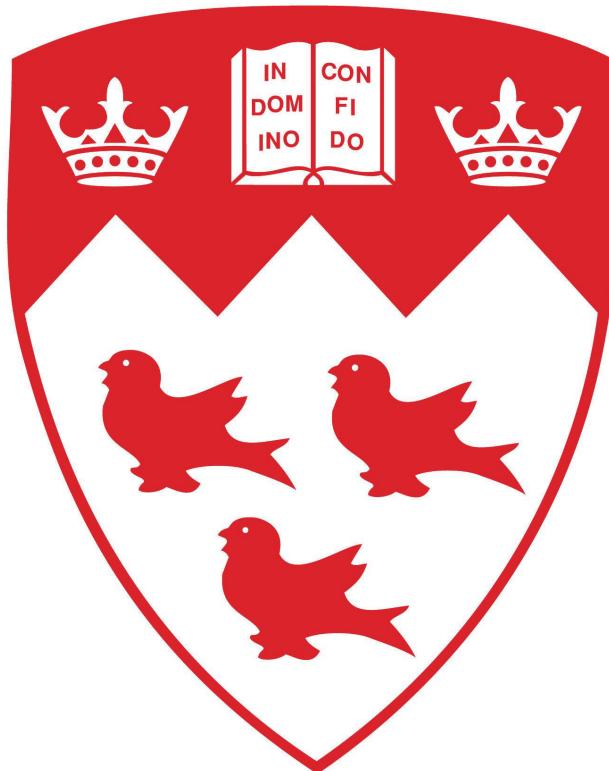


**NETWORK PROPERTIES UNDERLYING WORKING MEMORY
IN PRIMATE PREFRONTAL CORTEX**

Matthew Lewis Leavitt



Department of Physiology
McGill University
Montréal, Québec, Canada

April 2017

A thesis submitted to McGill University in partial fulfillment
of the requirements of the degree of Doctorate of Philosophy

Copyright © Matthew Lewis Leavitt 2017

TABLE OF CONTENTS

SUMMARY	vi
RÉSUMÉ	ix
ACKNOWLEDGEMENTS	xii
CONTRIBUTIONS OF AUTHORS	xiv
LIST OF SYMBOLS AND ABBREVIATIONS	xvii
<u>CHAPTER 1</u> Introduction, Literature Review, and Aims	1
1.1 Introduction	2
1.2 Structure and Function of the PFC	6
1.3 Origins of PFC Research	8
1.4 Neuropsychology of Working Memory	10
1.5 Electrophysiology of Working Memory and PFC	11
1.6 Beyond Working Memory: Sustained LPFC Activity Across Tasks	14
1.7 Network Structure and Topographic Representation in 8a	15
1.8 Non-linear Representation of Visual Mnemonic Space	18
1.9 Beyond the Single Neuron	20
1.10 Spike Count Correlations (r_{sc})	23
1.11 Research Aims	25
<u>CHAPTER 2</u> Spike Count Correlation Structure Reveals Intrinsic Functional Topography in Primate Prefrontal Cortex Area 8a	32
2.1 Abstract	34
2.2 Introduction	35
2.3 Methods	39
2.3.1 Ethics Statement	39
2.3.2 Task	40
2.3.3 Apparatus and recording procedures	40
2.3.4 Microelectrode array (MEA) implant	42
2.3.5 Recordings and spike detection	43
2.3.6 Estimation of inter-electrode distance	46
2.3.7 Data analysis	46

2.3.8 Visuospatial tuning	47
2.3.9 Control simulations	48
2.4 Results	49
2.4.1 Influence of distance between neurons and tuning properties on r_{sc}	54
2.4.2 Dissociating proximity and tuning similarity	63
2.5 Discussion	63
2.5.1 Previous correlation studies in dlPFC	64
2.5.2 Comparison with studies of r_{sc} in other brain areas	69
2.5.3 Factors that affect r_{sc}	71
2.6 Conclusion	72
2.7 Acknowledgements	72
<u>CHAPTER 3 A Quadrantic Bias in Prefrontal Representation of Visual-Mnemonic Space</u>	73
3.1 Abstract	75
3.2 Introduction	76
3.3 Materials and methods	80
3.3.1 Ethics Statement	80
3.3.2 Task	81
3.3.3 Experimental Setup	83
3.3.4 Microelectrode Array (MEA) Implant	84
3.3.5 Recordings and Spike Detection	84
3.3.6 Analysis Epochs	84
3.3.7 Single Unit Yield and Epoch Selectivity	85
3.3.8 Spatial Autocorrelation Analysis	85
3.3.9 Single Unit Firing Rate Meridian Effects	87
3.3.10 Stepwise Regression	88
3.3.11 Quadrantic Bias Visualization	89
3.3.12 Correlation (r_{sc}) Analysis	90
3.3.13 Population Decoding	91
3.3.14 Saccade Endpoint Distribution Variability	92
3.4 Results	93
3.4.1 Anatomical Topography of Mnemonic Representations in dlPFC	94
3.4.2 Spatial Bias in Single-Neuron Firing Rates	97

Table of Contents

3.4.3 Correlated Variability During WM Maintenance	103
3.4.4 Quadrantic Bias in Single-Trial Ensemble Representations	107
3.4.5 Quadrantic Biases in Behavior	110
3.5 Discussion	114
3.5.1 Clustering of Mnemonic Representations in dLPFC	114
3.5.2 Potential Origins of Quadrantic Biases in Visual and Mnemonic Space	117
3.5.3 Alternative Factors Affecting Memory-Guided Saccades	121
3.5.4 Meridian Effects Elucidate the Relationships Between WM, Attention, and Motor Activity	123
3.5.5 r_{sc} and WM Tasks	125
3.5.6 A Hemifield-Independent WM Mechanism	126
3.6 Conclusion	127
3.7 Acknowledgements	128
3.8 Supplemental Data	129
CHAPTER 4 Correlated Variability Modifies Working Memory Fidelity in Primate Prefrontal Neuronal Ensembles	150
4.1 Abstract	152
4.2 Introduction	153
4.3 Results	157
4.3.1 Task-related modulation of spike count correlations (r_{sc})	159
4.3.2 Quantifying information content in neuronal ensembles using linear decoders	161
4.3.3 Effects of r_{sc} and r_{signal} structures on WM coding efficiency	168
4.3.4 Different ensemble configurations optimize WM coding when the r_{sc} structure is intact vs. removed	172
4.3.5 Ensembles optimized for WM representation are r_{signal} diverse and anatomically dispersed	173
4.3.6 Non-selective units can improve WM coding by modifying the r_{sc} structure	176
4.4 Discussion	179
4.4.1 Recurrent network dynamics during WM coding	180
4.4.2 Decoding WM representations from LPFC neuronal ensembles	181
4.4.3 Effects of r_{sc} and r_{signal} structures on WM coding	182
4.4.4 Effects of the r_{sc} structure on information in non-WM tasks	183

