impact the addition of noise had on classification accuracy (i.e., change in overall percentage of test trials labeled correctly) over 100 additions of random noise. The horizontal dashed line shows the maximum value observed with the same procedure for a set of control voxels containing only noise. Voxels with values above this line are considered to be significantly informative for the classification.

the medial occipital and lateral occipital voxels showed significant NPS values.

Across the population and across the timepoints tested, on average fewer than 1% of parietal and frontal voxels exceeded the threshold, while an average of 28% (range 21%-40%) of lateral occipital voxels and 34% (range 23%-52%) of medial occipital exceeded the threshold. A comparison of those voxels that were deemed important by the RFE and NPS analyses showed strong agreement between the two, with on average 96% of voxels being labeled as important by the NPS approach also being present in the active feature set at the point of maximum decoding performance during the RFE approach.

Temporal Stability

While the overall stability of our cross-temporal generalization decoding results (see Figure 5.1) are suggestive of a stable pattern over time, when we looked previously at the important voxels on a timepoint-by-timepoint basis, it appeared that the influence of individual voxels varied considerably over time. We now discuss our efforts to investigate this observation, using three different approaches.

Comparison of RFE results across timepoints

First, we examined the extent to which the voxels included as most informative by the RFE analysis overlapped across classifiers trained on each individual timepoint of the delay period. We identified any voxels that appeared in the feature-set at the point of maximum classification performance for any timepoint during the period from 6 seconds following the delay to 20 s (8 timepoints). For each of these voxels we counted the number of times it appeared in the best-performing set for each of the 8 timepoints. Figure 5.4 shows the average proportion of voxels for each possible number of appearances as an important voxel (1-8). Only about 30% were included as informative voxels in all 8 timepoints, with another ~17% covering 7 of the 8 voxels. On the opposite side of the spectrum, a full 20% of voxels, on average, we only included on a single timepoint.

To gauge the approximate distribution of voxels that were active for more than 1, but not all, timepoints, we computed the average of the pairwise differences between all of the timepoints in which a voxel was included as informative. These differences tended to be small (< 4s), suggesting that when a voxel was only informative for a few of the timepoints, these timepoints tended to occur nearby in time..

Training on Multiple Timepoints

The next approach was aimed at more directly identifying stable patterns in particular. Rather than training the classifier on a single time-point, we trained it with data from multiple timepoints, the logic being that the classifier should be better able to separate noise from a stable signal with

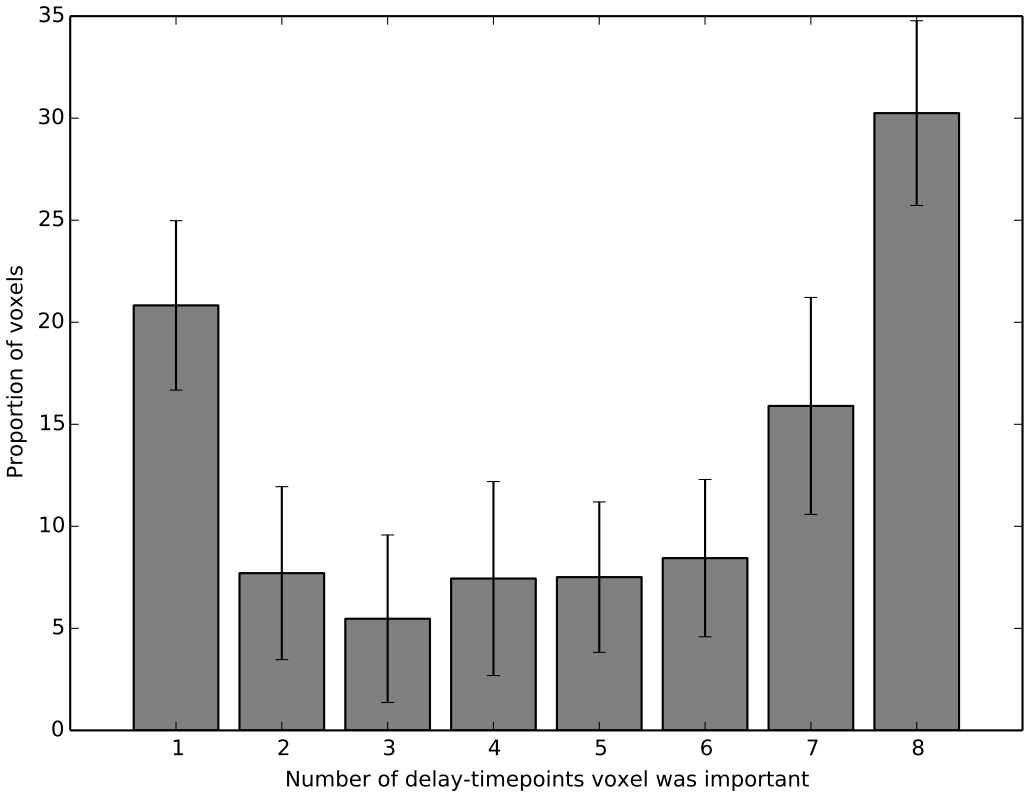

Figure 5.4: Mean cross-timepoint overlap of informative voxels across participants identified using RFE. Computed by taking the union of all voxels that were still included in the feature set at the point of maximum classification for each individual timepoint, then counting the number of timepoints each voxels was included across all the timepoints.

more data and a better estimate of the noise. To begin, we tried this approach by including every timepoint during the delay (from 6 s to 20 s) as separate exemplars. It is important to note that when we did this, we were still following our strict cross-validation procedure, so test trials were always from a completely different run from the training set, so there is no fear of biasing the results by including temporally adjacent train and test timepoints.

Treating each timepoint during the delay as a different exemplar produced a classifier which showed slightly improved overall classification performance compared to training on individual timepoints (see Figure 5.5, blue line). Important voxels, identified through the sameRFE approach

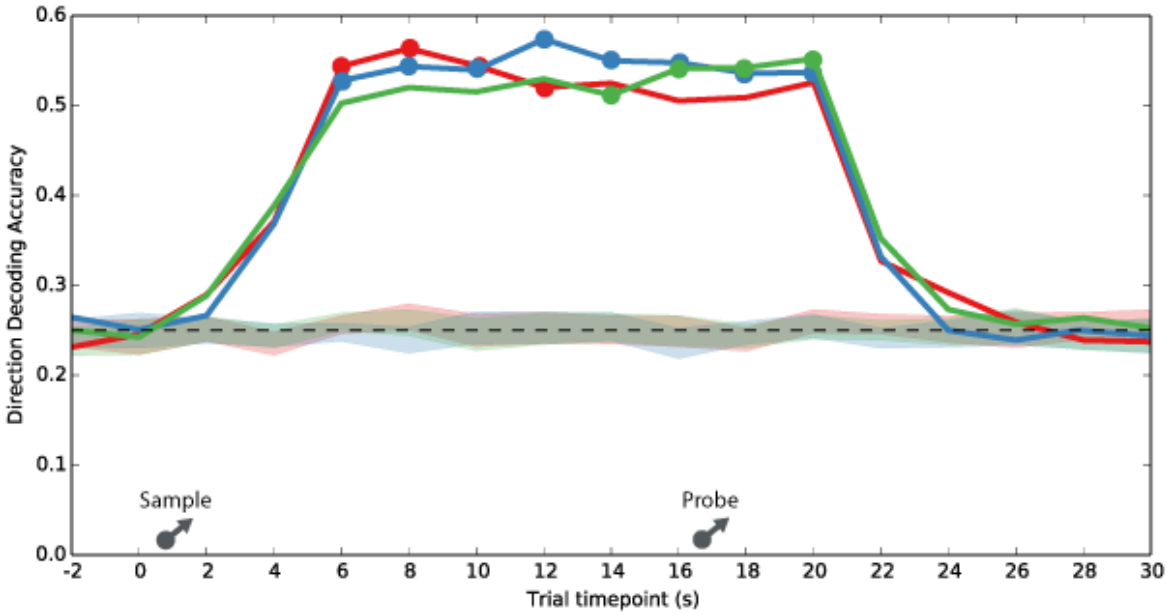

Figure 5.5: Mean decoding cross-temporal-generalization time courses after training classifiers to decode individual directions of motion (4 possible directions). Classifiers were trained on data from multiple timepoints during the trial. The blue line shows a classifier trained on all timepoints during the delay, red the first half of the delay, green the later half. Conventions the same as Figure 5.1.

used with the single timepoint data, were extremely similar to those seen in the single timepoint results (~94% overlap). There were no voxels identified in the multiple timepoint approach that had not been identified as important in at least one of the timepoints of the timepoint-by-timepoint approach.

To further assess the apparent stability of the representations, we repeating this processes, training on data from either the first 4 timepoints of the delay, or the last four. As can be seen in Figure 5.5, each group performed slightly better on the half it was trained on compared to the other half of the delay, but this difference was quite small. Overall both groups performed quite well on all timepoints during the delay. An inspection of the informative voxels showed considerable overlap with the informative voxels from the all-delay-timepoints analysis. The voxels that did not overlap with the all-delay set for each group also did not generally overlap with each other (only ~19% overlap), suggesting slightly different representations across the two halves of

the delay.

Spatiotemporal Classification

The final analysis we performed was a spatiotemporal decoding analysis (Mourão-Miranda et al., 2007). With this approach, the entire delay period of each trial was included as a single training exemplar. Each voxel's individual timepoints are treated as a separate features. That is, a 2000 voxel feature space would become a 16000 voxel-timepoint feature space (2000 voxels times 8 delay timepoints) in our case. The appeal of this approach being that it can identify the specific timepoints at which a voxel is important in the context of all other timepoints within a single model.

We repeated training using this spatiotemporal approach and relied on RFE to eliminate unnecessary features and maximize classifier performance. Overall the trained classifier performed substantially worse (mean of 38% accuracy across participants) compared to either the single timepoint analyses or the multiple-timepoint approach. Inspection of the important voxels showed only a small portion of them (fewer than 10%) across subjects were judged as reliably informative across the delay period. These results suggest that the classifier was unable to extract all of the information present in the much larger feature space, as a much higher proportion were shown to be stable through both of the other approaches.

5.4 Discussion

The aim of the present study was to investigate two important characteristics of the neural representation, as indexed by BOLD activity, of individual motion directions during STM: how spatially distributed and how temporally stable are these representations? Are they spatially limited to only posterior visual areas and consistent during the memory delay, or might they be broadly distributed throughout a large network of the brain and dynamically changing? In our previous work (Riggall and Postle, 2012) we demonstrated successful temporal-cross generalization of de-

coding performance (King and Dehaene, 2014), but only in posterior visual regions, suggesting a stable, spatially limited representation. However, looking at the voxel-by-voxel contributions to whole brain decoding (i.e., the importance maps or sensitivity maps) we found the opposite pattern: important voxels were found in broadly distributed regions and their importance varied throughout the delay period.

To attempt to reconcile these contradictory observations we undertook a detailed assessment of the underlying voxel-by-voxel contributions to classification performance. We looked at the spatial distribution of the representation with two different approaches, RFE and NPS, which both showed that few, if any, voxels outside the posterior visual regions were actually informative to the overall classification. These results provide additional evidence supporting the *sensory recruitment hypothesis*, whereby the short-term retention of information is accomplished using the same neural systems used for sensory processing, rather than specialized storage buffers (Awh and Jonides, 2001; Pasternak and Greenlee, 2005; Jonides et al., 2005; Postle, 2006; D'Esposito, 2007; Serences et al., 2009).

We assessed the temporal stability of the representations in several ways. First, we compared the informative voxels identified by the RFE and NPS across timepoints, finding a relatively even split between voxels that were important throughout the delay, voxels that were important for several points during the delay, and voxels that were only important on one or two timepoints. Next we assessed performance when training the classifier on all the timepoints in the trial as separate examples, which identified a set of voxels that represented a stable representation throughout the delay. Splitting this approach, and only including early delay or late delay voxels showed general agreement, but some evidence for additional information present in each half not present in the other. Finally we performed a combined spatiotemporal analysis, which provided generally complementary, though likely underpowered, results. The interpretation of these temporal stability results is less clear. On the one hand, they suggest a relatively stable representation, challenging recent results suggesting a much more dynamic representation (at the level of BOLD; Sreenivasan et al., 2014b). On the other hand, we do find evidence for a change in representation over the

delay, based on training with only early or late delay examples, suggestive of a recoding of at least some of the representation. This would be consistent with recent electrophysiological results in the monkey suggesting an initial dynamic representation that eventually settles to a stable point (Stokes et al., 2013).

An important unresolved question related to the present work is that of what, exactly, we are decoding when applying classification methods to BOLD data. Several proposals have been put forth, including: uneven distributions in cortical columns coding for information across voxels (Boynton, 2005; Kamitani and Tong, 2005; Chaimow et al., 2011), accumulated signal in draining veins (Gardner, 2010; Thompson et al., 2011), larger scaler, weak cortical organization (Op de Beeck, 2010; Freeman et al., 2011), or a spatiotemporal combination of multiple sources of information (Kriegeskorte et al., 2010). For the purposes our of current results, the specifics don't matter much, as we were looking for differing contributions from distinct, spatially separated areas. However, as studies move to higher-field strengths and better spatial resolutions these questions will become much more important for truly understanding the underlying representations, to the extent possible with BOLD.

In the current study we took a blunt-force approach to identifying informative voxels and stuck with the same classification approach we had used in our original report, L2-regularized logistic regression. We performed only course tuning of the regularization parameter λ to limit overfitting and promote cross-generalization of the data, without excluding possibly informative voxels (Rasmussen et al., 2012). We chose to stick with logistic regression because the weights produced during training are directly interpretable as the effect each voxel has on the overall likelihood of the classification decision. This is not true of other classification methods, such as SVM, where methods akin to NPS must be used to get a voxel-by-voxel measure of importance (LaConte et al., 2005; Rasmussen et al., 2011; Baehrens et al., 2010).

While we do not believe that the specific choice of classification algorithm determined the results observed in this study, it is possible that more sophisticate approaches, such as the ElasticNet (Zou and Hastie, 2005), group lasso Yuan and Lin (2006), or sparse overlapping sets lasso (Rao et

al., In Press), might provide clearer results, particularly with regard to groups of voxels that may be highly correlated. Another alternative would be an ensemble of classifiers approach (Cabral et al., 2012). Rather than relying on a single classification method which might be best suited for a single area, combining the results of classification with a number of different classifiers might provide a more complete extraction of any relevant information in the data, particularly when paired with NPS metrics to insure true importance is detected (Baehrens et al., 2010).

6

Frontoparietal contributions to the short-term retention of motion and color

Given the recent empirical support for the sensory recruitment hypothesis, which posits that short-term storage occurs in the same neural systems used for initial sensory processing, an important question remains: what functions, if not storage, are the frontoparietal areas that routinely show load-dependent elevated activity during the delay-period of short-term memory tasks doing, if not storage? One compelling possibility is that these regions are involved in directing top-down attention towards these internal representations in sensory regions to support their maintenance. Under this hypothesis we would expect to see increased involvement of these regions with increased memory load, as additional attentional resources are needed to maintain additional memory items. In the current study we manipulated load in a short-term delayed-recall task for visual motion or color. During an initial session we collected fMRI data while participants performed this task. These data were used to identify load-sensitive frontoparietal regions which were subsequently targeted during a transcranial magnetic stimulation (TMS)/electroencephalography (EEG) session. We found differing patterns of functional connectivity depending on task condition and load in the fMRI data and increases in effective connectivity, as measured by the TMS-evoked response, at higher loads while stimulating frontal and parietal regions, but not a visual region (area MT). These results are consistent with a top-down attentional mechanism, controlled by prefrontal cortex (PFC) and intraparietal sulcus (IPS), supporting short-term maintenance by "sensory" regions.

Adam C. Riggall, Nathan S. Rose, Michael J. Starrett, & Bradley R. Postle

6.1 Introduction

A growing body of literature finds evidence for visual stimulus representation within posterior visual regions during short-term memory (STM) (Harrison and Tong, 2009; Serences et al., 2009; Ester et al., 2009; Albers et al., 2013; Christophel et al., 2012; Christophel and Haynes, 2014; Vicente-Grabovetsky et al., 2014; Linden et al., 2012). Taken together, these results provide compelling evidence supporting the *sensory recruitment hypothesis*, which posits that the short-term retention of information is accomplished using the same neural systems used for sensory processing, rather than specialized storage buffers (Awh and Jonides, 2001; Pasternak and Greenlee, 2005; Jonides et al., 2005; Postle, 2006; D'Esposito, 2007; Serences et al., 2009; Sreenivasan et al., 2014a).

These results are in direct contrast to decades of results highlighting the involvement of frontal and parietal regions, which reliably show elevated, memory-load dependent activity during the delay period of STM tasks (reviewed in Curtis and D'Esposito, 2003). How, then, to explain the discrepancy? One compelling explanation is that these regions, rather then being involved in representation-related activities, are instead involved in directing attention internally to stored-representations (Lepsien and Nobre, 2007; Munneke et al., 2012). This idea helps explain the considerable overlap often seen within these regions across putatively different cognitive tasks (Ikkai and Curtis, 2011; Jerde et al., 2012; Lückmann et al., 2014).

In order to test this idea, the current study examined functional and effective connectivity during a short-term memory task. To do this we used two complementary approaches in two different modalities: with functional magnetic resonance imaging (fMRI) looking at pychophysiological interactions (PPI; Friston et al., 1997) and with transcranial magnetic stimulation (TMS)/electroencephalography (EEG) looking at the propagation of the TMS-evoked response (TMS-ER; Rogasch and Fitzgerald, 2012).

The current study tested the hypothesis that frontoparietal regions with elevated activity during the delay-period of a short-term memory task are providing a top-down attentional signal to stimulus-feature-specific posterior regions where the relevant information is being retained. We tested this by scanning subjects (fMRI) while they performed a task requiring delayed-recall of

directions of motion or of colors. In a second session we collected EEG data while administering TMS to functionally localized targets in lateral prefrontal cortex (lPFC), intraparietal sulcus (IPS), and the middle temporal area (MT).

We predicted that the top-down attentional signal from frontoparietal regions would be stronger during higher memory loads, leading to higher functional connectivity between frontal, parietal, and posterior visual regions at higher loads (load 1 vs. load 3) as measured with fMRI, but different patterns of connectivity when different features were being remembered (motion vs. color). That is, we expected a *quantitative* difference between memory loads within task condition and a *qualitative* difference in the pattern of connectivity between task conditions. We also expected higher effective connectivity at higher loads (as measured by the spread and strength of the TMS-evoked response (TMS-ER) in the EEG) after lPFC stimulation and IPS stimulation (top-down stimulation), but not following MT stimulation (bottom-up stimulation).

6.2 Materials and Methods

This study was composed of two sessions: an initial fMRI session and a follow-up TMS/EEG session. During the initial fMRI session participants performed a delayed-recall task for visual motion and color, as well as a functional localizer task for motion. Data from this session were used to generate TMS targeting locations for the TMS/EEG session, during which participants performed a similar delayed-recall task while TMS was applied to MT, IPS, and dorsolateral prefrontal cortex (dlPFC).

Participants

Seven volunteers (1 female) between 19 and 31 years of age (mean 25.4 years) were recruited from the University of Wisconsin-Madison community and paid for their participation. All subjects had normal or corrected-to-normal vision. Participants gave written informed consent according to procedures approved by the Health Sciences Institutional Review Board at the University of

Wisconsin-Madison and were screened for the presence of neurological and psychiatric conditions and other risk factors related to the application of TMS.

fMRI Session

Behavioral Paradigm

Participants performed eight runs of a delayed-recall task for directions of visual motion or colors (blocked by run; Fig. 6.1) while being scanned (fMRI). Before each block subjects were instructed to either remember motion directions or to remember dot colors.

Each trial began with the presentation of an endogenous cue arrow (cue, 1 s), directing participants to remember items on the left or right side of the visual display. Next, the cue was replaced with a white fixation cross and three dot patterns were presented sequentially on each side of the display (sample, 2 s). Each dot pattern was displayed for 500 ms and the patterns were separated by an interstimulus interval of 250 ms. During direction blocks, all three patches within each hemifield were colored the same random color, and one or all three could be moving 100% coherently in a random direction. During color blocks, all three patches within each hemifield were moving 100% coherently in the same random direction, and one or all three patches could be randomly colored. Participants were instructed to remember the motion direction or color of dots on the cued side of the display. A variable length delay followed (delay, 6, 8, or 10 s, evenly distributed).

After the delay a response ring appeared on each side of the display (response, 4 s). The ring on the cued side of the display contained an indicator bar pointing from the center of the ring to a random position on the ring. At the same time, a numeral replaced the fixation point, indicating the serial position of the item that participants needed to recall. Participants had to recall the item from the specified serial position by rolling the ball of an MR-compatible fiber-optic trackball (Trackball 2; Current Designs) left or right to rotate the response indicator bar. During direction blocks the response rings were colored the same colors as the dot patterns had been, and rotating the indicator also rotated a set of moving dots within the response ring that moved in the direction of the indicator. On color blocks, the response rings were composed of a color wheel containing

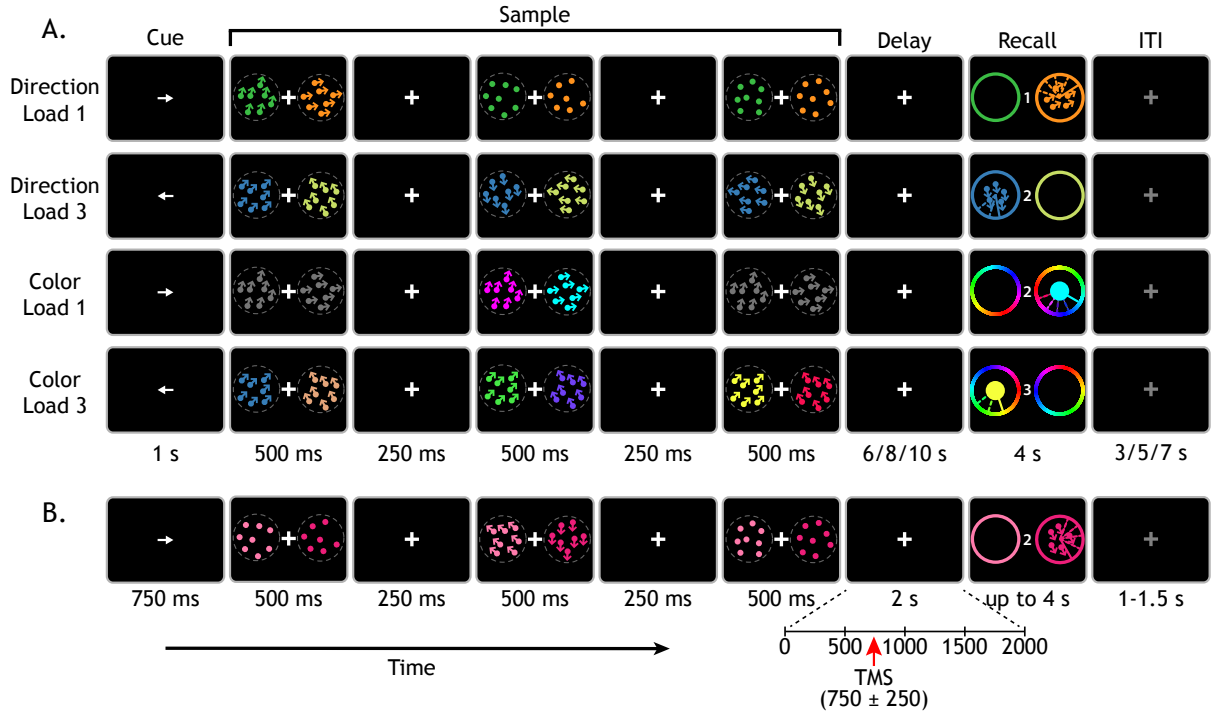

Figure 6.1: Behavioral task. A. fMRI task. Participants maintained the direction of motion or color of dots (depending on the block) within the cued half of the visual field over a variable delay period. At the end of the delay period they were required to recreate the item that had appeared in the indicated serial position by rotating a response indicator bar with a trackball. B. TMS/EEG task. Procedure and conditions were similar to the fMRI task, with changes in timing. TMS was applied during the delay period on every trial, to MT, IPS, or PFC (blocked).

all possible color responses, and rotating the indicator bar changed the color of a large circular patch in the center of the ring to match the color that the indicator bar pointed at. The color wheel was rotated from trial to trial to prevent the association of colors with specific spatial locations. Participants were instructed to recreate the indicated remembered item as quickly and accurately as possible, pressing a button on the trackball when they had finished. After the response period a variable intertrial interval (ITI) followed (3, 5, or 7 s, distributed evenly) during which a gray fixation cross was present.

Before scanning, subjects were trained on the trackball response for both direction and colors and practiced four blocks of 20 trials (2 each of direction and color) to familiarize themselves with the experimental procedure

All stimuli were presented on a black background. The fixation cross, cue arrow, and serial position indicator where presented in white and subtended ~0.13°. The sample stimuli consisted of two circular patches (8° diameter) of dots (~ 0.24°) centered 8° degrees on either side of the screen. The sample was composed of three independent sets of dots, of which either one or all three would need to be remembered.

On direction blocks, the directions of the sample were chosen randomly on every trial from the full 360° circle. For load 1 trials, one of the three sequential presentations (sequential position randomized) would move 100% coherently (at 15°/s) in a random direction while the other dot sets remained stationary. On load 3 trials, all three sets moved coherently in directions that were chosen randomly, constrained so that they were all at least 50° apart and not more than 165° to prevent overlapping or exactly opposite motions. Each hemifield was randomized separately, with the constraint that both patches would always either be moving or stationary, never one moving and one stationary (i.e, load and serial position were identical in the two hemifields). The color of the dots within each hemifield was chosen randomly each trial from the same color space as the color blocks (see below). All dot sets within a hemifield were the same color, leaving direction as the only distinguishing feature of the items.

On color blocks, the colors of the sample were chosen randomly from a circular CIE L $*$ a $*$ b color space. The circle was centered in the color space at (L=70, a=20, b=38) with a radius of 60. Thus the colors had equal luminance, varying only in hue and saturation. For load 1 trials, one of the three dot sets was colored randomly, while the other two dot sets were a dark gray (sequential position of the colored item was randomized). On load 3 trials, all three dot sets would be colored, constrained such that the colors were not within 50° in the circular color space to insure discriminability. As with direction trials, each hemifield was randomized separately, constrained so that load and serial position were the same in each hemifield. The direction of the dots within each hemifield was chosen randomly each trial and all dot sets moved, leaving color as the only distinguishing feature for the items.

Participants completed 240 trials over the course of 8 runs while in the scanner. Task condition

(2 levels: direction, color), attended hemifield (2 levels; left, right), memory load (2 levels, 1, 3), and probed-item sequential position (3 levels: 1st, 2nd, 3rd) were counterbalanced across runs, with participants completing 10 trials in each cell of the fully factorial design. Participants alternated blocks of direction and color trials, with the order counterbalanced across participants.

Experimental stimuli were controlled by the Psychophysics Toolbox (http://psychtoolbox.org; Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) running in MATLAB (MathWorks), rear projected onto a screen in the scanner bore near the participants head, at 60 Hz (Silent Vision 6011; Avotec), and viewed through a coil-mounted mirror.

Data Acquisition and Preprocessing

Whole brain images were acquired with the 3T scanner (Discovery MR750; GE Healthcare) at the Lane Neuroimaging Laboratory at the University of Wisconsin-Madison. A gradient-echo, echo-planar sequence (2 s TR, 25 ms TE) was used to acquire data sensitive to the blood-oxygen level-dependent (BOLD) signal within a 64 x 64 matrix (39 sagittal slices, 3.5 mm isotropic). Eight runs of the delayed-recognition task were obtained for each subject, each lasting 10 min, 26 sec (313 volumes). The first 8 seconds of all task runs were discarded to insure a steady state of tissue magnetization. A high-resolution T_1-weighted image was also acquired for all subjects with a fast spoiled gradient-recalled-echo (FSPGR) sequence (8.132 ms TR, 3.18 ms TE, 12° flip angle,156 axial slices, 256 x 256 in-plane, 1.0 mm isotropic).

Participants also performed two runs of a lateralized motion localizer. Two dot patterns were present on the screen in the same lateralized locations used in the main experiment. The localizer pseudorandomly cycled between both patches moving, only the left moving, only the right moving, or neither moving, changing every 18 seconds. Moving patches consisted of dots alternating every second between inward and outward motion, centered on the center of the circular aperture. Each run lasted 5 minutes (150 volumes), with the first 8 seconds discarded before further processing. During each run the four combinations of stimulus motion were shown four times each. During the localizer participants performed a luminance change-detection task on the fixation

cross to insure stable fixation.

The data were preprocessed using the Analysis of Functional NeuroImages (AFNI) software package (Cox, 1996). All functional volumes were spatially aligned to the first volume of the first run using a rigid-body realignment and corrected for slice time acquisition. Data were skull-stripped and spatially smoothed with a 6 mm FWHM Gaussian kernel. To facilitate group level analysis, individual subject statistical maps were transformed into Talairach space (Talairach and Tournoux, 1988).

Behavioral Analysis

Behavioral performance was assessed using a mixture-model (Zhang and Luck, 2008; Bays et al., 2009; Zokaei et al., 2011). This model breaks down the distribution of errors (the angular deviations between the samples and the responses) into estimates of target accuracy (proportion of responses toward the target direction), error (proportion of guesses, and in the case of load 3 trials, proportion of nontarget responses) and a measure of precision (the variability of target and nontarget responses). Responses are modeled to be a mixture of a Von Mises distribution (the circular analog of the Gaussian distribution) with a concentration of κ for all target and nontarget responses plus a uniform distribution of random responses (guesses). The concentration parameter, κ, corresponds to the precision of the target (and nontarget) responses, with larger values indicating more precision (i.e., less variability). The model is defined as follows:

$$p(\hat{\theta}) = P_T \phi_\kappa(\hat{\theta} - \theta) + P_{NT} \frac{1}{m} \sum_{i}^{m} \phi_\kappa(\hat{\theta} - \theta_i^*) + P_G \frac{1}{2\pi}$$

where θ is the target responses, $\hat{\theta}$ is the actual responses, P_T is the probability of reporting the target, and P_{NT} is the probability of reporting a nontarget (of m possible nontarget options). P_G is defined as $1 - P_T - P_{NT}$, and represents the probability of responding at random (Bays et al., 2009; Zokaei et al., 2011).

Parameter estimates for P_T, P_{NT}, P_G, and κ were obtained using expectation maximization using MATLAB routines (available at http://www.bayslab.com). Responses were entered for all

trials with a confirmed response, and separate estimates were obtained for each participant for each task condition (direction and color) at each load (1 and 3) separately.