4.4.5 Generalization to larger ensembles	184
4.4.6 Effects of spike sorting	186
4.4.7 Noise shaping units	187
4.5 Conclusion	188
4.6 Materials and methods	189
4.6.1 Ethics Statement	189
4.6.2 Task	190
4.6.3 Experimental Setup	191
4.6.4 Microelectrode Array (MEA) Implant	192
4.6.5 Recordings and Spike Detection	192
4.6.6 Analysis Epochs	193
4.6.7 Spatial Selectivity	193
4.6.8 Spike Count Correlation (r_{sc}) and Signal Correlation (r_{signal}) Analysis	194
4.6.9 Population Decoding	196
4.6.10 Functional anatomy	198
4.6.11 Non-selective noise-shaping neuron analysis	200
4.7 Acknowledgements	201
4.8 Supplemental Data	202
CHAPTER 5 General Discussion	213
5.1 Summary	214
5.2 Reconciling the relationship between tuning and r_{sc} across experiments	216
5.3 Stability of prefrontal WM representations	218
5.4 Layer-specific functional properties of dlPFC	222
5.5 FEF, 8a, and prefrontal cortex	226
5.6 Sustained activity underlying WM: A ubiquitous cortical phenomenon?	228
5.7 Reconciling fMRI and electrophysiological findings	231
5.8 Features of neural circuits allowing sustained activity and WM	233
5.9 Future directions for WM and cognition research in PFC	236
REFERENCES	244

SUMMARY

Working memory (WM) is the transient (i.e. milliseconds to seconds) maintenance and/or manipulation of information that is no longer available to the senses. WM capacity is strongly correlated with measures of intelligence and is a crucial component of goal-directed behavior. More than 80 years of neuropsychological and neurophysiological research have demonstrated the role of the prefrontal cortex (PFC) in WM. The WM-related activity in primate PFC is hypothesized to arise from the structure of the network in which the neurons are embedded. Recent studies have also shown that it is difficult to predict the properties of neuronal ensembles from the properties of individually-examined neurons. However, our current understanding of the ensemble-level mechanisms of WM comes largely from theoretical modeling and speculation. This doctoral thesis seeks to rectify this issue experimentally.

First, I examined the intrinsic network structure of macaque area 8a by using high-density microelectrode arrays to record from ensembles of simultaneously-active neurons during fixation behavior. We found a pattern of intrinsic functional connectivity between neurons indicative of a recurrent-

excitatory lateral-inhibitory network structure, which has been hypothesized to allow WM maintenance.

Next, I used the same microelectrode arrays to record ensemble activity while the animals performed a spatial working memory (SWM) task, and found neuronal correlates of a psychophysical bias in which visual-mnemonic space is divided by the meridians of the visual field. This bias is present at multiple levels of examination: in single neuron firing rates, the structure of correlated firing rate variability, and the neuronal ensemble representations. We also found that 8a neurons are anatomically clustered in a non-retinotopic manner that partially reflects the organization of visual space.

Finally, I examined how ensembles of simultaneously-recorded neurons in 8a represent SWM and how it differs from single neuron representations. We found that SWM maintenance modulated the structure of correlated variability in a manner indicative of increased coupling between similarly-tuned neurons and increased inhibition between dissimilarly-tuned neurons. This structure of correlated variability could facilitate or impair the readout of WM representations, depending on the size of the ensemble and tuning properties of its constituent neurons. We also found neurons that contained little WM information when

Summary

examined in isolation, but contributed to coding when part of an ensemble by shaping the ensemble's correlated variability structure. This final result is a powerful example of a phenomenon that is inaccessible in individually-recorded neurons, and emphasizes the importance of multi-neuron recordings for elucidating the neuronal mechanisms of cognition.

RÉSUMÉ

La mémoire de travail (MdT) est la maintenance temporaire et / ou la manipulation d'informations qui ne sont plus disponibles aux sens. La capacité de la MdT est fortement corrélée avec les mesures d'intelligence et constitue un élément crucial du comportement adaptatif humain. Plus de 80 ans de recherche neuropsychologique et neurophysiologique ont démontré le rôle du cortex préfrontal (CPF) dans la MdT. L'activité liée à la MdT dans le cortex préfrontal des primates découle supposément de la structure du réseau dans laquelle les neurones sont intégrés. Des études récentes ont montré qu'il est difficile de prédire les propriétés des ensembles neuronaux à partir des propriétés des neurones examinés individuellement. Cependant, notre compréhension des mécanismes de réseau de la MdT provient en grande partie de la modélisation théorique et de la spéculation. Cette thèse de doctorat vise à adresser cette question expérimentalement.

Tout d'abord, j'ai examiné la structure neuronale de la zone anatomique 8A du cerveau du singe en utilisant des puces de microélectrodes à haute densité pour enregistrer des ensembles de neurones simultanément actifs pendant le comportement de fixation visuelle. Nous avons trouvé un schéma de connectivité

Résumé

fonctionnelle intrinsèque entre les neurones indiquant une structure d'inhibition latérale récurrente, en support aux modèles théoriques de la MdT.

Ensuite, j'ai utilisé les mêmes microélectrodes pour enregistrer la méthodologie de l'activité d'ensemble neuronal tandis que l'animal effectuait une tâche de mémoire de travail spatial (MdTS) et j'ai trouvé des corrélats neuronaux d'un biais psychophysique dans lequel l'espace visuel-mnémonique du champ visuel. Ce biais est présent à plusieurs niveaux d'examen : dans les taux d'activation des neurones individuels, la structure de la variabilité corrélée de l'activation et les représentations d'ensemble neuronal. Nous avons également constaté que les neurones de 8A sont anatomiquement regroupés de manière non-rétinotopique.

Enfin, j'ai examiné comment les ensembles de neurones simultanément enregistrés dans 8A représentent la MdTS et comment ils diffèrent des représentations de neurones individuels. Nous avons constaté que la maintenance de la MdTS module la structure de la variabilité corrélée d'une manière indicative d'un couplage accru entre des neurones similaires et d'une inhibition accrue entre des neurones dissemblables. Cette structure de variabilité corrélée pourrait faciliter ou nuire à la lecture des représentations de la MdTS dépendamment de la taille de l'ensemble et des propriétés d'accord de ses neurones constitutifs. Nous avons

également trouvé des neurones qui contiennent peu d'information lorsqu'ils sont examinés isolément, mais qui contribuent au codage lorsqu'ils font partie d'un ensemble en façonnant la structure de la variabilité corrélée de l'ensemble. Ce résultat final est un exemple puissant d'un phénomène inaccessible par l'enregistrement individuel des neurones et souligne l'importance d'analyser la coopération au sein d'ensembles neuronaux pour élucider les mécanismes neuronaux de la cognition.

ACKNOWLEDGEMENTS

I'd like to start by thanking my supervisor, Dr. Julio Martinez-Trujillo. I am forever grateful for the chance he took on me as an inexperienced undergrad. His contagious passion and unbridled excitement about scientific discovery are nonpareil, and I aspire to embody his motivation, dedication, and enthusiasm.

I'd also like to thank my advisory committee: Dr. Erik Cook, for calling out my hand waving, and showing me the value of distilling an idea down to a form that can be drawn out on a whiteboard; Dr. Chris Pack, whose dry wit I can always rely on to add a layer of refreshing humor to his understated insights; and Dr. Paul Cisek, for allowing me to waylay him with what has amounted to hours of fascinating discussions in hallways, conferences, and even the produce market.