Univariate Analyses

A univariate general linear model (GLM) was used to identify load-sensitive delay-period activity. The model included regressors for each within-trial event of the task (e.g., cue, each individual dot pattern, delay, response). Direction and color blocks were modeled separately, as were remember left and remember right trials. The cue was modeled with a 1 s boxcar. The sample presentations were modeled as a set of 0.5 s boxcars, with different regressors for each aspect of the dot motion (i.e., moving, static, colored, uncolored). The delays were modeled with a pair of regressors, one which captured overall changes during the delay, and a second amplitude-modulated regressors which directly contrasted load 1 and load 3 delays. Both of these were boxcars covering the delay period (6, 8, or 10 s). The response period was modeled with four regressors: a 4 s boxcar representing the presence of the response aperture, a boxcar whose duration lasted as long as the response time on the trial, an amplitude-modulated boxcar that contrasted loads whose duration lasted as long as the response time on each trial, and an impulse response at the time of the button press response. All event regressors were convolved with a canonical hemodynamic response function (HRF). The model also included nuisance regressors for run, head-motion, and within run polynomial drift (up to 5th order). Each of these independent regressors was entered into a GLM for analysis using AFNI. Individual subject contrast images were aligned to Tailarach space and entered into a second-level (random effects) analyses. To control for family-wise type I errors, the results were cluster-thresholded based on results from 10,000 Monte Carlo simulations using AFNI's AlphaSim, with a minimum individual voxel threshold of $p < 0.05$ and a minimum cluster size of 32 contiguous voxels.

The localizer task was modeled with four boxcars, representing the moving and stationary periods in the left and right visual fields separately. The duration of the boxcars was dependent on the duration of continuous stimulation for each period, which varied given the random nature

of the stimulus presentation, but could take on one of four values: 18, 36, 54, or 72 s. All block regressors were convolved with a canonical HRF.

Functional Connectivity Analyses

To assess functional connectivity, a PPI analysis was used (Friston et al., 1997; O'Reilly et al., 2012). Seed timeseries were created by extracting mean timeseries from seed regions using a 7mm sphere region centered on the seed voxel (See table 6.1). Prior to averaging over the voxels within the sphere, the voxel-by-voxel effects of no interest were removed by subtracting the effects-of-no-interest fits from the original GLM. Regressors for the psychological constructs of interest were created by labeling all the points during the delay period with 1 or -1 for direction against color (for all trials) and load 1 against load 3 (direction and color separately). These psychological-construct regressors were then convolved with a canonical HRF. Finally, the convolved construct regressors were multiplied timepoint-by-timepoint with the individual seed regressors to create interaction regressors. For each construct of interest the original GLM was refit, with the inclusion of three additional regressors representing the seed of interest, the convolved construct of interest, and the interaction between the two.

TMS/EEG Session

Behavioral Paradigm

Participants performed the same delayed-recall task for motion direction and color as in the fMRI session, with the following modifications: The cue arrow was presented for 750 ms, the delay lasted 2 s, the response period ended immediately after the subject pressed the button to confirm their response (or timed out after 4 s), and the ITI varied from 1 to 1.5 s (rectangular distribution). Single-pulse TMS was applied during the delay at 750 ± 250 ms after final stimulus offset.

Participants completed 192 trials over 6 blocks for each TMS site (MT, IPS, dlPFC). Order of TMS sites was randomized and counterbalanced across subjects. Within site, task condition (2 levels: direction, color), attended hemifield (2 levels; left, right), memory load (2 levels, 1, 3),

and probed item sequential position (3 levels: 1st, 2nd, 3rd) were fully counterbalanced across blocks, with participants completing 24 trials in each cell of the fully factorial design. Participants alternated blocks of direction and color trials, with the order counterbalanced across participants.

Experimental stimuli were controlled by the Psychophysics Toolbox running in MATLAB, presented on a 24" LCD monitor at 120 Hz (VG248QE, Asus).

TMS Stimulation and Targeting

TMS targeting was controlled using a Navigated Brain Stimulation (NBS) system (Nexstim) which employs 3D, infrared-based frameless stereotaxy to align the positions of TMS coil and subject's head within the reference space of the individual's structural MRI. Optimal alignment between MRI fiducials and digitized scalp landmarks (nasion, left and right tragii) was verified prior to each experiment session. NBS also calculates on-line the distribution and intensity (expressed in V/m) of the intracranial electric field induced by TMS, using a model of subjects' head and brain and taking into account the exact shape, 3D position and orientation of the coil. Stimulation intensity was adjusted for each participant and TMS site to achieve 95-105 V/m at the specified site.

TMS was targeted to three different locations in the left hemisphere: MT, IPS, and PFC, all derived from the data collected during the fMRI session. The MT stimulation site was localized by using the motion localizer, focusing on the spot near the intersection of the lateral occipital sulcus and the ascending limb of the inferior temporal sulcus (Dumoulin et al., 2000) showing the strongest difference when comparing moving to static blocks from the right visual field regressors. The IPS and PFC target sites were both identified using data from the main experimental task, isolating regions that showed a load-dependent change in activity for both the direction and color trials, collapsed across hemifields. The IPS site was chose as the most posterior and lateral activation peak within the IPS. The PFC target was isolated as the peak near the junction of the precentral sulcus and the inferior frontal sulcus, the inferior frontal junction (IFJ). The reproducibility of the stimulation coordinates across sessions was guaranteed by a virtual aiming device that indicated in real-time any deviation from the desired target greater than 2 mm. To avoid con-

tamination of the EEG by auditory artifacts induced by the click of the TMS discharge, participants listened to masking white noise through in-earbud headphones. The volume of the masking noise was adjusted immediately prior to the experimental session for each subject until they could not hear the TMS discharge.

EEG Data Acquisition and Preprocessing

EEG was recorded with a 60-channel TMS-compatible amplifier (Nexstim). This amplifier gates the input during the TMS pulse by means of a sample-and-hold circuit (Virtanen et al., 1999), holding the amplified output constant from 100μs before to 2 ms after TMS discharge, limiting the direct TMS-induced artifact. Electrode impedance was kept below 5kΩ. A single electrodes placed on the forehead was used as the reference and another two electrodes placed outside the lateral canthi were used to record the horizontal electrooculogram (EOG). The EEG signals were filtered at (0.1-500 Hz) and were sampled at 1450 Hz with 16-bit resolution. At the end of the experiment, electrode positions on the participants head were digitized using NBS.

Data were processed off-line with EEGLAB (version 13.2.1b Delorme and Makeig, 2004) within MATLAB (MathWorks). Data were band-pass filtered between 1 and 100 Hz. A notch filter with a stop band of 55 to 65 Hz was used to remove line noise. The data were then downsampled to 290 Hz and epoched by trial. A four step procedure was used to remove non-neural noise from the data. First, epochs with large artifacts were removed via visual inspection. Next, independent component analysis (ICA) was used to separate the EEG into putative sources (Jung et al., 2000; Delorme et al., 2007). The results of this separation were visually inspected and epochs with large artifacts in the component timeseries were removed. ICA was performed again on the pruned dataset and components reflecting blinks, eye movements, residual muscle activity and TMS-related artifacts were removed. Finally, the data were rerefrenced to the average of all 60 electrodes.

Source Modeling

Source modeling was performed using the methods described by Casali et al. (2010). Individual cortical meshes (3,004 vertices) were created with the Statistical Parametric Mapping software package (SPM5, http://www.fil.ion.bpmf.ac.uk/spm). This involved warping binary masks of the skull, scalp and cortex obtained from individual MRIs to the Montreal Neurological Institute (MNI) atlas, allowing for the creation of skull, scalp and cortex meshes in MNI space. The meshes were coregistered to the positions of the EEG sensors by rigid-body rotations and translations to coregister digitized landmarks with MRI landmarks (nasion, left and right tragus). Next, a model of the conductive head volume was created using the three-spheres Berg method (Berg and Scherg, 1994) using the Brainstorm software package (http://neuroimage.usc.edu/brainstorm Tadel et al., 2011). The model was constrained to a source space consisting of 3004 fixed dipoles oriented normally to the surface of each node of the cortex mesh.

The inverse solution was then calculated on a trial-by-trial basis with an empirical Bayesian approach as implemented in SPM5 (Friston et al., 2002; Phillips et al., 2005; Tikhonov and Arsenin, 1977). For this analysis, the covariance matrix was assumed independent across EEG electrodes, and covariance components were modeled by two priors: the weighted minimum norm constraint and a Gaussian distribution of source covariance along the geodesic (smoothness parameter = 8 mm), which enforced correlation among neighboring sources. These priors were estimated directly from the data with restricted maximum likelihood (Friston et al., 2002, 2006; Phillips et al., 2005) Finally, to compute the overall current evoked by TMS in different cortical areas, individual cortical surface regions were attributed to different Broadmann areas (BAs) with an automatic anatomical classification method that mapped the individual cortical surface regions to the region of interest (ROI) masks provided by the WFUPickAtlas tool (http://ansir.wfubmc.edu Maldjian et al., 2003).

Effective Connectivity

Given the causality introduced by the TMS-ER, measures of effective connectivity can be derived from TMS/EEG data. To quantitatively evaluate task-driven differences in effective connectivity, the source-localized data were submitted a standardized, data-driven procedure that characterizes the electrical response of the brain to TMS by means of two synthetic indices: significant current density (SCD) and significant current scattering (SCS) (Casali et al., 2010). SCD represents the sum of the absolute amplitude of all significant TMS-evoked currents observed, in $\mu A/mm^2$, over a given time interval and/or cortical region, which were identified with a nonparametric statistical procedure. SCS is calculated as the sum of the geodesic distances (in mm) between the stimulated brain region and any significant current source over a given time interval and cortical volume. Thus this index captures the spatial spread of TMS-evoked currents to distal brain regions, growing larger as significant TMS activations spread farther away from the stimulated brain area. The comparison of these two values across different task conditions allows one to characterize changes in effective connectivity under different neural. To that end, the inverse solution was generated separately for different combinations of conditions (e.g., TMS to MT, direction trials, remember left vs. remember right) to allow for comparisons of the effect of task condition and memory load on connectivity between MT, IPS, and PFC.

6.3 Results

fMRI Session

Behavioral Results

Participants performed a delayed-recall task for directions of motion or colors which required recreating a remembered direction or color within a 4 s response window. Participants were significantly more likely to fail to respond within the 4 s window on color trials compared to direction trials (Fig. 6.2A; ($t_{(6)} = 3.45$, $p < 0.014$). When considering only trials with a valid response,

response times were significantly faster on direction trials compared to color trials ($t_{(6)} = 6.15$, $p < 0.001$). Within direction trials with valid responses, a two-way repeated measures ANOVA with load (1 and 3) and hemifield (left and right) as factors revealed a significant effect of load (Fig. 6.2B; $F_{(1,6)} = 13.55$, $p = 0.010$), a trending effect of hemifield ($F_{(1,6)} = 5.64$, $p = 0.055$), and no interaction ($F_{(1,6)} = 1.97$, $p = 0.21$). For color trials with valid responses, the same analysis revealed no significant effects (load: $F_{(1,6)} = 0.45$, $p = 0.52$, hemifield: $F_{(1,6)} = 2.30$, $p = 0.18$, interaction: $F_{(1,6)} = 0.02$, $p = 0.90$)

Memory precision was analyzed using the method of Bays et al. (Bays et al., 2009; Zokaei et al., 2011). This method fits a mixture model to the behavioral response errors (the circular distances between response values and the actual values of the probed memory stimuli) to obtain estimates for the proportion of correctly recalled targets, the precision of those responses (the distribution of error around the target), and the proportion of errors (both nontarget responses and guesses; see Materials and Methods). The model was fit separately using the relevant trials of specific conditions of interest.

We examined the precision estimates (κ, the concentration parameters of the von Mises distribution) using a two-way repeated measures ANOVA with load (1 or 3) and hemifield (left or right) as factors. For direction trials we observed a significant effect of load (Fig. 6.2C; $F_{(1,6)} = 7.11$, $p = 0.037$), no effect of hemifield ($F_{(1,6)} = 0.51$, $p = 0.50$), and no interaction between the two ($F_{(1,6)} = 0.37$, $p = 0.56$). A similar pattern was observed for color trials, with a significant effect of load ($F_{(1,6)} = 24.5$, $p < 0.003$), no effect of hemifield ($F_{(1,6)} = 0.57$, $p = 0.48$), and no significant interaction ($F_{(1,6)} = 0.63$, $p = 0.46$).

Having found no evidence that the remembered hemifield influenced behavior, we collapsed responses across hemifields so we could further analyze behavioral performance. We refit the mixture-models with the collapsed data and extracted the proportion of target responses (Fig. 6.2D). A two-way repeated measures ANOVA with load (1 and 3) and condition (direction and color) as factors revealed no significant differences in the proportion of target responses (load: $F_{(1,6)} = 1.99$, $p = 0.21$, condition: $F_{(1,6)} = 2.08$, $p = 0.20$, interaction: $F_{(1,6)} = 1.26$, $p = 0.30$).

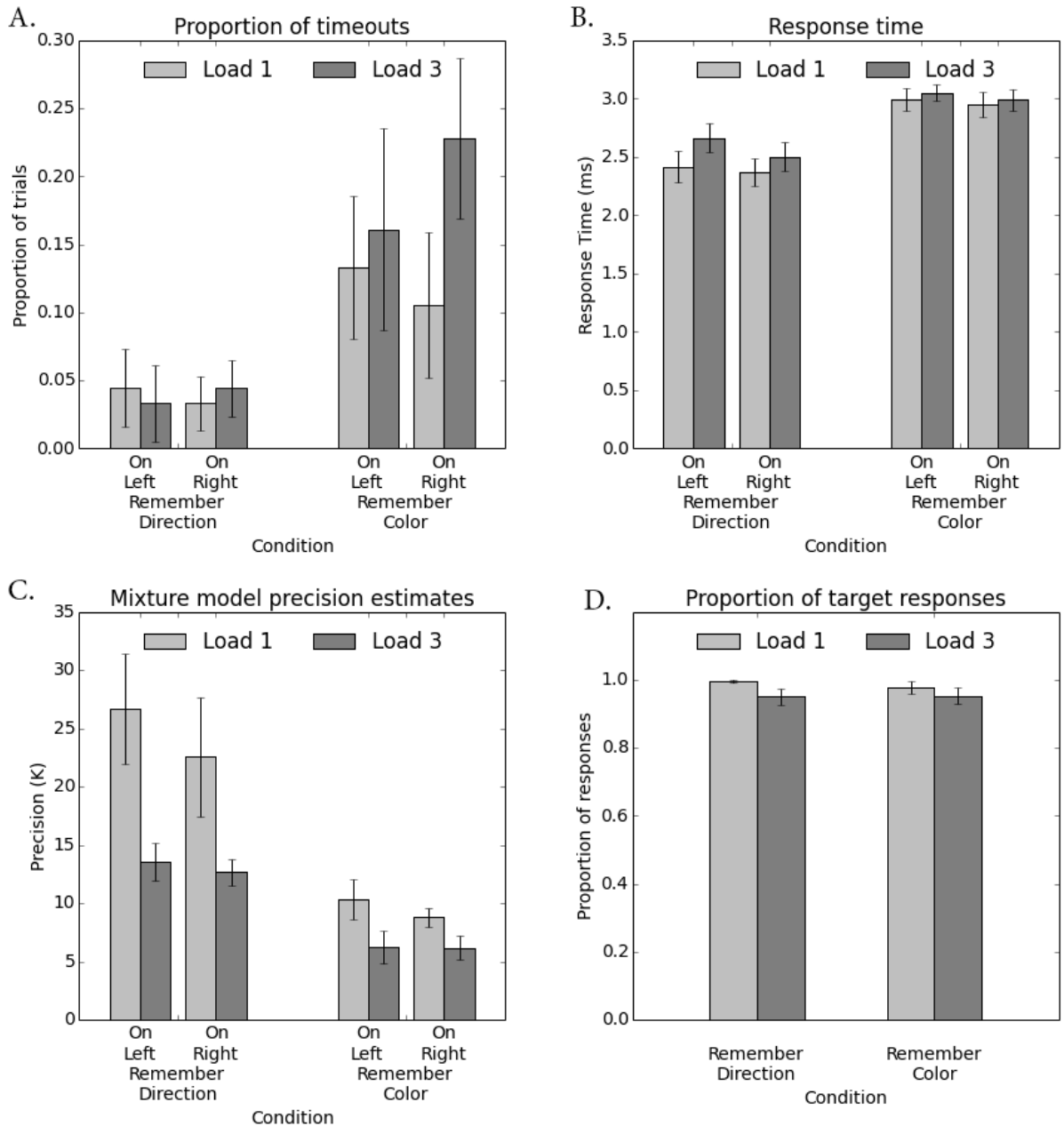

Figure 6.2: Behavioral results from fMRI session. (*A*) Mean proportion of trials which timed out after participants (n=7) did not respond within the 4 s response window, which occurred more often on color trials compared to directions trials. (*B*) Mean response times on trials with a validly recorded response, which were longer for color trials than direction trials. (*C*) Estimates of the precision (κ) of responses from the mixture model, which showed a significant effect of load. (*D*) Estimates of the proportion of responses (P_T) to the target from the mixture model fits, which showed no differences across conditions or loads. For all plots load 1 results are in light gray, load 3 results in dark gray.