I cannot sufficiently express my gratitude to and adoration of the lab family. They create an environment that is intellectually and socially enriching. Their motivation and scrutiny have been indispensable in the creation of this thesis and my growth as a scientist, and their contributions too numerous to list. This thesis is a truly a collective effort.

I will forever cherish the scientific and personal support provided by Thomas Quail, Rishi Rajalingham, Roberto Gulli, and Lyndon Duong.

I'd like to express my heartfelt gratitude to M. Lipton and K. Carter, for broadband intellectual stimulation, emotional support, editorial assistance, and being models of resourcefulness, dedication, thoughtfulness, and creativity.

Finally, I thank my family: Catherine Lewis and Andrew Leavitt, for providing me a nurturing environment and encouraging me to pursue my curiosity, both explicitly and by example; Daniel, for providing sage guidance, holding me accountable for my claims and actions, and leading by example; and Allison MacKinnon, for her boundless love and support.

CONTRIBUTIONS OF AUTHORS

This thesis contains material from four published manuscripts:

Leavitt, M., Pieper, F., Sachs, A., Joober, R., Martinez-Trujillo, J.C. (2013) Structure of spike count correlations reveals functional interactions between neurons in dorsolateral prefrontal cortex area 8a of behaving primates, *PLoS ONE*, 8(4): e61503 (Leavitt et al., 2013). This study was conceived and designed by Drs. Pieper, Sachs, Joober, and Martinez-Trujillo. Drs. Sachs, Pieper, and Martinez-Trujillo, and I performed the surgical procedures. Dr. Pieper trained the animals and collected the neurophysiological recordings. Dr. Pieper and I analyzed the data. Dr. Martinez-Trujillo and I wrote the manuscript, and all authors contributed to revisions and editing.

Leavitt, M., Pieper, F., Sachs, A., Martinez-Trujillo, J.C. (2017), Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles, *PNAS*, 114(12): E294-E205 (Leavitt et al., 2017c). This study was conceived and designed by Drs. Pieper, Sachs, and Martinez-Trujillo, and I. Drs. Sachs, Pieper, and Martinez-Trujillo, and I performed the surgical procedures. I trained the animals, collected the neurophysiological recordings, and analyzed the

data. Dr. Martinez-Trujillo and I wrote the manuscript, and all authors contributed to revisions and editing.

Leavitt, M., Pieper, F., Sachs, A., Martinez-Trujillo, J.C. A quadrantic bias in prefrontal representation of visual-mnemonic space (2017), *Cerebral Cortex*, 1-17 (Leavitt et al., 2017b). This study was conceived and designed by Drs. Pieper, Sachs, and Martinez-Trujillo, and I. Drs. Sachs, Pieper, and Martinez-Trujillo, and I performed the surgical procedures. I trained the animals, collected the neurophysiological recordings, and analyzed the data. Dr. Martinez-Trujillo and I wrote the manuscript, and all authors contributed to revisions and editing.

Leavitt, M., Mendoza-Halliday, D., Martinez-Trujillo, J.C. Sustained activity encoding working memories: not fully distributed (2017), *Trends in Neurosciences*, 40(6): 328-346 (Leavitt et al., 2017a). Dr. Martinez-Trujillo conceived of the topic for the review, performed the bulk of the literature aggregation and analysis, and wrote the initial draft of the manuscript. I assembled the manuscript components and performed substantial editing. All authors contributed to revisions, editing, and figure creation.

Chapter 1 contains excerpts and modifications of material from {Leavitt:2017jk, Leavitt:2017jl, Leavitt:2013ir}.

Contributions of Authors

Chapter 2 contains Leavitt et al., (2013) *PLoS ONE*.

Chapter 3 contains Leavitt et al., (2017) *Cerebral Cortex*.

Chapter 4 contains Leavitt et al., (2017) *PNAS*

Chapter 5 contains excerpts and modifications of material from Leavitt et al., (2017)

PNAS, Leavitt et al., (2017) *Cerebral Cortex*, and Leavitt et al., (2017) *Trends in Neurosciences*.

Megan Schneiderman provided technical assistance in primate training and data acquisition.

Roberto Gulli, Theda Backen, Benjamin Corrigan, M. Lipton, and K. Carter provided editorial assistance in the preparation of this thesis.

This thesis uses the TeX Gyre Schola typeface (<https://www.fontsquirrel.com/fonts/tex-gyre-schola>), and a template provided by Georgia Tech (<http://grad.gatech.edu/theses-dissertations-templates>).

LIST OF SYMBOLS AND ABBREVIATIONS

dLPFC/dlPFC dorsolateral prefrontal cortex

LPFC lateral prefrontal cortex

WM working memory

r_{sc} spike count correlation

r_{signal} signal correlation

SVM support vector machine

FEF frontal eye fields

ODR oculomotor delayed-response

CHAPTER 1

INTRODUCTION, LITERATURE REVIEW, AND AIMS

1.1 Introduction

A driver thinks little about the routines he or she encounters on the road during a daily commute. However, a successful transit from home to work could be considered a lesson in prefrontal cortex (PFC) function. A driver might first depress the brake while starting the car, waiting until the audible response issues from the motor before removing pressure from the pedal. A driver must recognize cross traffic at a stop sign and determine yield conditions, manipulating the controls for speed and direction without looking away from the road. A driver must maintain the positional information of other vehicles outside the immediate field of view. A driver confronted with a split second decision to drive directly over either a pothole or an open manhole will choose the safer of the two obstacles.

In the face of all this active decision making, it is as remarkable that the driver not only makes the correct choices when forced to decide, but maintains a steady state of non-decision making while not presented with new, material data. Additionally, the driver does not randomly stop the car or veer off the road in response to novel or unexpected stimuli. This vigilant reassessment when confronted with trivial data and the ability to reason out decision-critical information from noise or repetitious inputs is an equally important capability. A

flexible resting state with respect to the range of sensory inputs allows humans to decipher meaning from the dense reality around them.

This scenario involves coordinating detailed sensory and motor information with internal goals and motivations in a manner that is robust to the multitude of alternative possibilities. Flexibility and diversity of behaviors are considered hallmarks of intelligence. The cognitive control required to do this involves coordination across many areas of the brain, but PFC is widely considered to be paramount (Passingham, 1993; Petrides, 1994; Miller and Cohen, 2001; Fuster, 2008). On a pragmatic level, research into the neural mechanisms underlying these identifiable cognitive capabilities continues to aid countless individuals, giving insight into their pathologies and providing novel therapeutic routes toward the mitigation of neurological illness. More broadly, however, developing accountable mechanistic descriptions of cognition has generated as much insight into the biological foundations of our shared humanity as it has provided avenues for understanding any specific etiology.

Given the myriad phenomena relevant to “cognitive control”, a tractable research program requires selecting specific, clearly-defined exemplars to study. Primates are known to have sophisticated behavioral repertoires, and PFC

development differentiates primates from other animals (Miller and Cohen, 2001; Fuster, 2008). Thus, I have chosen to focus on one model phenomenon that is reliant on the PFC in non-human primates: working memory (WM), the transient (i.e. milliseconds to seconds) maintenance of information that is no longer available to the senses (Baddeley and Hitch, 1974; Baddeley, 1992). More specifically, I will be studying spatial working memory (SWM—sometimes called visuospatial working memory or visuospatial short term memory), WM for a location in space (Fuster and Alexander, 1971; Goldman-Rakic, 2011). SWM is used, for example, when one rests their utensil on the table while dining, perhaps to take a sip of wine, and then picks up the utensil again without ever having to break eye contact with their dining partner.

While the need to respect social mores during a fine dining experience may not be the most scientifically compelling or generalizable scenario, the ubiquity and importance of WM cannot be understated. Goldman-Rakic expressed this eloquently:

At the most elementary level, our basic conceptual ability to appreciate that an object exists when out of view depends on the capacity to keep events in mind beyond the direct experience of those events. For some organisms, including most humans under certain conditions, "out of sight" is equivalent to "out of mind." However, working

memory is generally available to provide the temporal and spatial continuity between our past experience and present actions. (Goldman-Rakic, 1995)

The necessity for complex behavior can also be understood in the context of evolutionary pressure to optimize the tradeoff between metabolic costs and information processing capacity (Niven and Laughlin, 2008; Atick, 2011). Information processing in the nervous system is metabolically costly. The human brain accounts for approximately 20% of resting oxygen consumption, despite constituting only 2% of body mass (Kety, 1957). Increasing the information processing capacity of the nervous system would rapidly become metabolically untenable. One effect of this bandwidth limitation is that organisms must obtain information about the environment by selectively sampling subsections of it. These subsections are then integrated over time to construct a cohesive representation of the external environment. Much of this processing and integration occurs below the level of awareness, and this component is commonly conceptualized as “perception”. Nevertheless, this process also applies to quantities in the “cognitive” domain, which typically involve greater abstraction, flexibility, and/or occur over longer timescales, and of which the subject is aware. WM can be considered the application of this process to cognitive variables. Re-examining the previously presented example of driving, one can find countless examples of its usage: maintaining a

representation of other vehicles that are not currently visible; the most recently observed speed limit sign; the street number of the destination, repeated to oneself in order to constantly refresh the memory. The complexity of the situation rapidly becomes overwhelming. It should come as no surprise that WM capacity is strongly correlated with measures of human intelligence (Engle et al., 1999; Conway et al., 2003).

1.2 Structure and Function of the PFC

WM is one of myriad cognitive phenomena in which the PFC is involved (Miller and Cohen, 2001; D'Esposito and Postle, 2015; Riley and Constantinidis, 2015). Other phenomena associated with the PFC include associative learning (Wise and Murray, 2000), attentional filtering (Noudoost et al., 2010), and strategic reasoning (Lee, 2008), all across a variety of modalities. Miller and Cohen have attempted to integrate the diverse set of phenomena and findings into a model of the PFC, a brain region they describe as serving

... a specific function in cognitive control: the active maintenance of patterns of activity that represent goals and the means to achieve them. They provide bias signals throughout much of the rest of the brain, affecting not only visual processes but also other sensory modalities, as well as systems responsible for response execution, memory retrieval, emotional evaluation, etc. The aggregate effect of these

bias signals is to guide the flow of neural activity along pathways that establish the proper mappings between inputs, internal states, and outputs needed to perform a given task (Miller and Cohen, 2001).

While Miller and Cohen were not the first to propose the PFC as the locus of goal-directed behavior (Batuev et al., 1980; Batuev, 1986), their model successfully articulates the commonalities in the diversity of phenomena associated with the PFC. Indeed, “*active maintenance of patterns of activity that represent goals and the means to achieve them*” succinctly encapsulates WM—assuming the external information is not currently available to the senses.

Working memory-related activity has been localized to specific sub-regions of the PFC in human and non-human primates (Passingham, 1993; Miller and Cohen, 2001; Petrides, 2005a; 2005b; Fuster, 2008; Goldman-Rakic, 2011). Electrophysiological, neuropsychological, and functional imaging studies have shown that lateral prefrontal cortex (LPFC), specifically macaque and human areas 46, 9/46d, 9/46v, and 8a (Figure 1.1), are involved—correlatively and causally—in SWM.

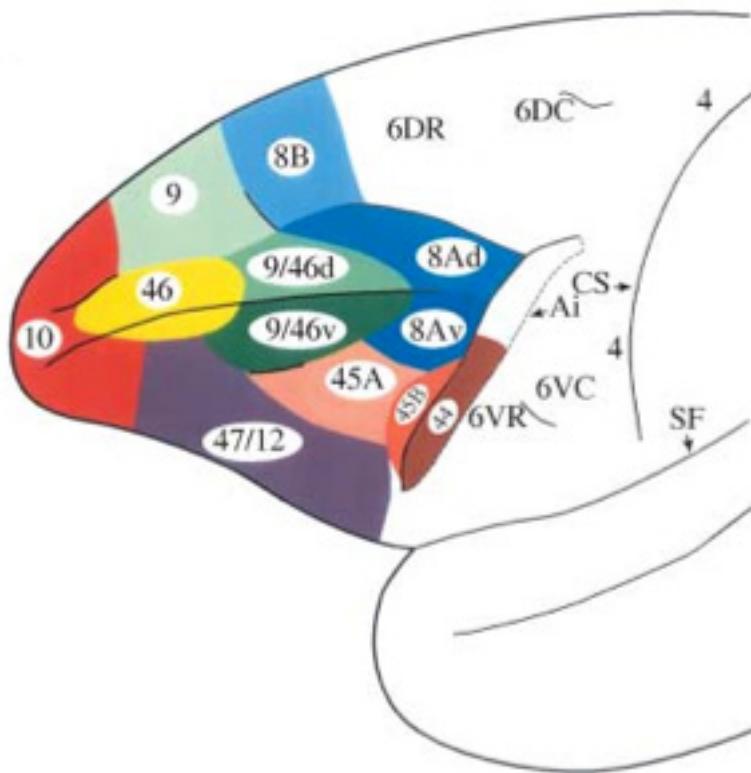


Figure 1.1: Cytoarchitectonic map of the lateral surface of macaque PFC. From (Petrides, 2005a). Abbreviations: Ai, inferior arcuate sulcus; CS, central sulcus; SF, Sylvian fissure.

1.3 Origins of PFC Research

The PFC's role in behavior was succinctly encapsulated by the neurologist Bianchi in 1895, when describing “frontal lobe syndrome” as resulting in a “failure of synthesis” (Bianchi, 1895; 1922). He then enumerates the symptoms:

Loss of ‘perceptive power’, leading to defective attention and object recognition; reduction in memory; reduction in ‘associative power’, leading to lack of coordination of the individual steps leading towards a given goal, and thus to severe difficulty

solving anything but the most simple problem; altered emotional attachments, leading to serious changes in ‘sociality’; disruption of focal consciousness and purposive behavior, leading to apathy and/or distractibility (Bianchi, 1895; 1922; Smith, 2003)

This account is remarkably similar to modern assessments of PFC function (Milner, 1982; Milner and Petrides, 1984; Stuss and Benson, 1984; Petrides, 1994; Miller and Cohen, 2001).