Table 6.1: Average PPI seed region locations across participants in Tailarach coordinates, for illustration purposes only. PPI analyses were done with seed regions identified in individual subjects data.

	Left			Right		
Seed	X	Y	Z	X	Y	Z
V1	4	80	-8	-8	78	-4
MT	46	73	-3	-42	69	0
pIPS	16	73	47	-19	66	42
aIPS	36	45	46	-32	47	43
IFJ	52	6	40	-51	10	40
FEF	27	11	54	-26	11	52

Univariate Results

A univariate GLM was used to identify regions in the brain that showed a parametric change in signal intensity with memory load during the delay-period for each of the two task conditions (direction and color) and for each remembered visual field (left and right). Group results from the four condition combinations can be seen in Figure 6.3. For direction trials, the results were very similar across visual fields. Significant load-sensitive increases in activation were observed in bilateral posterior intraparietal sulcus (pIPS), anterior intraparietal sulcus (aIPS), IFJ, superior precentral sulcus (sPCS), and right insula. For color trials, significant load-sensitive increases in activation were observed in bilateral pIPS, left IFJ, left inferior frontal gyrus (IFG), and left middle temporal gyrus across the visual field conditions, and in bilateral striate and peristriate cortex in the remember-left condition.

The results from the univariate GLM were used to identify seed regions for the PPI analysis on a subject by subject basis. The unthresholded statistical maps for each participant from the load-sensitive regressors for the four conditions of interest were combined to isolate regions that showed increased activity with memory load across the conditions. From this combined map 8 seed locations were identified: bilateral pIPS, aIPS, IFJ, and sPCS (See table 6.1). An additional 4 seeds were created from the results of the motion localizer: bilateral V1 and MT.

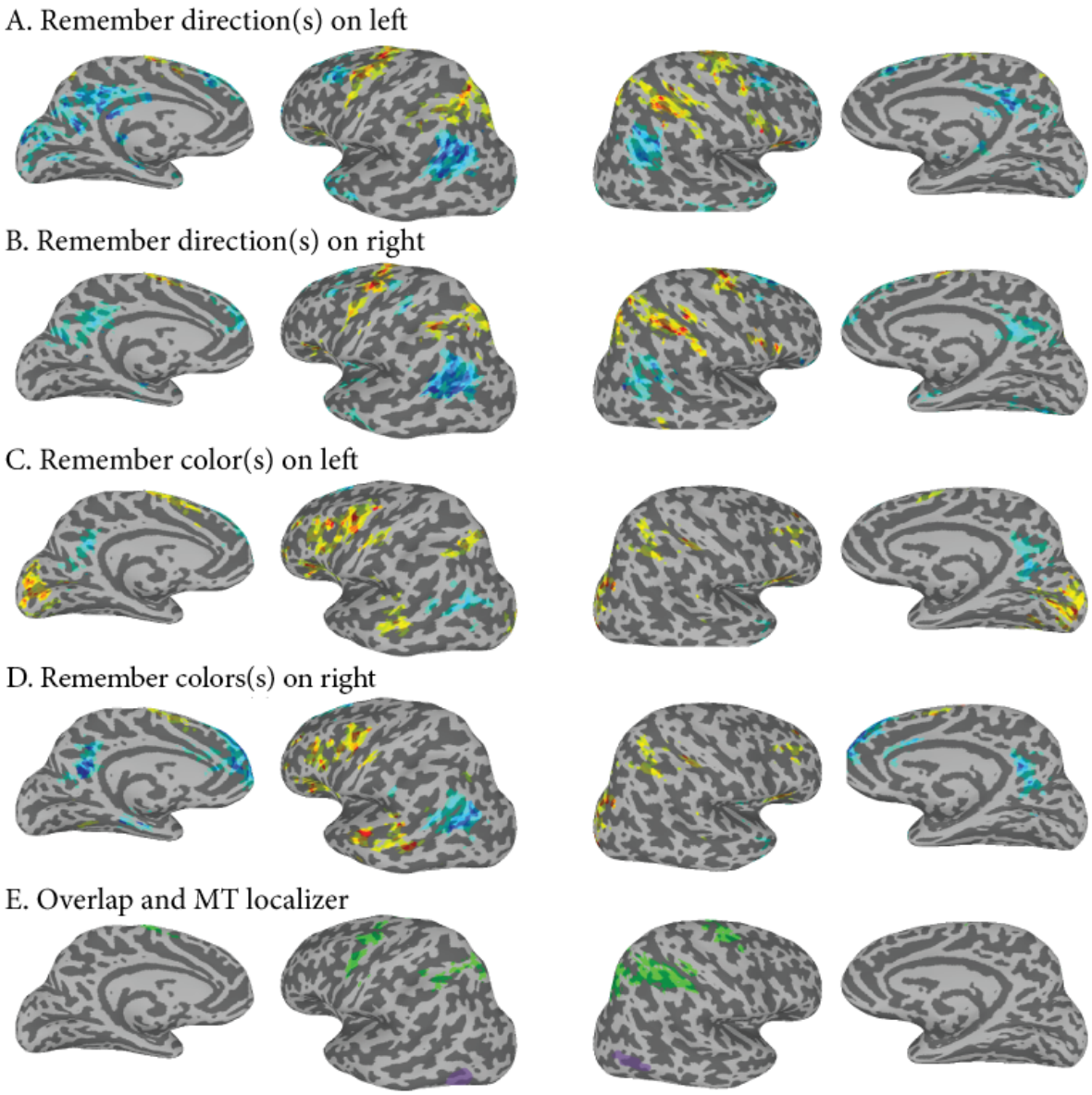

Figure 6.3: Group univariate GLM results. Regions showing load-depending changes in delay-period activity on (*A*) direction trials while remembering directions on the left, (*B*) direction trials while remembering on the right, (*C*) color trials while remembering on the left, and (*D*) color trials on the right. Warm colors are regions with more higher activity on load 3 compared to load 1, cool colors the reverse. (*E*) regions showing overlap across the four conditions (green) and MT as identified by the localizer (purple). Group results are for illustrative purposes, as individual seed regions were chosen based on individual subject GLM results. All shown on inflated cortical surfaces, from left to right: medial left hemisphere, lateral left hemisphere, later right hemisphere, medial right hemisphere.

Functional Connectivity Results

In order to identify functionally connected brain regions involved in the performance of the task, we conducted a PPI analysis using the seed regions identified in the univariate GLM. For a given condition of interest and seed region, we refit the GLM with 3 additional regressors: the seed timeseries, the condition regressor, and the interaction between the seed and condition. This allowed us to generate maps of overall seed-based connectivity as well as to identify those regions where information in the seed area provided predictive information about other areas above and beyond what could be explained by shared task effects.

We began by looking at interactions between seed regions and a condition regressor representing the delay period of direction tasks against color tasks. The interaction regressor in this model should identify those voxels whose timeseries are more correlated with the seed voxel's timeseries during direction trials compared to color trials (or vice versa). Given the relative variability in the location of seed regions across participants, we opted to extract the coefficients from the individual fits from the subject specific seed regions and perform statistical tests on these seed specific values rather than try to observe the results in template space group analysis. Results from the interaction regressor based on task condition can be seen in Figure 6.4B. In this matrix, rows represent the t-values of the group interaction effect from all seed regions for a specific seed timeseries. For example, the first row shows the extent to which the activity in each of the seed regions is more or less correlated with the timeseries from left V1 on direction trials versus color trials. Red values indicated regions where the correlation is higher between the region and the respective seed during direction trials compared to color trials, blue the opposite. These results show that on the whole, most of seed regions are more correlated during direction trials compared to color trials.

Given the large differences in connectivity between tasks, we next split the data by task to look for more subtle changes in connectivity with task demands. Figure 6.4C and D show the results for each of the two task conditions for a condition regressor representing the delay period of load 3 trials against load 1 trials. Red values indicate regions where the correlation with the seed region was higher on load 3 trials compared to load 1 trials, blue the opposite. Comparing across the two

tasks we see quite different results. On direction trials the correlations between the visual regions (V1 and MT), including those in the opposite hemisphere, were stronger on load 3 with the other visual regions, but lower for all other regions, suggesting a coordination of activity at higher loads that was not driven by any of the other regions. On color trials the pattern was quite different. The activity between the parietal and frontal regions was generally more correlated during load 3, and the visual regions showed little change. Particularly strong connectivity was observed between right aIPS and right IFJ

TMS/EEG Session

Behavioral Results

Five participants from the fMRI session returned for the TMS/EEG session. They performed the same delayed-recall task for visual motion or color during the TMS/EEG session, while TMS was applied to MT, IPS, and PFC in separate blocks. As in the fMRI session, participants were significantly more likely to fail to respond within the 4 s response window on color trials compared to direction trials (Fig. 6.5; $F_{(2,8)} = 8.93, p = 0.04$), however this did not vary across TMS targets ($F_{(2,8)} = 0.74, p > 0.50$) and there was no interaction ($F_{(2,8)} = 0.10, p > 0.90$). Response times showed a similar pattern (Fig. 6.5; condition: $F_{(2,8)} = 70.36, p = 0.001$, TMS target: $F_{(2,8)} = 0.80$, $p = 0.48$, interaction: $F_{(2,8)} = 1.19, p = 0.35$).

We again examined the mixture-model precision estimates using a two-way repeated measures ANOVA with load (1 or 3) and hemifield (left or right) as factors separately for each TMS site and task condition to investigate the impact of TMS and hemifield (Fig. 6.5). Across the conditions and site, only the load difference was significant on direction trials for IPS stimulation ($F_{(1,4)} = 11.9$, $p = 0.026$) and PFC stimulation ($F_{(1,4)} = 17.4, p = 0.014$). All others tests were not significant.

We next collapsed responses across hemifields so we could further analyze performance. We refit the mixture-models with the collapsed data and extracted the proportion of target responses for each of the TMS stimulation sites (Fig. 6.5). As in the fMRI session, no differences were found between loads or conditions across all of the TMS sites (all $p > 0.05$).

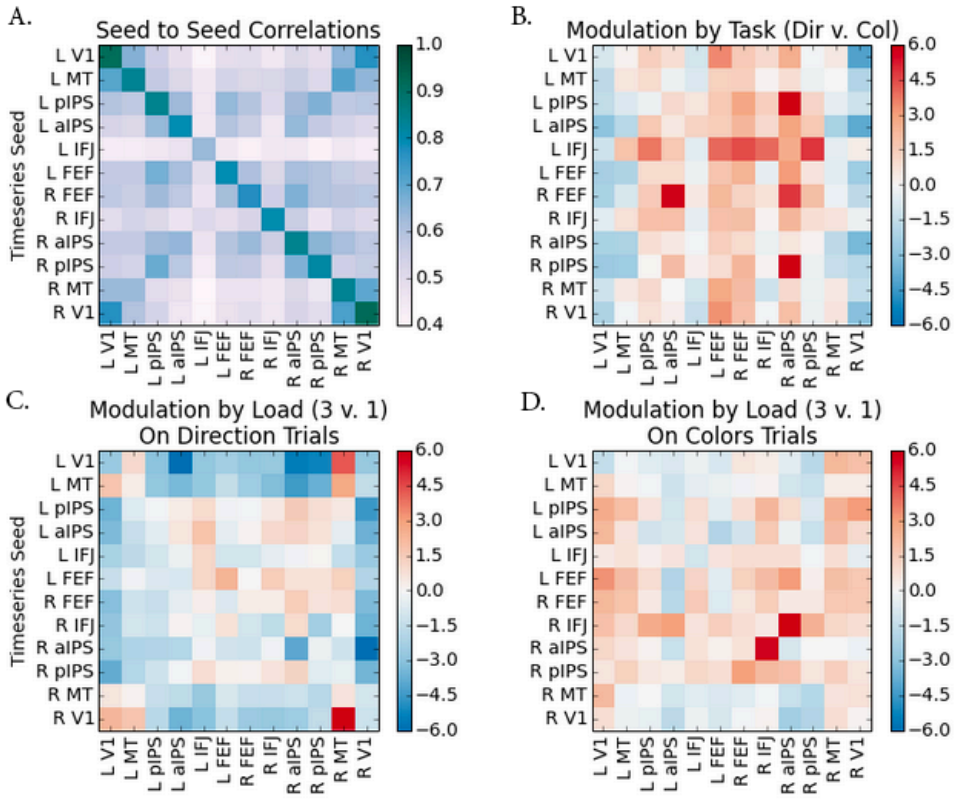

Figure 6.4: Results from fMRI function connectivity analysis. (*A*) seed-based correlations between specific seed timeseries and other seed regions. Each row represents the results from correlations between a single seed timeseries and all other regions. Colors indicate average *r*-values across participants. (*B*) Modulation of the correlation during the delay period between between seed timeseries and all other regions on direction trials against color trials. Red areas indicate regions where the correlation to the specific seed timeseries (identified by row) was higher during direction trials compared to color trials, blue areas the opposite. Colors indicate *t*-values comparing the PPI interaction regressor coefficients across participants. (*C-D*) Modulations of the correlation during the delay period on load 3 trials against load 1 trials, for direction (*C*) and color (*D*) trials separately. Red areas indicate higher correlations during load 3 compared to load 1 trials, blue the opposite. Colorbar same as *C*.