The relationship between the PFC and cognition was also compelling to Lashley, who proposed prescient research questions in 1920, addressing functional neuranatomical distinctions: “*the question of prime importance ... is that of the existence of anatomical divisions of the nervous system to which the function of learning is restricted*” (Lashley, 1920), as well as the causal role of PFC in learning: “*learned reactions are mediated by ... the frontal and parietal association areas in higher mammals ... [but] do the association areas have a directive function in learning or are they masses of conductive tissue through which nerve impulses penetrate at random from afferent to efferent projection areas?*” (Lashley, 1920).

While Lashley’s conclusions that “memory” is unitary and stored diffusely across the whole cortex proved inaccurate, due largely to imprecise lesioning methods and multifaceted tasks, his work is still considered foundational.

1.4 Neuropsychology of Working Memory

The foundations of neuropsychological WM research can be traced to Jacobsen, who tested chimps and monkeys with PFC lesions on two kinds of WM tasks, variants of which are still in use (Jacobsen, 1935; Jacobsen et al., 1935; Jacobsen and Nissen, 1937). In the “delayed memory response” task, the monkey watches while a food reward is placed under one of two identical cups. A screen is then lowered that occludes the food wells from view, and after a variable delay epoch of seconds to minutes, the screen is lifted and the subject is allowed to choose one of the two cups. A correct response grants access to the food reward, while an incorrect response yields no reward. The “delayed memory alternation” task is very similar, except the cup hiding the food reward is always the cup that did not contain the reward in the most recent successful trial. Jacobsen found that monkeys with unilateral lesions of PFC were not noticeably impaired on either task compared to pre-lesion performance, while monkeys with bilateral lesions were severely impaired on both tasks, and did not improve with additional training.

Further research by Blum (Blum, 1952) showed that lesions of macaque PFC impaired performance on visual and auditory delayed memory response tasks, as well as visual and auditory delayed-response tasks without a memory component

(i.e. the stimulus was visible throughout the delay). Crucially, Blum also demonstrated that the impairment was most severe in subjects with focal lesions of the principal sulcus and prearcuate areas (i.e. LPFC), and that the impairment was proportional to the lesion size. A series of studies by Mishkin and Pribram (Mishkin and Pribram, 1955; 1956; Pribram and Mishkin, 1956) demonstrated that similar lesions also cause performance deficits with modified numbers of choices, spatial layouts, and in object-based WM, in both alternation and non-alternation tasks.

These neuropsychological studies are the foundations of WM research, and primed the field for the electrophysiological techniques that were widely adopted in the following decades.

1.5 Electrophysiology of Working Memory and PFC

A potential neural mechanism for WM was first proposed in 1949, when Hebb postulated that sustained neuronal activity in the absence of stimulus input could serve as the neural mechanism for WM maintenance (Hebb, 1949; 2005). Two decades later, in 1971, Fuster and Alexander discovered neurons in macaque LPFC (and the medial dorsal nucleus of the hypothalamus) that exhibited sustained activity during the delay epoch of a delayed memory response task (Fuster and Alexander, 1971) (WM-selective neurons). That same year, Kubota and Niki found

neurons that were more active during the delay portion of a delayed memory alternation task (Kubota and Niki, 1971). Additional studies by Niki found WM-selective neurons that were also selective for the absolute and relative spatial positions of the stimuli (and thus could be considered SWM-selective neurons), and responses (Niki, 1974a; 1974b; Niki and Watanabe, 1976), and additional studies by Fuster found that the activity of WM-selective neurons also reflected accuracy of performance and was attenuated by distracting stimuli. Subsequent studies corroborated these results (Batuev et al., 1979; 1980; Kojima and Goldman-Rakic, 1982; 1984; Batuev et al., 1985), described temporal dynamics to the WM activity (e.g. ramping and decay), and also found similar signals in parietal cortex.

In 1989, Funahashi et al. recorded from the principal and arcuate regions of macaques while they performed an oculomotor delayed response (ODR) task, which has proven to be one of the most widely used tasks in SWM research (Figure 1.2).

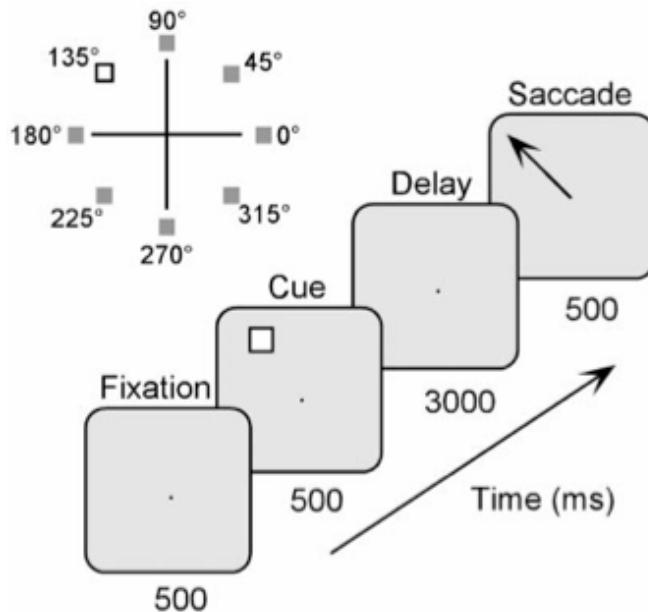


Figure 1.2: Oculomotor delayed response (ODR) task. In this task, subjects first fixate on a central point (fixation epoch), then are presented with a cue (cue epoch) that disappears, followed by a delay (delay epoch), and finally the fixation point disappears, cuing the animal to saccade to the remembered location (saccade epoch). From (Constantinidis et al., 2001a).

This study made a number of important discoveries: First, it found neurons that were tuned to precise locations in visual space (typically in the hemifield contralateral to the hemisphere being recorded from), in contrast to previous studies that used only a small number of stimuli; second, it found that this tuning was relatively parametric, in that the responses of a neuron to all the different stimulus locations could be approximated with a Gaussian function; finally, it demonstrated the existence of inhibitory delay epoch activity—neurons that fired significantly less during WM maintenance.

1.6 Beyond Working Memory: Sustained LPFC Activity Across Tasks

Further experiments found neurons in LPFC that exhibit delay-epoch selectivity for nearly every feature of a WM task (e.g. object/sound identity, spatial position, motor act, behavioral relevance) (Passingham, 1993; Goldman-Rakic, 1995; Petrides, 2000b; Courtney, 2004; Petrides, 2005a; Postle, 2006; Fuster, 2008; Funahashi, 2013). Results implying the separation of object and spatial WM led Goldman-Rakic to propose that PFC is the neural substrate for the storage of WM, and different modalities are relegated to different anatomical sub-areas (Wilson et al., 1993; Goldman-Rakic, 1995). This seems like a clear, tangible definition, but it begins to wither when considering that the separation may have been the result of training the animals on object and spatial WM tasks at different times, and was not replicated when this issue was controlled for experimentally (Rao et al., 1997). Furthermore, the issue of how to dissociate different features to different cortical domains becomes increasingly untenable as one counts the “features” for which specific WM capabilities have been identified (e.g. color, motion, egocentric space, allocentric space, tactile stimuli, faces, animals... see (Postle, 2006) for a review).

Postle proposes an alternative theory: WM is an emergent property of PFC, that the retention of information in WM is associated with sustained activity in the

same brain regions that are responsible for the representation of that information, and that the delay activity observed in PFC is better attributed to the maintenance of abstract behavioral goals or task rules (Postle, 2006). This reconciles studies that report signals in LPFC related to, among other things, motivation and reward expectancy, transformation/response preparation, and attention(selection, and integrates “WM” activity parsimoniously with Miller and Cohen’s model of PFC. Indeed, prior studies from my research group analyzed signals from the same arrays implanted in the same region in the same animals as the studies constituting my thesis, but the animals performed a change detection task that required sustained spatial attention (Tremblay et al., 2014; 2015). Thus the neuronal activity observed in LPFC during SWM tasks can be considered as the allocation of behavioral significance and computational resources to a specific region of visual space.

1.7 Network Structure and Topographic Representation in 8a

While Postle’s model of WM as an emergent phenomenon renders Goldman-Rakic’s proposition of domain-specific WM regions in PFC tenuous, her basic model of LPFC cellular network architecture is still well-supported (Goldman-Rakic, 1995; Kritzer and Goldman-Rakic, 1995; Goldman-Rakic, 1996; Rao et al., 1999). She proposed

that LPFC is organized in a microcolumnar manner, such that groups of cells within the same ~.7mm region share recurrent excitatory connections, while inhibitory connections to other microcolumns extend laterally up to 7mm (Figure 1.3) (Goldman-Rakic, 1995; Kritzer and Goldman-Rakic, 1995; Goldman-Rakic, 1996; Rao et al., 1999). Only the excitatory functional interactions predicted by this connection scheme have been observed, and only at distances of less than 2mm (Constantinidis and Goldman-Rakic, 2002; Constantinidis et al., 2002). It has been proposed that the sustained activity underlying WM maintenance is effected by increasing the strength of the recurrent excitation and lateral inhibition of these microcolumns (Amit and Brunel, 1997; Camperi and Wang, 1998; Compte et al., 2000; Durstewitz et al., 2000; Wang, 2001; Constantinidis and Wang, 2004; Wimmer et al., 2014).

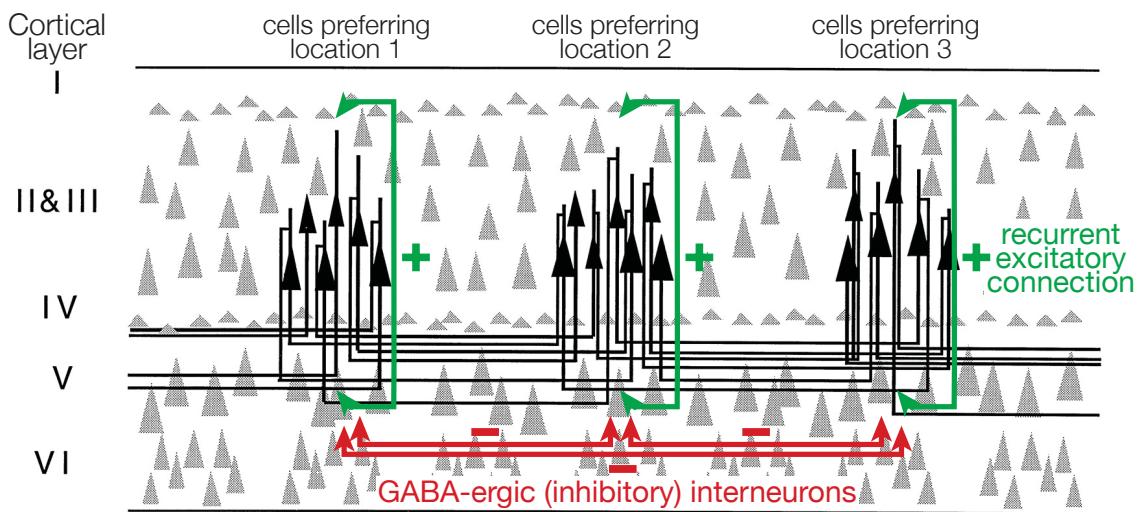


Figure 1.3: Model of LPFC network structure. Adapted from (Constantinidis and Wang, 2004).

Goldman-Rakic also posited that these microcolumns were organized by spatial tuning, which seems implausibly rigid in light of the discussed evidence supporting a more modality-agnostic role for LPFC. But it remains a possibility that the microcolumns are organized by some less distinct or more flexible property. While most electrophysiological studies relate neurons' functional properties to electrode penetration sites, few verify their claims statistically. Additionally, day-to-day movement of the brain inside the skull makes functional localization with single electrodes less certain. Studies examining the functional topography and intrinsic network structure in 8a are scant (Suzuki and Azuma, 1983), likely due in

part to the inherent difficulty of providing concrete descriptions of a region that's known for its flexibility and context sensitivity.

1.8 Non-linear Representation of Visual Mnemonic Space

Most SWM studies have parameterized visual space as either binary (e.g. left/right) or unidimensional (e.g. degrees of angle across the same eccentricity) (Funahashi and Kubota, 1994; Goldman-Rakic, 1995; Funahashi and Takeda, 2002). Although such studies have substantially advanced our understanding of SWM, they have also led to models that assume a continuous and/or homogenous representation of the visual field (Goldman-Rakic, 1995; Camperi and Wang, 1998; Compte et al., 2000; Constantinidis and Wang, 2004; Wimmer et al., 2014).

A number of behavioral and physiological studies examining WM capacity have demonstrated varying degrees of independence between the left and right visual hemifields (Vogel and Machizawa, 2004; Delvenne, 2005; Buschman et al., 2011; Delvenne et al., 2011; Matsushima and Tanaka, 2014). However, these studies have treated visual space as a binary variable, restricting their ability to make conclusions about visual mnemonic space beyond that it is represented separately for each hemifield.

Psychophysical studies have revealed that maintaining visuospatial information in WM distorts it in consistent, stereotyped ways. Saccades to remembered target locations exhibit biases in their endpoint distributions, some of which are proportional to the length of the memory delay, but all of which vanish when saccade targets remain visible (White et al., 1994). The horizontal and vertical meridians of the visual field also appear to exert biases on the contents of spatial WM: remembered locations are repelled away from the meridians, towards the center of a quadrant (Huttenlocher et al., 1991; 2004; Merchant et al., 2004; Haun et al., 2005). These results suggest inhomogeneities in the representation of remembered locations across the visual field.

Additional psychophysical research has shown that attentional capabilities seem to be somewhat independent for different visual hemifields (Alvarez et al., 2012) and/or quadrants (Carlson et al., 2007; Liu et al., 2009), and that shifting the focus of attention across a meridian incurs a substantial reaction time penalty (Rizzolatti et al., 1987). If WM and attention are as similar as Postle's model claims, then they should share similar constraints. Furthermore, considering that homogeneity of SWM representations is taken for granted, it is quite possible that

WM representations of visual space exhibit hemifield or quadratic biases that have never been explicitly tested for.