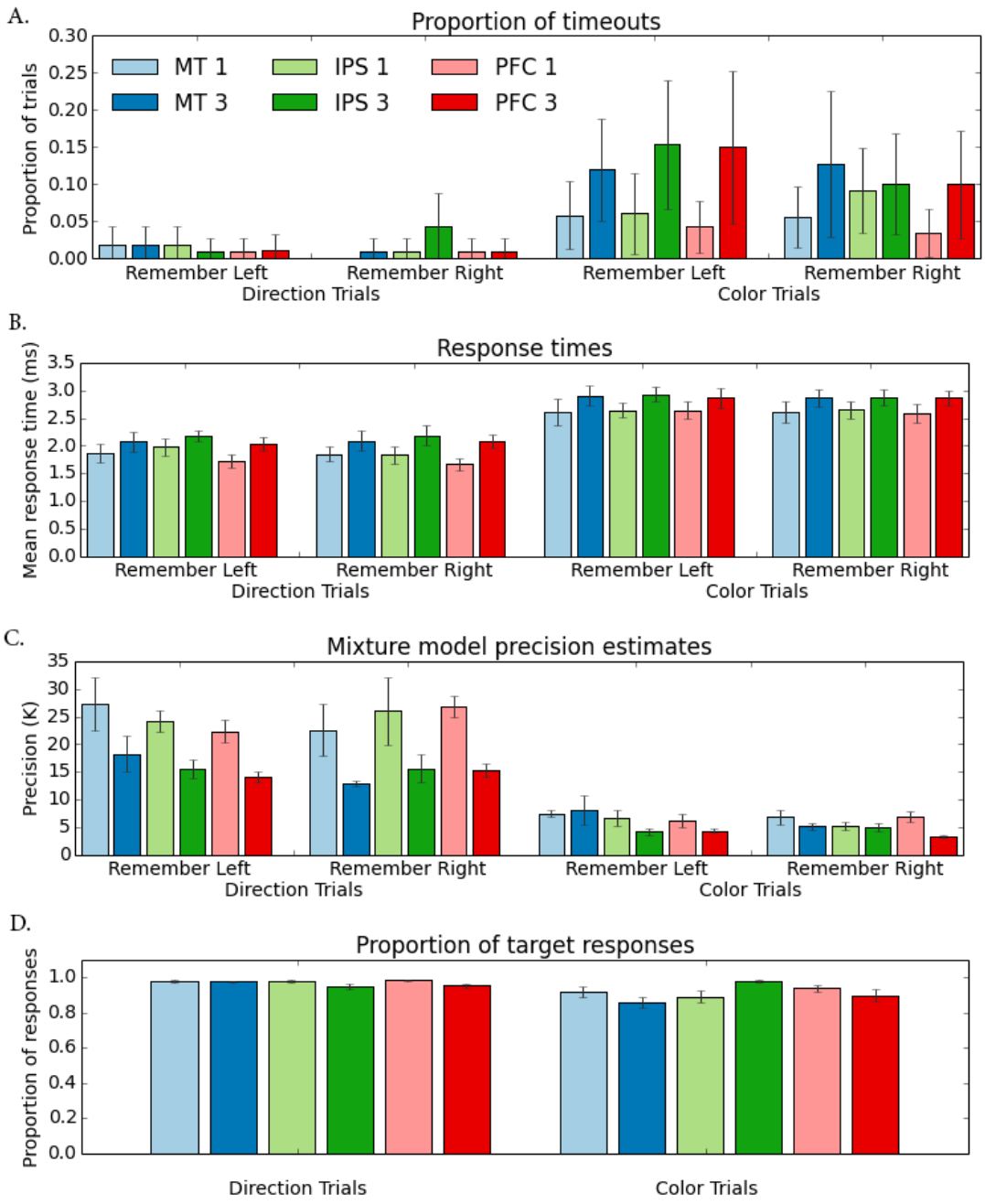

Figure 6.5: Behavioral results from TMS/EEG session. (*A*) Mean proportion of trials which timed out after participants (n=5) did not respond within the 4 s response window, which occurred more often of color trials compared to directions trials. (*B*) Mean response times on trials with a validly recorded response, which were longer for color trials than direction trials. (*C*) Estimates of the precision (κ) of responses from the mixture model, which showed a significant effect of load for both task conditions. (*D*) Estimates of the proportion of responses (P_T) to the target from the mixture model fits, which showed no differences across conditions or loads. Colors correspond to sites of TMS stimulation: blue for MT, green for IPS, and red for PFC. Lighter colored bars are for load 1 trials, darker bars for load 3.

Effective Connectivity Results

To explore the influence of task condition on the connectivity between frontoparietal regions showing load-sensitive activity and posterior visual regions we analysed the TMS-ER using the method developed by Casali et al. (2010). This method computes a series of synthetic measures characterizing the strength and spread of the brains response to TMS stimulation in source space. Figure 6.6 shows an illustration of these measures for one example participant.

To assess the impact of task conditions on effective connectivity, we computed the SCD and SCS for each combination of task condition, remembered visual field, memory load separately and TMS stimulation site separately (Fig. 6.7). We then entered the results from combinations of TMS site and task condition into two-way repeated measures ANOVAs to look for effects of memory load and hemifield.

For IFJ stimulation, looking at SCD, we observed a significant main effect for load for both task conditions (direction: $F_{(1,4)} = 11.6, p = 0.03$, color: $F_{(1,4)} = 15.3, p = 0.02$), no effect of visual field (direction: $F_{(1,4)} = 0.09, p = 0.78$, color: $F_{(1,4)} = 0.01, p = 0.93$) and no interaction (direction: $F_{(1,4)} = 2.93, p = 0.16$, color: $F_{(1,4)} = 2.94, p = 0.16$) for either condition. Looking at SCS we observed a slightly different pattern, with the main effect of load only significant for color trials, (direction: $F_{(1,4)} = 2.8, p = 0.17$, color: $F_{(1,4)} = 8.4, p = 0.045$), and no effect of visual field (direction: $F_{(1,4)} = 1.5, p = 0.29$, color: $F_{(1,4)} = 1.1, p = 0.45$) or their interaction (direction: $F_{(1,4)} = 3.1, p = 0.15$, color: $F_{(1,4)} = 1.7, p = 0.26$) for either condition.

With IPS stimulation we observed different patterns for the two tasks. Looking at SCD, on direction trials we observed a significant main effect of load ($F_{(1,4)} = 9.4, p = 0.04$), a significant main effect of remembered hemifield, ($F_{(1,4)} = 11.6, p = 0.03$), and no interaction effect ($F_{(1,4)} = 0.33, p = 0.60$). For colors trials we did not observe any significant results (load: $F_{(1,4)} = 0.64$, $p = 0.47$, hemifield: $F_{(1,4)} = 0.79, p = 0.42$, interaction: $F_{(1,4)} = 0.91, p = 0.39$). For SCS, the pattern of results was again slightly different. For direction trials a trend towards a significant effect of load ($F_{(1,4)} = 7.4, p = 0.06$), no effect of hemifield, ($F_{(1,4)} = 2.1, p = 0.22$), and no interaction ($F_{(1,4)} = 0.98, p = 0.38$). For colors trials we again did not observe any significant

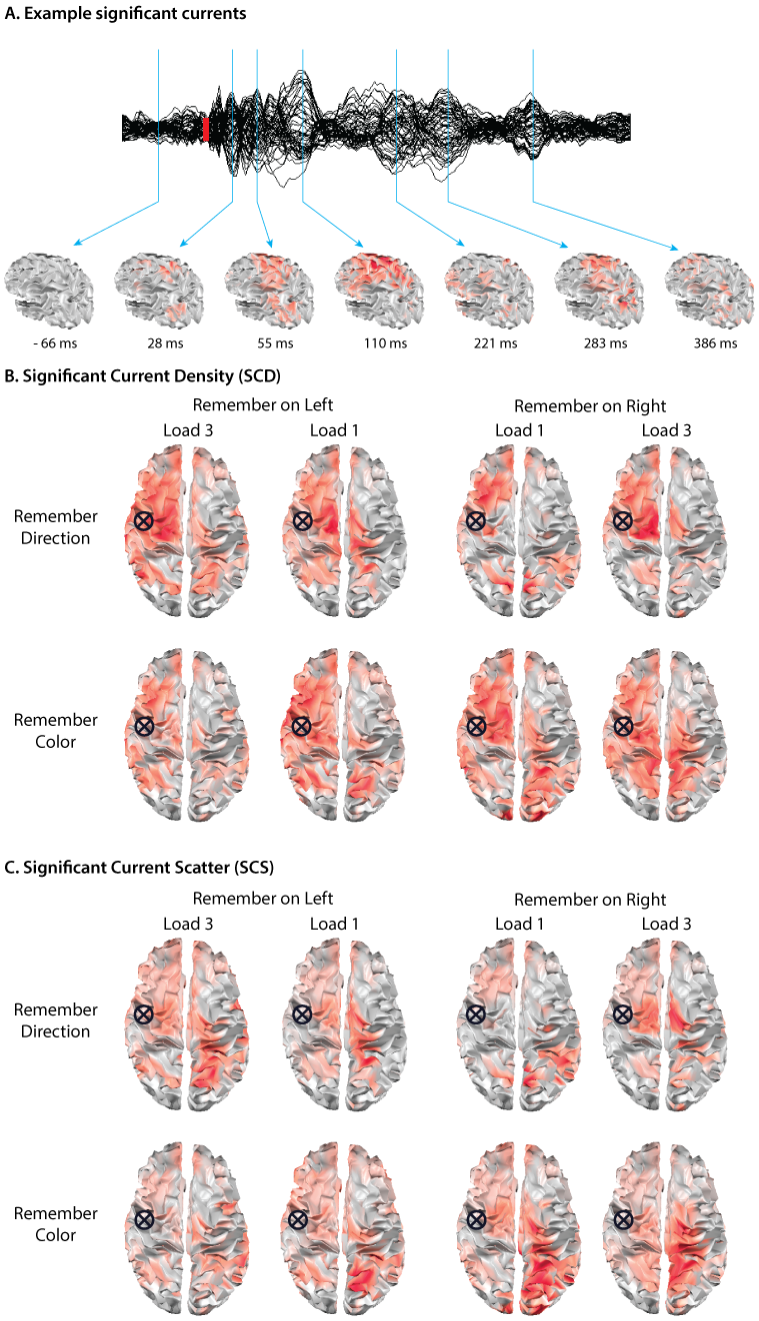

Figure 6.6: Example TMS-evoked results from a representative subject. (*A*) Top: Average TMS-ER responses for each channel after stimulation of IFJ. Bottom: Cortical currents estimated at individual time points, with increasingly strong significant currents shown in darker reds. Cortical surface is shown as though viewed from behind and to the left, such that the frontal pole is up and to the left, occipital pole on the lower right. (*B*) Average SCD computed for the evoked potentials in *A* across an interval from 40ms-400ms post-TMS. (*C*) Average SCS for the same data over the same interval. Cortical surfaces shown from directly above, with the frontal pole towards the top of the page, occipital pole towards the bottom. Location of TMS stimulation indicated by the circle and cross.

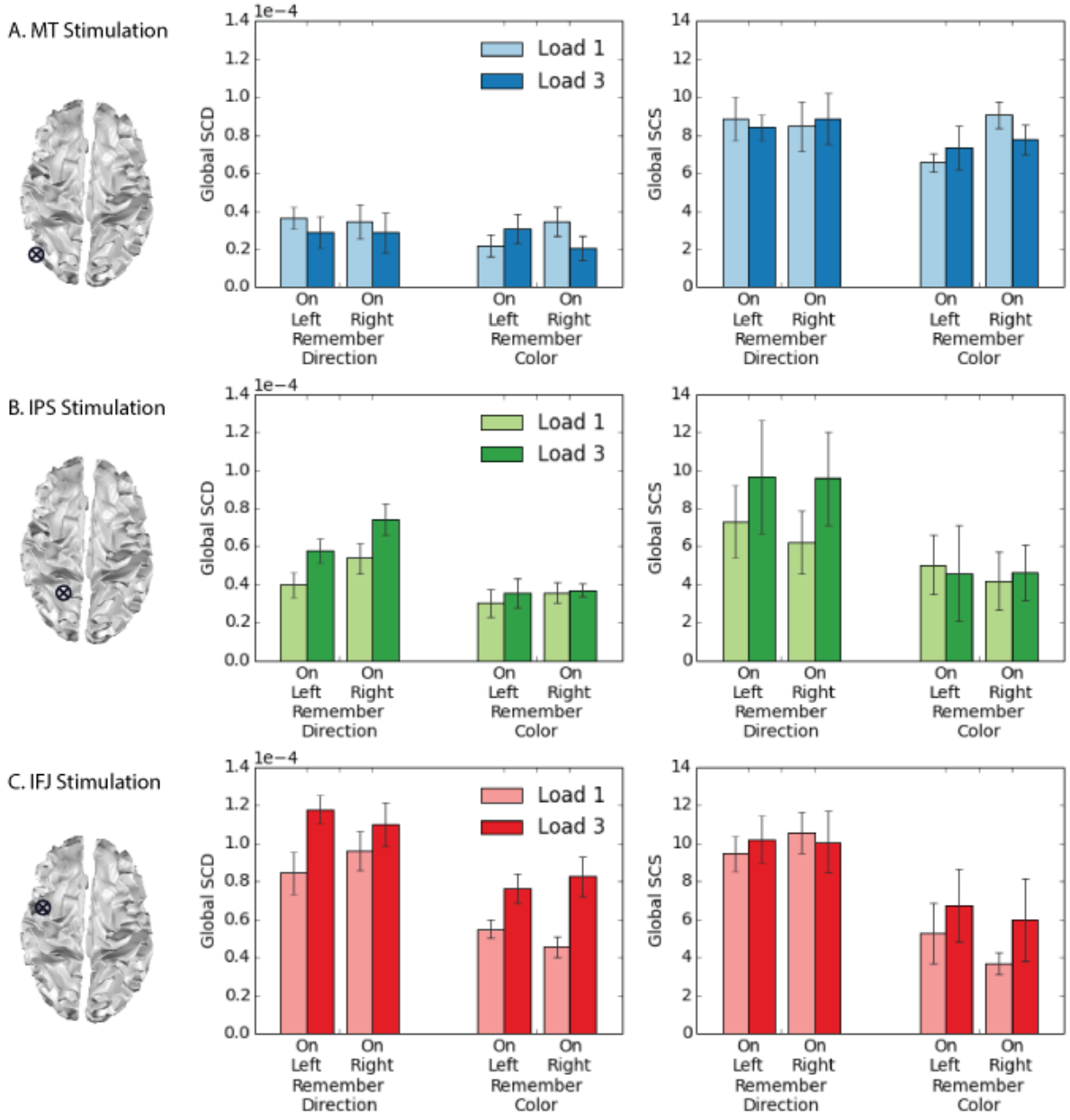

Figure 6.7: Global TMS-evoked response synthetic measures following delay-period stimulation to MT (*A*), pIPS (*B*), or IFJ (*C*). Left-most column shows cortical surfaces directly from above with location of TMS stimulation indicated by circle and cross. Middle column shows SCD results. Left column shows SCS results.

results (load: $F_{(1,4)} = 1.8, p = 0.25$, hemifield: $F_{(1,4)} = 1.6, p = 0.27$, interaction: $F_{(1,4)} = 2.2, p = 0.21$)

Following stimulation to MT we did not observe any significant effects for either task condition for either measure (All Fs $< 2.5, p > 0.18$).

6.4 Discussion

In the current work we aimed to explore the connectivity between frontoparietal regions that show load-dependent delay-period activity and visual regions during visual short-term memory. Participants performed a delayed-recall task requiring them to remember either the motion direction or color of moving dot stimuli (loads 1 or 3). We collected fMRI data during an initial session, and in a second session applied TMS to several regions functionally localized with the fMRI data while collecting EEG. We aimed to test the hypothesis that this load-dependent activity represents processes involved in allocating attention towards internal representations that are maintained in the same neural systems used for sensory processing (Lückmann et al., 2014). We predicted that these frontoparietal regions should show higher connectivity with posterior visual regions under conditions that should require greater internal attention, i.e., higher memory loads.

Using a univariate GLM we identified robust load-sensitive delay-period BOLD activation in frontal (sPCS and IFJ) and parietal cortex (pIPS) across both direction and color trials, in accordance with other results for direction (Kawasaki et al., 2008; Riggall and Postle, 2012; Emrich, Riggall et al., 2013) and color Todd and Marois (2004); Xu and Chun (2006); Kawasaki et al. (2008). We used seeds placed in these regions as well as locations identified in an independent motion localizer to assess connectivity during the delay-period of the task using a PPI approach (Friston et al., 1997). We observed large differences in the amount to which the specific task being performed (remember direction or remember color) impacted the interaction effect between most seed regions. Looking only at direction trials, correlations increased with increased load between visual regions (V1 and MT), and decreased between parietal and frontal seeds and V1 at higher loads.

On color trials, the pattern was different, with an increase in load leading to higher correlations between seeds and other brain regions for most frontoparietal sites. These results are at somewhat at odds with other recent work looking at frontoparietal connectivity during short-term memory, which have tended to find consistent connectivity, particularly between IFJ and IPS during short-term memory (Zanto et al., 2010; Majerus et al., 2010). Possible explanations for this difference will be discussed below.