1.9 Beyond the Single Neuron

Traditional neurophysiological experiments record the activity of a single neuron over many repetitions (or trials) of the same stimulus or behavioral condition. This activity is then averaged across trials to remove the trial-to-trial variability that is assumed to be extraneous noise. While this approach is the bedrock of neurophysiology and has yielded countless invaluable scientific insights, it has three major shortcomings: First, very few ecologically valid scenarios involve repeating many instances of the same action, using the repetitions to determine the prototypical representation of that action, and then accepting this prototype as the true outcome. The second shortcoming of serial recordings of individual neurons is that the neuronal computations underlying sophisticated behaviors such as WM require the coordinated activity of many neurons within and across brain networks (Quijan Quiroga and Panzeri, 2009; Cunningham and Yu, 2014). The third shortcoming is that recording individual neurons separately and then aggregating them into a “pseudopopulation” of many neurons assumes that trial-to-trial

variability of each individual neuron is independent of the rest; this assumption is incorrect, and carries significant implications that will be discussed shortly.

Researchers are not ignorant of these shortcomings. Technologies capable of recording the simultaneous activity of large neuronal populations *in-vivo* only became widely accessible to the neuroscience community in the 2000s. These advancements have created novel research opportunities for the field of neuroscience, which Cunningham and Yu describe lucidly:

These technological advances have enabled researchers to reconsider the types of scientific questions that are being posed and how neural activity is analyzed, even with classical behavioral tasks and in brain areas that have been studied for decades. Indeed, many studies of neural systems are undergoing a paradigm shift from single-neuron to population-level hypotheses and analyses...there are settings in which data fundamentally cannot be understood on a single-neuron basis (Cunningham and Yu, 2014).

WM and the PFC are a prime example of a “classical behavioral task in [a] brain area that has been studied for decades”. The preponderance of electrophysiological studies discussed here relied on single electrode recordings. And while the 40+ years of research into the neuronal mechanisms of WM have been extremely informative, single electrode recordings are fundamentally limited in their ability to investigate phenomena that only exist in populations of neurons.

Our understanding of the ensemble-level mechanisms of WM and how single neuron coding properties scale to larger populations comes largely from theoretical modeling and speculation based on results from large-scale recordings in other brain regions and investigating other behaviors.

One such prediction from modeling work is that the recurrent-excitatory lateral-inhibitory network structure thought to underlie WM maintenance should result in a stereotyped pattern of correlated firing rate variability between neurons—i.e. that the correlation in trial-to-trial variability between neurons changes depending on the relationship of the neurons' tuning (Constantinidis and Wang, 2004; Polk et al., 2012). Such a pattern can be characterized by two measurements: The first is signal correlation (r_{signal}), a measure of the similarity of two neurons' responses to a set of different stimuli or experimental conditions. The second is the spike count correlation (r_{sc} ; sometimes called noise correlation or spike rate correlation), which measures shared variability or co-fluctuation in firing rate between neurons that is not ascribable to changes in stimuli or behavioral conditions (Zohary et al., 1994; Averbeck et al., 2006; Cohen and Kohn, 2011). Formulae for these quantities are provided in the Methods sections of the relevant chapters.

1.10 Spike Count Correlations (r_{sc})

r_{sc} are often used to quantify functional connectivity (Cohen and Kohn, 2011), but they can also have a profound effect on the information content of a neuronal population, meaning that conclusions derived from analyses of pseudopopulations may be inaccurate (Shadlen and Newsome, 1994; Zohary et al., 1994; Shadlen and Newsome, 1998; Averbeck et al., 2006; Cohen and Kohn, 2011; Moreno-Bote et al., 2014) (Figure 1.4).

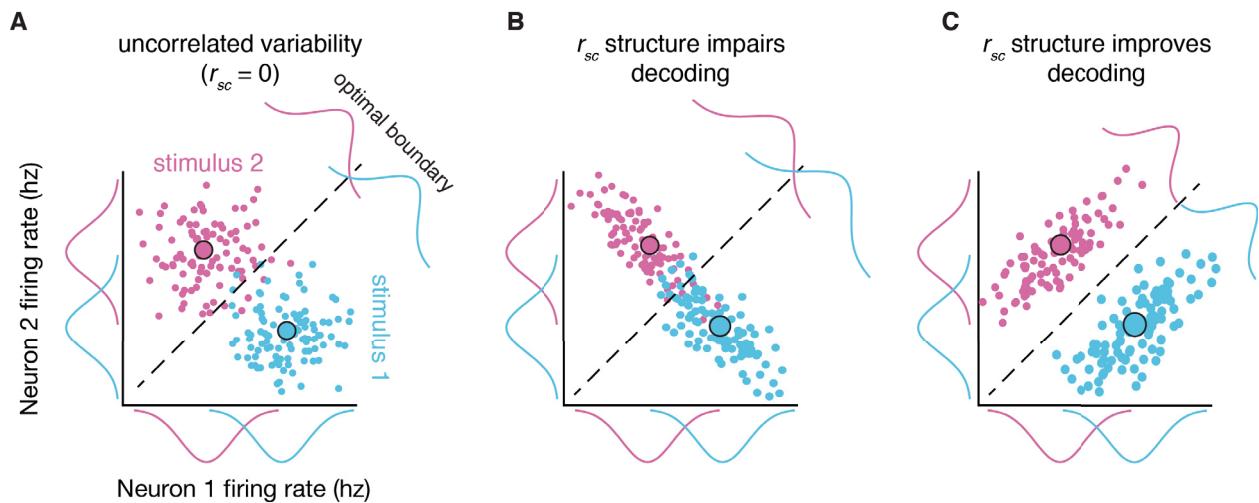


Figure 1.4: Effects of r_{sc} on decoding in a two-neuron example model. A) The firing rate of neuron 1 (x-axis) is plotted against the firing rate of neuron 2 (y-axis). The marginal distributions of firing rates for each neuron in response to each of two stimuli (pink vs. blue) are plotted along each axis. The smaller colored dots represent the firing rates for single trials, while the two larger dots represent the mean responses to the pink and blue stimuli. In this example, the variability of the two neurons' responses is uncorrelated (i.e. $r_{sc} = 0$). The dashed line denotes the optimal boundary for decoding the identity of the stimulus using the activity of the two neurons, and the joint firing rate distributions for each of the two stimuli as

projected along the discrimination boundary are shown at the top right. B) Same format as A), but the r_{sc} between the two neurons has changed to impair decoding, visible as the increased overlap of the joint distributions. Note that the mean responses and marginal distributions are identical to A). C) Same format as B), but the r_{sc} between the two neurons has changed to improve decoding. Again, the mean responses and marginal distributions remain identical to the previous examples.