Using the results of the fMRI, we identified target locations for TMS (MT, pIPS, and IFJ) and stimulated during the delay-period of the task while collecting EEG. Analysis of the TMS-evoked response provides a direct a measure of network connectivity (Rogasch and Fitzgerald, 2012), and has been used to show differences in connectivity during different states of consciousness (Ferrarelli et al., 2010; Casali et al., 2013) as well as between different task conditions (Johnson et al., 2012). Using the synthetic measures developed by Casali et al. (2010), which quantify the strength and spread of the TMS-ER, we found several notable results. For stimulation to IFJ, SCD, the strength of significantly induced currents, was greater in load 3 compared to load 1 for both direction and color trials, in agreement with our prediction that higher loads require more top-down control. Interestingly, SCS, the scatter of the response, did not differ for direction trials, but was larger for load 3 for colors trials, suggesting a recruitment of more neural resources for higher loads on the behaviorally more difficult color trials. For pIPS stimulation we observed a significant increase with load for both strength and scatter of the TMS-ER for direction trials, but no effect for color trials. Finally, we observed no significants effects on the strength or scatter of the TMS-ER after stimulation to MT. These results are consistent with a network where top-down control is initiated from PFC in coordination with IPS during the maintenance.

Overall our results are similar to those of previous groups. Sneve et al. (2013) have used Granger causality analyses to show top-down activation from IFJ and frontal eye field (FEF) to IPS and higher-order visual areas with a delayed orientation discrimination. Majerus and colleagues (Majerus et al., 2006, 2007, 2010) have shown similar connectivity between IFJ, IPS and posterior regions during short-term memory for verbal information, for faces, and for temporal

order information. Finally, Zanto et al. (2010) have shown IFJ to be the critical source of top-down control during a short-term memory task for colors and motion.

Given these other results, our finding that stimulation of pIPS only had a load-dependent effect on the TMS response during direction trials is surprising. One possible explanation, though only speculative, could be that this result is related to differences in maintenance strategy across the two task conditions. From the behavioral performance it is clear that the difficultly of the two conditions was not well matched, with color trials being much more difficult for all trials, as measured by the precision of responses and overall response times. Given the particularities of the task, one might conjecture that direction maintenance is partly supported by spatial attention processes, while color maintenance is partly supported by verbalization and articulatory rehearsal. As no measures were taken to control either of these, beyond the requirement of overt fixation controls, we can again, only speculate. But given recent results identifying these same frontoparietal regions as also containing "priority maps" of space (Jerde et al., 2012), it's possible that TMS to IPS interacted with this process when it was being used to support direction representation. Taking this speculative path to it's full conclusion, we could hypothesize that TMS to a region involved in verbal articulation, such as IFG, which was active only during our color task, would show load sensitive effects in the TMS response during color trials, but not direction trials.

Taken together, these results support the theory that frontoparietal regions are involved in the control of top-down attention towards internal mental representations maintained in the sensory regions that initially stored them Postle (2006); Lepsien and Nobre (2007); Lückmann et al. (2014). Within this framework, many questions still remain. What are the specific roles of individual subregions within the network (e.g., IFJ vs. sPCS vs. aIPS vs. pIPS)? What parts of this network are necessary and sufficient for short-term memory storage? Recent work in the monkey suggests that sPCS may not be, as muscimol inactivation of this region, which showed reliable sustained responses during the delay-period, produced no impact on task performance for an object-based STM task (Clark et al., 2014). The evidence from human studies are mixed, some have shown decreased STM memory performance after disruptions to PFC with TMS (Zanto et al., 2011; Lee and

D'Esposito, 2012), while other have shown no effect, or even improvements, after TMS (Hamidi et al., 2008, 2009). A final question relates to the extent to which the processing of these regions is specialized by hemisphere versus being flexibly bilateral. A number of results seem to point to hemispheric specialization depending on the particular information being attended to (Majerus et al., 2006, 2007, 2010; Zanto et al., 2010), but recent work using theta-burst TMS to disrupt PFC before an STM task showed this may be more flexible, as individual differences in the amount of compensatory activity seen in homologous area predicted differences in the impact that TMS had on behavioral performance.

In closing, our results help to clarify the network of brain regions involved in the short-term retention of visual motion information. We have previously shown posterior visual areas, encompassing striate and extrastriate context (including MT), to be the site of information storage. In the current work we have shown that the reliable frontoparietal activity observed during this retention is a marker of top-down attentional mechanisms supporting the representations in posterior regions.

7

Conclusions

The present work describes a series of experiments that were conducted to investigate the neural basis of visual short-term memory. Questions such as how information is maintained in the brain and what specific neural processes support this retention, have long tantalized researchers. Only relatively recently have we developed the necessary methods to directly measure the information content in regions of the brain during short-term memory. In particular, the development of multi-voxel pattern analysis (MVPA) approaches applied to blood-oxygen level-dependent (BOLD) functional magnetic resonance imaging (fMRI) data (Haxby et al., 2001; Haynes and Rees, 2006; Norman et al., 2006; Naselaris et al., 2011; Tong and Pratte, 2012), have allowed us to directly measure information-content, rather than overall activity level, within brain regions.

With these tools in hand, we set out to investigate the neural processes supporting the active retention of visual motion information. The results from this work provide strong support to the growing consensus that short-term storage does not rely on specialize storage subsystems, as long believed by supported of the multicomponent model of working memory first proposed by Baddeley and Hitch (1974). Rather, it appears that memory maintenance is supported by the same neural systems specialized for the initial processing of visual information (*the sensory recruitment hypothesis*; Awh and Jonides, 2001; Pasternak and Greenlee, 2005; Jonides et al., 2005; Postle, 2006; D'Esposito, 2007; Serences et al., 2009).

7.1 Summary of Findings

The aim of the first study, discussed in Chapter 2, was to directly test the relationship between elevated delay-period activity and information storage during short-term memory. The importance of regions demonstrating sustained, elevated delay-period activity during short-term memory for memory storage has been inferred from the reliable presence of these regions across a wide variety of short-term memory tasks (for a review see Curtis and D'Esposito, 2003). The particular assumption of a role for these regions in storage has been reinforced by their sensitivity to manipulations of memory-influencing factors, such as persistence across varying delay lengths, and variation of signal magnitude with memory load (Vogel and Machizawa, 2004; Xu and Chun, 2006; Postle, 2006). Given this framework, we hypothesized that if these regions are the locus of storage, application of MVPA to data from these regions during the memory delay should succeed in identifying stimulus-specific patterns of activity in these regions.

We scanned participants (fMRI) while they performed a delayed-recognition task for simple dot motion stimuli that required them to remember both the direction and speed of the dot motion stimulus over a long delay period, then make a match/non-match decision about either the direction of motion or the speed of motion of a probe dot motion stimulus. The dimension to be probed was indicated halfway through the delay-period. To test the hypothesis that elevated, persistent delay-period activity is a mark of information storage, we trained pattern classifiers to identify the direction of motion in the sample motion stimulus, and by inference, in memory (*stimulus-specific decoding*) using data from frontal and parietal regions showing sustained, elevated activity during the memory delay. Using this approach we failed to evidence for stimulus-specific representation at any timepoint during the trial in these frontal and parietal regions.

Next we tested an alternative hypothesis, that stimulus-specific information could be encoded in distributed patterns of substhreshold activity. Such decoding had been previously demonstrated in posterior visual regions for orientation information (Harrison and Tong, 2009; Serences et al., 2009). To test this hypothesis we retrained the classifiers on data from posterior regions that showed a large initial sample-evoked response, but no evidence for sustained activity during the

delay-period. These classifiers performed well, reliably decoding the individual direction in memory on a trial-by-trial basis. Using a temporal cross-generalization procedure (King and Dehaene, 2014), which involved training the classifier with data from a single timepoint during the delay and then testing it on all the timepoints from a held-out test trial, we showed successful decoding throughout the entire delay-period. Critically, this decoding was only successful in the later part of the delay on trials where direction had been cued as important. That is, the memory representation was task specific. When participants were cued that they would be making a comparison about speed, evidence for the direction of motion vanished.

To eliminate the possibility that there is something about these persistently active delay-regions that precludes applying MVPA methods to them, we showed successful decoding of the task-context (direction or speed to be probed) in all regions, but only after the cue had been presented midway through the memory delay.

Thus we concluded from the first study that it was indeed possible to decode specific directions of motion held in short-term memory using MVPA, but only from posterior visual regions, not frontoparietal regions that showed elevated delay-period activity. These decodable representations were task-specific and stable throughout the delay.

For the second study (Chapter 3), we performed a follow-up fMRI study examining the impact of remembering multiple motion directions on each trial. Several studies using univariate approaches with BOLD data have suggested that regions that show a load-sensitive modulation in delay-period activity, particularly locations where that activity modulation saturates at an individual's memory capacity, must be the location of storage (Todd and Marois, 2004, 2005; Xu and Chun, 2006). We hypothesized that if this is true, then we should be able to decode stimulus-specific representations from regions showing load-sensitive delay-period activity. We tested this hypothesis with a similar approach to the first study, training classifiers to decode the direction of motion in memory. As in the first study, we found no evidence for stimulus representation in the previously-hypothesized storage activity. Rather, again in line with the results from the first study, we only found evidence for stimulus-specific representations in striate and extrastriate

visual regions.

Having identified regions containing memory representations during multi-item memory, we explored the extent to which these representations where dependent on the trial-to-trial memory load. We hypothesized that if short-term memory representations rely on a shared-memory resource, as has been proposed by several groups (Wilken and Ma, 2004; Bays and Husain, 2008; Bays et al., 2009; van den Berg et al., 2012), then the evidence for a specific representation, as observed with MVPA, should also depend on load, with decreasing evidence under higher load as the resources available for any given representation are reduced. To test this hypothesis we compared the classifier evidence for trials at each load, finding a decrease in classifier evidence with an increase in memory load.

Our third study (Chapter 4) was another follow-up fMRI study investigating the locus of storage during short-term memory. In part, this study was designed to address two limitations in the design of the second study: the short delay length and the limited number of responses required. In the second study we used a relatively short (for fMRI) delay period (8 s). While this shorter delay allowed us to acquire more trials and thus more training exemplars for our classification approach, it prevented us from being able to fully examine changes in representations throughout the delay period. To improve upon this, we switched back to a longer delay length (12 s). We had also asked participants to recall the target motion directions we were planning to use to classify on 90% of the trials. Thus it is possible that participants learned that only three directions were ever probed and may have started to implicitly ignore non-target directions (as they were rarely probed) on load 2 and 3 trials. We don't believe this was actually the case, given that the estimates of participant's memory precision decreased with increases in load and were comparable to the results from a study with a similar design and completely random probe selection (Zokaei et al., 2011). Nonetheless, to remove this as a possible confound, we altered the task so that the probed item was evenly selected from all the options on a trial, and introduced more trials without any of the target directions to further limit the ability of participants to perform the task without maintaining all the items in the sample. Finally, we made two addition changes to the task to test that

our results would generalize to slightly different task contexts. We switched from sequential presentation of multiple motion directions to simultaneous presentation using transparent motion stimuli. This type of stimulus, which consists of two or more directions presented within the same visual aperture at the the same time, produces the perception of independent, semitransparent surfaces moving in different directions. This change eliminated the temporal and spatial cues that can be used to help differentiate the individual items, presenting a much more challenging problem for the visual system. Finally we changed the response from a delayed-recall response to a delayed-discrimination, which required participants to indicate which direction a probe motion stimulus had been rotated relative to one of the remembered stimuli. This provided two benefits over delayed-recall, allowing us to provide immediate feedback on the performance on each trial while also providing a measure of the precision of representation at each memory load.

Participants again performed this task in the scanner (fMRI) and we analyzed that data with a temporal cross-generalization MVPA approach. Based on our earlier two results, we hypothesized that individual directions of motion would be recoverable from patterns of activity in posterior visual regions, rather than in regions showing load-sensitive or elevated delay-period activity. Results from the decoding analysis bore out this prediction, as we were only able to find evidence for memory representations in posterior visual regions. We also hypothesized, based on our initial results with one item memory, that the representation would remain stable throughout the memory delay. The results showed a different pattern, with a clear initial represenation shortly after the sample presentation followed by a different representation developing later during the delay period, strongly suggestive of a recoding process happening after the initial sample presentation during the delay. Importantly, this recoding was occurring entirely within the same visual regions; it did not involve a transfer of information to any another regions.

Given our repeated successful detection of stimulus-specific patterns in BOLD activity during short-term memory in posterior visual areas, a natural next question was what, exactly, underlies these patterns to allow the successful decoding of items in memory. For this we had to look "under the hood" of our classification algorithm and inspect which voxels it was relying on in or-

der to perform successful classification. A first glance at the most informative voxels suggested a seemingly contradictory result to what we had observed in the decoding accuracies: the most informative voxels during classification appeared to vary greatly over the trial, and extended into parietal and frontal areas during wholebrain decoding, completely at odds with our initial results suggesting a temporally stable representation confined to posterior visual cortex.

Chapter 5 looked in detail at the specifics of exactly what it was that the pattern classifiers were using to identify and separate specific directions of motion in memory. With this analysis we hoped to address two separate points: the spatial distribution of voxels involved in these representations and the temporal stability of the representations. To assess the spatial distribution of the representations we used two approaches: recursive feature eliminination (RFE) and noise-perturbation sensitivity (NPS), which each attempt to identify the informativeness of each voxel towards the overall classification performance. These two methods provided very similar results, showing that voxels outside of striate and extrastriate cortical areas did not contain information useful to the classifier for predicting the memory representations. All the informative voxels were located within these posterior visual regions. We looked at the temporal stability of the representation with several complementary approaches. These results provided mixed results, with some voxels being informative about the remembered item at all timepoints during the delay while others were only informative intermittently.

Together, the first four studies provide compelling evidence that the elevated activity observed in frontoparietal regions in all of these studies does not identify these regions as the location of information storage. This leads to the obvious question of what, then, does this activity represent? Why are these regions consistently activated during short-term memory?

The final study (Chapter 6) sought to directly address this question of what exactly these delay-active regions are doing. We hypothesized that these regions coordinate the recruitment of sensory resources for storage and protect remembered information from disruption through the direction of top-down attention to the remembered information. This hypothesis suggests that these elevated delay-period regions should be functionally connected to the posterior visual regions in-

volved in the actual maintenance of information. As the memory demands increase, (e.g., when multiple items need to be remembered) more resources would be needed to support the representations and protect them from interference, leading to higher connectivity between these regions.

To test this hypothesis, we had participants perform a delayed-recall task that required them, on separate blocks, to remember either the directions of moving dot stimuli or the colors that the moving dot stimuli were presented in. We hypothesized that if this elevated delay-period activity represents a global attentional mechanism in the service of memory, it should be similarly active in both conditions, but there should be qualitative differences in the pattern of connectivity between these regions and the posterior visual regions involved in storing the particular features relevant to the two different tasks. Participants initially performed this task in the scanner (fMRI), allowing us to identify common regions of elevated activity across the tasks and perform functional connectivity analyses. The results from these analyses identified several regions in frontoparietal cortex that were active across both tasks. The results from the connectivity analyses from these overlapping regions showed quite different patterns of connectivity between the tasks.

We next tested participants with the same task (for directions and colors), while recording electroencephalography (EEG) and stimulating with single-pulse transcranial magnetic stimulation (TMS) during the delay period of the task. We targeted TMS to three different locations in each participant: the middle temporal area (MT), a posterior visual region primarily involved in the processing of motion; posterior intraparietal sulcus (pIPS), a location in the parietal cortex showing elevated activity during the delay-period of both tasks; and inferior frontal junction (IFJ), a region in the lateral prefrontal cortex (lPFC) that was also active for both conditions. We hypothesized that TMS to the frontal and parietal sites should evoke a stronger TMS-evoked response (TMS-ER) at higher memory loads, as these regions would be involved in providing the top-down attentional signals necessary for maintenance of higher memory loads. TMS to MT, on the other hand, should not differ by load. The results from these analyses, which relied on the synthetic measures developed by Casali et al. (2010), were consistent with our hypotheses. In general, significant current scattering (SCS) and significant current density (SCD) of the TMS-ER

was larger on load 3 trials compared to load 1 trials after stimulation to pIPS or IFJ. There was no load effect after stimulation to MT.

7.2 Concluding Remarks

Taken together, this collection of studies provides strong evidence pointing to the role of specialized visual processing regions in the maintenance of visual information during short-term memory. By decoding *stimulus-specific* patterns of BOLD activity during the memory delay of several short-term memory paradigms, we have provided considerable additional support to the *sensory recruitment hypothesis* (Awh and Jonides, 2001; Pasternak and Greenlee, 2005; Jonides et al., 2005; Postle, 2006; D'Esposito, 2007; Serences et al., 2009; Sreenivasan et al., 2014a). Given our findings and the recent findings of many other groups (Harrison and Tong, 2009; Serences et al., 2009; Ester et al., 2009; Albers et al., 2013; Christophel et al., 2012; Christophel and Haynes, 2014; Vicente-Grabovetsky et al., 2014; Linden et al., 2012), it seems without doubt that posterior visual "sensory" regions are recruited in the service of memory storage during short-term memory.

Our results also provide supporting evidence for an alternative explanation of what functions the frontoparietal areas that show elevated activity during the delay-period are performing, if not storage. We argue that these regions play a critical role in controlling the top-down allocation of attention towards the internal representations stored in specialized visual regions (Postle, 2006; Lepsien and Nobre, 2007; Lückmann et al., 2014).

Bibliography

Albers AM, Kok P, Toni I, Dijkerman HC, Lange FP (2013) Shared representations for working memory and mental imagery in early visual cortex. Curr Biol 23:1427–1431.

Alvarez GA, Cavanagh P (2004) The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychol Sci 15:106–111.

Anderson DE, Vogel EK, Awh E (2011) Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. J Neurosci 31:1128–1138.

Armstrong KM, Chang MH, Moore T (2009) Selection and maintenance of spatial information by frontal eye field neurons. J Neurosci 29:15621–15629.

Awh E, Barton B, Vogel EK (2007) Visual working memory represents a fixed number of items regardless of complexity. Psychol Sci 18:622–628.

Awh E, Jonides J (2001) Overlapping mechanisms of attention and spatial working memory. Trends Cogn Sci 5:119–126.

Axmacher N, Mormann F, Fernández G, Elger CE, Fell J (2006) Memory formation by neuronal synchronization. Brain Res Rev 52:170–182.

Baddeley AD, Hitch GJ (1974). Working memory. In Bower GH, editor, *The psychology of learning and motivation: Advances in research and theory, Vol 8* pp. 47–89. Academic Press.

Baddeley AD (1987). *Working memory*. Psychology Series. Oxford University Press.

Baehrens D, Schroeter T, Harmeling S, Kawanabe M, Hansen K, Muller K-R (2010) How to explain individual classification decisions. J Mach Learn Res 11:1803–1831.

Barton B, Ester EF, Awh E (2009) Discrete resource allocation in visual working memory. J Exp Psychol Human 35:1359–1367.

Baylis GC, Rolls ET (1987) Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. Exp Brain Res 65:614–622.

Bays PM, Catalao R. FG, Husain M (2009) The precision of visual working memory is set by allocation of a shared resource. J Vision 9:7.

Bays PM, Husain M (2008) Dynamic shifts of limited working memory resources in human vision. Science 321:851–854.

Beck DM, Kastner S (2009) Top-down and bottom-up mechanisms in biasing competition in the human brain. Vision Res 49:1154–1165.

Berg P, Scherg M (1994) A fast method for forward computation of multiple-shell spherical head models. Electroen Clin Neuro 90:58–64.

Bisley JW, Zaksas D, Droll JA, Pasternak T (2004) Activity of neurons in cortical area MT during a memory for motion task. J Neurophysiol 91:286–300.

Bland JM, Altman DG (1995) Calculating correlation coefficients with repeated observations: Part 1–Correlation within subjects. BMJ 310:446.

Born RT, Bradley DC (2005) Structure and function of visual area MT. Annu Rev Neurosci 28:157–189.

Boynton GM (2005) Imaging orientation selectivity: decoding conscious perception in V1. Nature Neuroscience 8:541–542.

Braddick OJ, Wishart KA, Curran W (2002) Directional performance in motion transparency. Vision Res 42:1237–1248.

Brainard DH (1997) The Psychophysics toolbox. Spatial Vision 10:433–436.

Bullmore ET, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186–198.

Buschman TJ, Siegel M, Roy JE, Miller EK (2011) Neural substrates of cognitive capacity limitations. Proc Natl Acad Sci U S A 108:11252–11255.

Cabral C, Silveira M, Figueiredo P (2012) Decoding visual brain states from fMRI using an ensemble of classifiers. Pattern Recogn 45:2064–2074.

Casali AG, Casarotto S, Rosanova M, Mariotti M, Massimini M (2010) General indices to characterize the electrical response of the cerebral cortex to TMS. NeuroImage 49:1459–1468.

Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, Casarotto S, Bruno M-A, Laureys S, Tononi G, Massimini M (2013) A theoretically based index of consciousness independent of sensory processing and behavior. Sci Transl Med 5:198ra105.

Chaimow D, Yacoub E, Uğurbil K, Shmuel A (2011) Modeling and analysis of mechanisms underlying fMRI-based decoding of information conveyed in cortical columns. NeuroImage 56:627–642.

Chao LL, Knight RT (1998) Contribution of human prefrontal cortex to delay performance. J Cognitive Neurosci 10:167–177.

Christophel TB, Haynes J-D (2014) Decoding complex flow-field patterns in visual working memory. NeuroImage 91:43–51.

Christophel TB, Hebart MN, Haynes J-D (2012) Decoding the contents of visual short-term memory from human visual and parietal cortex. J Neurosci 32:12983–12989.

Clark KL, Noudoost B, Moore T (2014) Persistent spatial information in the FEF during object-based short-term memory does not contribute to task performance. J Cognitive Neurosci 26:1292–1299.

Cohen MX (2011) It's about time. Front Hum Neurosci 5:2.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201–215.

Courtney SM, Ungerleider LG, Keil K, Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. Nature 386:608–611.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Crowe DA, Averbeck BB, Chafee MV (2010) Rapid sequences of population activity patterns dynamically encode task-critical spatial information in parietal cortex. J Neurosci 30:11640–11653.

Curtis CE, D'Esposito M (2003) Persistent activity in the prefrontal cortex during working memory. Trends Cogn Sci 7:415–423.

Curtis CE, Lee D (2010) Beyond working memory: the role of persistent activity in decision making. Trends Cogn Sci 14:216–222.

Cusack R, Mitchell DJ, Duncan J (2010) Discrete object representation, attention switching, and task difficulty in the parietal lobe. J Cognitive Neurosci 22:32–47.

De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. NeuroImage 43:44–58.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Meth 134:9–21.

Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage 34:1443–1449.

D'Esposito M (2007) From cognitive to neural models of working memory. Philos T R Soc B 362:761–772.

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

Dumoulin SO, Bittar RG, Kabani NJ, Baker CL, Le Goualher G, Pike GB, Evans AC (2000) A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. Cereb Cortex 10:454–463.

Durstewitz D (2009) Implications of synaptic biophysics for recurrent network dynamics and active memory. Neural Networks 22:1189–1200.

Emrich SM, Al-Aidroos N, Pratt J, Ferber S (2009) Visual search elicits the electrophysiological marker of visual working memory. PLoS ONE 4:e8042.

Emrich SM, Ferber S (2012) Competition increases binding errors in visual working memory. J Vision 12:12.

Emrich SM, Riggall AC, LaRocque JJ, Postle BR (2013) Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. J Neurosci 33:6516–6523.

Erickson MA, Maramara LA, Lisman JE (2010) A single brief burst induces GluR1-dependent associative short-term potentiation: A potential mechanism for short-term memory. J Cognitive Neurosci 22:2530–2540.

Ester EF, Anderson DE, Serences JT, Awh E (2013) A neural measure of precision in visual working memory. J Cognitive Neurosci 25:754–761.

Ester EF, Serences JT, Awh E (2009) Spatially global representations in human primary visual cortex during working memory maintenance. J Neurosci 29:15258–15265.

Felisberti FM, Zanker JM (2005) Attention modulates perception of transparent motion. Vision Res 45:2587–2599.

Feredoes E, Postle BR (2007) Localization of load sensitivity of working memory storage: quantitatively and qualitatively discrepant results yielded by single-subject and group-averaged approaches to fMRI group analysis. NeuroImage 35:881–903.

Feredoes E, Tononi G, Postle BR (2007) The neural bases of the short-term storage of verbal information are anatomically variable across individuals. J Neurosci 27:11003–11008.

Ferrarelli F, Massimini M, Sarasso S, Casali A, Riedner BA, Angelini G, Tononi G, Pearce RA (2010) Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. Proc Natl Acad Sci U S A 107:2681–2686.

Fischl B, Rajendran N, Busa E, Augustinack J, Hinds O, Yeo B. TT, Mohlberg H, Amunts K, Zilles K (2008) Cortical folding patterns and predicting cytoarchitecture. Cereb Cortex 18:1973–1980.

Fougnie D, Suchow JW, Alvarez GA (2012) Variability in the quality of visual working memory. Nat Commun 3:1229.

Fransén E, Tahvildari B, Egorov A, Hasselmo M, Alonso AA (2006) Mechanism of graded persistent cellular activity of entorhinal cortex layer V neurons. Neuron 49:735–746.

Freedman DJ, Assad JA (2006) Experience-dependent representation of visual categories in parietal cortex. Nature 443:85–88.

Freeman J, Brouwer GJ, Heeger DJ, Merriam EP (2011) Orientation decoding depends on maps, not columns. J Neurosci 31:4792–4804.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6:218–229.

Friston KJ, Henson R, Phillips C, Mattout J (2006) Bayesian estimation of evoked and induced responses. Hum Brain Mapp 27:722–735.

Friston KJ, Penny W, Phillips C, Kiebel S, Hinton GE, Ashburner J (2002) Classical and Bayesian Inference in Neuroimaging: Theory. NeuroImage 16:465–483.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331–349.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1993a) Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas". J Neurosci 13:1479–1497.

Funahashi S, Chafee MV, Goldman-Rakic PS (1993b) Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. Nature 365:753–756.

Fuster JM (2002). Physiology of executive function: The perception-action cycle. In Struss D, Knight RT, editors, *Principles of Frontal Lobe Function* pp. 96–108. Oxford University Press.

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173:652–654.

Gardner JL (2010) Is cortical vasculature functionally organized? NeuroImage 49:1953–1956.

Glover GH (1999) Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9:416–429.

Gnadt J, Andersen RA (1988) Memory related motor planning activity in posterior parietal cortex of macaque. Exp Brain Res 70:216–220.

Goldman-Rakic PS (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In Mountcastle VB, Plum F, Geiger SR, editors, *Handbook of neurobiology* pp. 373–417. American Physiological Society.

Goldman-Rakic PS (1990). Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates. In Uylings H, Van Eden C, DeBruin J, Corner MA, Feenstra M, editors, *Progress in brain research* pp. 325–336. Elsevier Science Publishers.

Goldman-Rakic PS (1995) Cellular basis of working memory. Neuron 14:477–485.

Golland P, Fischl B (2003) Permutation tests for classification: towards statistical significance in image-based studies. Inform Process Med Imaging Proc 18:330–341.

Graziano MS, Hu XT, Gross CG (1997) Coding the locations of objects in the dark. Science 277:239–241.

Hamidi M, Tononi G, Postle BR (2008) Evaluating frontal and parietal contributions to spatial working memory with repetitive transcranial magnetic stimulation. Brain Res 1230:202–210.

Hamidi M, Tononi G, Postle BR (2009) Evaluating the role of prefrontal and parietal cortices in memory-guided response with repetitive transcranial magnetic stimulation. Neuropsychologia 47:295–302.

Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby JV, Pollmann S (2009) PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics 7:37–53.

Hanson SJ, Halchenko YO (2008) Brain reading using full brain support vector machines for object recognition: there is no "face" identification area. Neural Comput 20:486–503.

Hanson SJ, Matsuka T, Haxby JV (2004) Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? NeuroImage 23:156–166.

Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458:632–635.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430.

Haynes J-D (2011) Multivariate decoding and brain reading: introduction to the special issue. NeuroImage 56:385–386.

Haynes J-D, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7:523–534.

Hebb DO (1949). *The Organization of Behavior: A Neuropsychological Theory*. John Wiley & Sons Hoboken, NJ.

Huk AC, Dougherty RF, Heeger DJ (2002) Retinotopy and functional subdivision of human areas MT and MST. J Neurosci 22:7195–7205.

Hussar CR, Pasternak T (2012) Memory-guided sensory comparisons in the prefrontal cortex: contribution of putative pyramidal cells and interneurons. J Neurosci 32:2747–2761.

Ihssen N, Linden D. EJ, Miller CE, Shapiro KL (In Press) Neural mechanisms underlying visual short-term memory gain for temporally distinct objects. Cereb Cortex .

Ikkai A, Curtis CE (2011) Common neural mechanisms supporting spatial working memory, attention and motor intention. Neuropsychologia 49:1428–1434.

Jacobson C (1936) Studies of cerebral functions in primates: I. The functions of the frontal association areas in monkies. Comp Psych Monog 13:1–60.

Jensen O (2006) Maintenance of multiple working memory items by temporal segmentation. Neuroscience 139:237–249.

Jerde TA, Merriam EP, Riggall AC, Hedges JH, Curtis CE (2012) Prioritized maps of space in human frontoparietal cortex. J Neurosci 32:17382.

Jha AP, McCarthy G (2000) The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional MRI study. J Cognitive Neurosci 12 Suppl 2:90–105.

Jimura K, Poldrack RA (2012) Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia 50:544–552.

Johnson JS, Kundu B, Casali AG, Postle BR (2012) Task-dependent changes in cortical excitability and effective connectivity: a combined TMS-EEG study. J Neurophysiol 107:2383–2392.

Jonides J, Lacey SC, Nee DE (2005) Processes of working memory in mind and brain. Curr Dir Psychol Sci 14:2–5.

Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ (2000) Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37:163–178.

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nature Neuroscience 8:679–685.

Kawasaki M, Watanabe M, Okuda J, Sakagami M, Aihara K (2008) Human posterior parietal cortex maintains color, shape and motion in visual short-term memory. Brain Res 1213:91–97.

King J-R, Dehaene S (2014) Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci 18:203–210.

Kleiner M, Brainard D, Pelli D, Ingling A, Murray R (2007) What's new in Psychtoolbox-3. Perception 36:ECVP Abstract Supplement.

Knight RT, D'Esposito M (2003). Lateral prefrontal syndrome: A disorder of executive control. In D'Esposito M, editor, *Neurological foundations of cognitive neuroscience* pp. 259–279. MIT Press.

Koenigs M, Barbey AK, Postle BR, Grafman J (2009) Superior parietal cortex is critical for the manipulation of information in working memory. J Neurosci 29:14980–14986.

Kriegeskorte N (2011) Pattern-information analysis: from stimulus decoding to computational-model testing. NeuroImage 56:411–421.

Kriegeskorte N, Cusack R, Bandettini PA (2010) How does an fMRI voxel sample the neuronal activity pattern: Compact-kernel or complex spatiotemporal filter? NeuroImage 49:1965–1976.

Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc Natl Acad Sci U S A 104:20600–20605.

Kriegeskorte N, Goebel R, Bandettini PA (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868.

Kubota K, Niki H (1971) Prefrontal cortical unit activity and delayed alternation performance in monkeys. J Neurophysiol 34:337–347.

LaConte S, Strother S, Cherkassky VL, Anderson J, Hu X (2005) Support vector machines for temporal classification of block design fMRI data. NeuroImage 26:317–329.

Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT (2000) Automated Talairach atlas labels for functional brain mapping. Hum Brain Mapp 10:120–131.

LaRocque JJ, Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR (2013) Decoding attended information in short-term memory: an EEG study. J Cognitive Neurosci 25:127–142.

Larzelere RE, Mulaik SA (1977) Single-sample tests for many correlations. Psychol Bull 84:557–569.

Lebedev MA, Messinger A, Kralik JD, Wise SP (2004) Representation of attended versus remembered locations in prefrontal cortex. PLoS Biol 2:e365.

Lee TG, D'Esposito M (2012) The dynamic nature of top-down signals originating from prefrontal cortex: a combined fMRI-TMS study. J Neurosci 32:15458–15466.

Lemm S, Blankertz B, Dickhaus T, Muller K-R (2011) Introduction to machine learning for brain imaging. NeuroImage 56:387–399.

Lepsien J, Nobre AC (2007) Attentional modulation of object representations in working memory. Cereb Cortex 17:2072–2083.

Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am 49:Suppl 2:467+.

Lewis-Peacock JA, Postle BR (2008) Temporary activation of long-term memory supports working memory. J Neurosci 28:8765–8771.

Lewis-Peacock JA, Postle BR (2012) Decoding the internal focus of attention. Neuropsychologia 50:470–478.

Linden D. EJ, Bittner RA, Muckli L, Waltz JA, Kriegeskorte N, Goebel R, Singer W, Munk M. HJ (2003) Cortical capacity constraints for visual working memory: dissociation of fMRI load effects in a fronto-parietal network. NeuroImage 20:1518–1530.

Linden D. EJ, Oosterhof NN, Klein C, Downing PE (2012) Mapping brain activation and information during category-specific visual working memory. J Neurophysiol 107:628–639.

Linke AC, Vicente-Grabovetsky A, Mitchell DJ, Cusack R (2011a) Encoding strategy accounts for individual differences in change detection measures of VSTM. Neuropsychologia 49:1476–1486.

Linke AC, Vicente-Grabovetsky A, Cusack R (2011b) Stimulus-specific suppression preserves information in auditory short-term memory. Proc Natl Acad Sci U S A 108:12961–12966.

Lisman JE, Idiart MA (1995) Storage of 7 +/- 2 short-term memories in oscillatory subcycles. Science 267:1512–1515.

Liu T, Hospadaruk L, Zhu DC, Gardner JL (2011) Feature-specific attentional priority signals in human cortex. J Neurosci 31:4484–4495.

Lorente de Nó R (1933) Vestibulo-ocular reflex arc. Arch Neuro Psychiatr 30:245–291.

Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. Nature 390:279–281.

Lückmann HC, Jacobs H. IL, Sack AT (2014) The cross-functional role of frontoparietal regions in cognition: internal attention as the overarching mechanism. Prog Neurobiol 116:66–86.

MacEvoy SP, Epstein RA (2009) Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. Curr Biol 19:943–947.

Magen H, Emmanouil T-A, McMains SA, Kastner S, Treisman A (2009) Attentional demands predict short-term memory load response in posterior parietal cortex. Neuropsychologia 47:1790–1798.

Mahmoudi A, Takerkart S, Regragui F, Boussaoud D, Brovelli A (2012) Multivoxel pattern analysis for FMRI data: a review. Comput Math Methods Med 2012:961257.

Majerus S, Bastin C, Poncelet M, Linden M, Salmon E, Collette F, Maquet P (2007) Short-term memory and the left intraparietal sulcus: focus of attention? Further evidence from a face short-term memory paradigm. NeuroImage 35:353–367.

Majerus S, D'Argembeau A, Martinez Perez T, Belayachi S, Linden M, Collette F, Salmon E, Seurinck R, Fias W, Maquet P (2010) The commonality of neural networks for verbal and visual short-term memory. J Cognitive Neurosci 22:2570–2593.

Majerus S, Poncelet M, Linden M, Albouy G, Salmon E, Sterpenich V, Vandewalle G, Collette F, Maquet P (2006) The left intraparietal sulcus and verbal short-term memory: focus of attention or serial order? NeuroImage 32:880–891.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. NeuroImage 19:1233–1239.

Malmo RB (1942) Interference factors in delayed response in monkeys after removal of frontal lobes. J Neurophysiol 5:295–308.

Mazyar H, Berg R, Ma WJ (2012) Does precision decrease with set size? J Vision 12:10.

Miller EK, Li L, Desimone R (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. J Neurosci 13:1460–1478.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.

Milner B (1963) Effects of different brain lesions on card sorting: The role of the frontal lobes. Arch Neurol 9:90–100.

Mitchell DJ, Cusack R (2008) Flexible, capacity-limited activity of posterior parietal cortex in perceptual as well as visual short-term memory tasks. Cereb Cortex 18:1788–1798.

Miyashita Y, Chang HS (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature 331:68–70.

Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. Science 319:1543–1496.

Mourão-Miranda J, Friston KJ, Brammer M (2007) Dynamic discrimination analysis: a spatial-temporal SVM. NeuroImage 36:88–99.

Munneke J, Belopolsky AV, Theeuwes J (2012) Shifting attention within memory representations involves early visual areas. PLoS ONE 7:e35528.

Mur M, Bandettini PA, Kriegeskorte N (2008) Revealing representational content with pattern-information fMRI–an introductory guide. Soc Cogn Affect Neur 4:101–109.

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. NeuroImage 56:400–410.

Newsome WT, Britten KH, Salzman CD, Movshon JA (1990) Neuronal mechanisms of motion perception. Cold Spring Harb Sym 55:697–705.

Niki H (1974a) Differential activity of prefrontal units during right and left delayed response trials. Brain Res 70:346–349.

Niki H (1974b) Prefrontal unit activity during delayed alternation in the monkey. I. Relation to direction of response. Brain Res 68:185–196.

Niki H (1974c) Prefrontal unit activity during delayed alternation in the monkey. II. Relation to absolute versus relative direction of response. Brain Res 68:197–204.

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424–430.

Noudoost B, Chang MH, Steinmetz NA, Moore T (2010) Top-down control of visual attention. Curr Opin Neurobiol 20:183–190.

Op de Beeck HP (2010) Probing the mysterious underpinnings of multi-voxel fMRI analyses. NeuroImage 50:567–571.

O'Reilly JX, Woolrich MW, Behrens T. EJ, Smith SM, Johansen-Berg H (2012) Tools of the trade: psychophysiological interactions and functional connectivity. Soc Cogn Affect Neur 7:604–609.

Palmer J (1990) Attentional limits on the perception and memory of visual information. J Exp Psychol Human 16:332–350.

Pashler H (1988) Familiarity and visual change detection. Percept Psychophys 44:369–378.

Pasternak T, Greenlee MW (2005) Working memory in primate sensory systems. Nat Rev Neurosci 6:97–107.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision 10:437–442.

Pereira F, Mitchell TM, Botvinick MM (2009) Machine learning classifiers and fMRI: a tutorial overview. NeuroImage 45:S199–209.

Phillips C, Mattout J, Rugg MD, Maquet P, Friston KJ (2005) An empirical Bayesian solution to the source reconstruction problem in EEG. NeuroImage 24:997–1011.

Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci 10:59–63.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. Science 310:1963–1966.

Postle BR (2005) Delay-period activity in the prefrontal cortex: one function is sensory gating. J Cognitive Neurosci 17:1679–1690.

Postle BR (2006) Working memory as an emergent property of the mind and brain. Neuroscience 139:23–38.

Postle BR (2011). What underlies the ability to guide action with spatial information that is no longer present in the environment? In Vandierendonck A, Szmalec A, editors, *Spatial Working Memory* pp. 897–901. Psychology Press.

Postle BR, Berger JS, D'Esposito M (1999) Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. Proc Natl Acad Sci U S A 96:12959–12964.

Postle BR, Ferrarelli F, Hamidi M, Feredoes E, Massimini M, Peterson M, Alexander AL, Tononi G (2006) Repetitive transcranial magnetic stimulation dissociates working memory manipulation from retention functions in the prefrontal, but not posterior parietal, cortex. J Cognitive Neurosci 18:1712–1722.

Postle BR, Zarahn E, D'Esposito M (2000) Using event-related fMRI to assess delay-period activity during performance of spatial and nonspatial working memory tasks. Brain Res Protoc 5:57–66.

Rao N, Nowak R, Cox C, Rogers T (In Press) Logistic regression with structured sparsity. J Mach Learn Res .

Rasmussen PM, Hansen LK, Madsen KH, Churchill NW, Strother SC (2012) Model sparsity and brain pattern interpretation of classification models in neuroimaging. Pattern Recogn 45:2085–2100.

Rasmussen PM, Madsen KH, Lund TE, Hansen LK (2011) Visualization of nonlinear kernel models in neuroimaging by sensitivity maps. NeuroImage 55:1120–1131.

Reddy L, Kanwisher NG, VanRullen R (2009) Attention and biased competition in multi-voxel object representations. Proc Natl Acad Sci U S A 106:21447–21452.

Reinhart R. MG, Heitz RP, Purcell BA, Weigand PK, Schall JD, Woodman GF (2012) Homologous mechanisms of visuospatial working memory maintenance in macaque and human: properties and sources. J Neurosci 32:7711–7722.

Riggall AC, Postle BR (2012) The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. J Neurosci 32:12990–12998.

Rogasch NC, Fitzgerald PB (2012) Assessing cortical network properties using TMS-EEG. Hum Brain Mapp 34:1652–1669.

Serences JT (2004) A comparison of methods for characterizing the event-related BOLD time-series in rapid fMRI. NeuroImage 21:1690–1700.

Serences JT, Ester EF, Vogel EK, Awh E (2009) Stimulus-specific delay activity in human primary visual cortex. Psychol Sci 20:207–214.

Sligte IG, Scholte HS, Lamme V. AF (2008) Are there multiple visual short-term memory stores? PLoS ONE 3:e1699.

Sligte IG, Scholte HS, Lamme V. AF (2009) V4 activity predicts the strength of visual short-term memory representations. J Neurosci 29:7432–7438.

Sligte IG, Vandenbroucke A. RE, Scholte HS, Lamme V. AF (2010) Detailed sensory memory, sloppy working memory. Front Psychol 1:175.

Sneve MH, Magnussen S, Alnæs D, Endestad T, D'Esposito M (2013) Top-down modulation from inferior frontal junction to FEFs and intraparietal sulcus during short-term memory for visual features. J Cognitive Neurosci 25:1944–1956.

Soon CS, Brass M, Heinze H-J, Haynes J-D (2008) Unconscious determinants of free decisions in the human brain. Nature Neuroscience 11:543–545.

Sreenivasan KK, Curtis CE, D'Esposito M (2014a) Revisiting the role of persistent neural activity during working memory. Trends Cogn Sci 18:82–89.

Sreenivasan KK, Vytlacil J, D'Esposito M (2014b) Distributed and dynamic storage of working memory stimulus information in extrastriate cortex. J Cognitive Neurosci 26:1141–1153.

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic coding for cognitive control in prefrontal cortex. Neuron 78:364–375.

Suzuki N, Watanabe O (2009) Perceptual costs for motion transparency evaluated by two performance measures. Vision Res 49:2217–2224.

Swaminathan SK, Freedman DJ (2012) Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. Nature Neuroscience 15:315–320.

Tadel F, Baillet S, Mosher JC, Pantazis D (2011) Brainstorm: a user-friendly application for MEG/EEG analysis. Comp Intel Neurosci 2011:879716.

Talairach J, Tournoux P (1988). *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme.

Thompson R, Correia M, Cusack R (2011) Vascular contributions to pattern analysis: comparing gradient and spin echo fMRI at 3T. NeuroImage 56:643–650.

Tikhonov AN, Arsenin V (1977). *Solutions of Ill-posed Problems*. Halsted Press.

Todd JJ, Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428:751–754.

Todd JJ, Marois R (2005) Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. Cogn Affect Behav Ne 5:144–155.

Tong F, Harrison SA, Dewey JA, Kamitani Y (2012) Relationship between BOLD amplitude and pattern classification of orientation-selective activity in the human visual cortex. NeuroImage 63:1212–1222.

Tong F, Pratte MS (2012) Decoding patterns of human brain activity. Annu Rev Psychol 63:483–509.

Tsujimoto S, Postle BR (2012) The prefrontal cortex and oculomotor delayed response: a reconsideration of the "mnemonic scotoma". J Cognitive Neurosci 24:627–635.

Uhlhaas PJ, Pipa G, Lima B, Melloni L, Neuenschwander S, Nikolić D, Singer W (2009) Neural synchrony in cortical networks: History, concept and current status. Front Integr Neurosci 3:1–19.

van den Berg R, Shin H, Chou W-C, George R, Ma WJ (2012) Variability in encoding precision accounts for visual short-term memory limitations. Proc Natl Acad Sci U S A 109:8780–8785.

Vicente-Grabovetsky A, Carlin JD, Cusack R (2014) Strength of retinotopic representation of visual memories is modulated by strategy. Cereb Cortex 24:281–292.

Virtanen J, Ruohonen J, Näätänen R, Ilmoniemi RJ (1999) Instrumentation for the measurement of electric brain responses to transcranial magnetic stimulation. Med Biol Eng Comput 37:322–326.

Vogel EK, Machizawa MG (2004) Neural activity predicts individual differences in visual working memory capacity. Nature 428:748–751.

Vogel EK, McCollough AW, Machizawa MG (2005) Neural measures reveal individual differences in controlling access to working memory. Nature 438:500–503.

Wei Z, Wang X-J, Wang D-H (2012) From distributed resources to limited slots in multiple-item working memory: A spiking network model with normalization. J Neurosci 32:11228–11240.

Wilken P, Ma WJ (2004) A detection theory account of change detection. J Vision 4:1120–1135.

Woolgar A, Thompson R, Bor D, Duncan J (2011) Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. NeuroImage 56:744–752.

Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 440:91–95.

Yang Z, Fang F, Weng X (2012) Recent developments in multivariate pattern analysis for functional MRI. Neurosci Bull 28:399–408.

Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. J Roy Stat Soc B Met 68:49–67.

Zaksas D, Bisley JW, Pasternak T (2001) Motion information is spatially localized in a visual working-memory task. J Neurophysiol 86:912–921.

Zaksas D, Pasternak T (2006) Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. J Neurosci 26:11726–11742.

Zanto TP, Rubens MT, Bollinger J, Gazzaley A (2010) Top-down modulation of visual feature processing: the role of the inferior frontal junction. NeuroImage 53:736–745.

Zanto TP, Rubens MT, Thangavel A, Gazzaley A (2011) Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. Nature Neuroscience 14:656–661.

Zarahn E, Aguirre GK, D'Esposito M (1999) Temporal isolation of the neural correlates of spatial mnemonic processing with fMRI. Cognitive Brain Res 7:255–268.

Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. Nature 453:233–235.

Zhang W, Luck SJ (2011) The number and quality of representations in working memory. Psychol Sci 22:1434–1441.

Zokaei N, Gorgoraptis N, Bahrami B, Bays PM, Husain M (2011) Precision of working memory for visual motion sequences and transparent motion surfaces. J Vision 11:2.

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J Roy Stat Soc B Met 67:301–320.