In some circumstances, r_{sc} can impose a limit on the amount of information that can be represented in a neuronal population; beyond a certain number of neurons, adding additional neurons will not improve the capabilities of the population (Zohary et al., 1994). Another example of the importance of r_{sc} comes from studies of visual attention. The improvements in behavioral performance effected by visual attention were once thought to result from improvements in the signal-to-noise ratio of the firing rates of individual neurons in visual cortex (Reynolds and Chelazzi, 2004). However, recent studies have shown that the majority of the signal-to-noise improvement in the neuronal population comes from changes in r_{sc} (Cohen and Maunsell, 2009; Mitchell et al., 2009). However, these results provide limited insight into the effects of r_{sc} on WM coding in the PFC. This issue is of particular importance. Tuning-dependent changes in r_{sc} are predicted by the recurrent-excitatory lateral-inhibitory model of WM maintenance, but given that r_{sc} can dramatically influence neuronal coding, it is possible that the pattern of r_{sc} generated by WM maintenance could affect the coding of the information

maintained in WM. Unfortunately, there are relatively few studies investigating r_{sc} in the LPFC, and none that directly examine the effects of r_{sc} on WM coding (Constantinidis et al., 2001a; Constantinidis and Goldman-Rakic, 2002; Cohen and Kohn, 2011; Qi and Constantinidis, 2012; Katsuki et al., 2014; Tremblay et al., 2014; Markowitz et al., 2015). Previous studies of pairwise correlations are also difficult to extrapolate to larger neuronal ensembles, which have a complex, multidimensional r_{sc} structure ($n!/(2(n-2)!)$ pairs of neurons) that is insufficiently characterized by pairwise measurements alone (e.g. the mean r_{sc}) (Moreno-Bote et al., 2014). Thus it is currently unknown whether and how the r_{sc} structure modulates the fidelity of WM coding in LPFC neuronal ensembles.

1.11 Research Aims

Having provided a background on non-human primate working memory research and identified open questions in the field, I now present the primary aims of my research program:

Aim 1: Determine how ensembles of simultaneously-recorded neurons in LPFC area 8a represent SWM, and how it differs from single neuron representations.

Electrophysiological research has traditionally relied on recording from single (or a few) neurons at once, thus we are largely ignorant of the network properties

underlying SWM, and how it is represented by the coordinated activity of neuronal ensembles in LPFC area 8a. Complex neural computations are enacted through the interactions of many neurons in and across networks (Dayan and Abbott, 2001; Quiroga and Panzeri, 2009), thus we expected to observe properties that exist in simultaneously-recorded neural ensembles that are not visible in pseudopopulations. Specifically, we expected to observe changes in the ensemble r_{sc} structure resulting from the modulation of recurrent excitation and lateral inhibition predicted by models of LPFC structure and WM maintenance (Constantinidis and Wang, 2004; Polk et al., 2012). We also expected that the pattern of r_{sc} should influence the fidelity of ensemble WM representations.

Aim 2: Determine the magnitude and spatial scale of functional interactions in LPFC in the absence of stimulus input.

Based on the proposed network model of LPFC (Goldman-Rakic, 1995; 2011), we expected that functional interactions between neurons in 8a extend beyond the previously measured range of 1mm, and should reflect the presence of both excitatory and inhibitory connections between neurons, as measured using r_{sc} .

Aim 3: Quantify whether functional (i.e. task-related) properties of neurons in LPFC exhibit topography.

Abundant examples of topographic organization in other brain regions, theoretical models of LPFC structure (Goldman-Rakic, 1995), and qualitative reports of such organization in 8a (Suzuki and Azuma, 1983) led us to hypothesize that quantifiable topography should be present. Recording from larger neuronal ensembles and across broader areas of cortex could reveal functional topography in area 8a that had so far remained elusive.

Aim 4: Examine whether neuronal tuning for visual space during WM is biased by the meridians of the visual field.

Given the psychophysical literature on biases in SWM (Huttenlocher et al., 1991; 2004; Merchant et al., 2004; Haun et al., 2005), the relative hemifield-independence of WM capacity (Vogel and Machizawa, 2004; Delvenne, 2005; Buschman et al., 2011; Delvenne et al., 2011; Matsushima and Tanaka, 2014), the similarity between SWM and spatial attention (Miller and Cohen, 2001; Postle, 2006; Theeuwes et al., 2009; Smith and Schenk, 2012), and the independence of attentional resources across hemifields and/or quadrants, we hypothesized that SWM representations should be biased by quadrants/meridians of the visual field. Specifically, representations of locations within a quadrant should be more similar than equidistant representations across quadrants.

In order to address these aims, we used 96-channel microelectrode arrays (Figure 1.5; (Maynard et al., 1997; Normann et al., 1999)) to record from ensembles of LPFC area 8a neurons in two *Macaca fascicularis* monkeys while they performed two different tasks. These recordings constitute the basis of three separate experiments, each of which constitutes a chapter in this thesis.

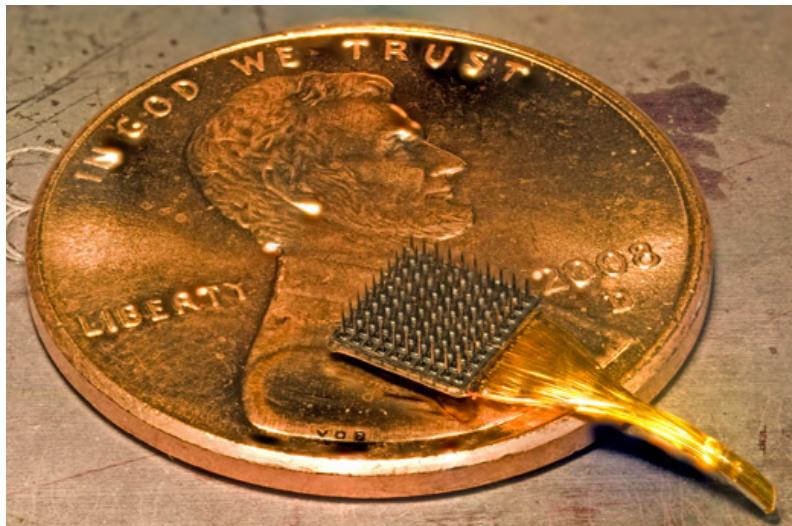


Figure 1.5: 96-channel Utah microelectrode array. This array is chronically implanted subdurally in order to record the spiking activity (i.e. action potentials) of large numbers of neurons simultaneously. Its capacity for large-scale simultaneous recordings makes it particularly well-suited for examining ensemble-level phenomena. Image courtesy of (Jones, 2009).

The first experiment addresses Aim 2, determining the magnitude and spatial scale of functional interactions in LPFC in the absence of stimulus input. In this experiment we determined the tuning properties of area 8a neuron during a simple visually-guided saccade task. We then examined the relationship between

neurons' tuning and r_{sc} during fixation, prior to the presentation of visual stimuli. We found that r_{sc} decreased as a function of distance between neurons, positive r_{sc} were stronger between similarly-tuned neurons, and negative r_{sc} were stronger between dissimilarly-tuned neurons. Most importantly, we found that r_{sc} between anatomically distant ($>3\text{mm}$) neurons with dissimilar tuning were predominantly negative, suggestive of tonic resting-state inhibition. This pattern of intrinsic functional connectivity supports the model of a recurrent-excitatory lateral-inhibitory network structure in LPFC.

Experiment 2 was designed to address three aims: Aim 1, to determine how ensembles of simultaneously-recorded neurons in LPFC area 8a represent SWM, and how it differs from single neuron representations; Aim 3, quantifying whether functional (i.e. task-related) properties of neurons in LPFC exhibit topography; and Aim 4, to examine whether neuronal tuning for visual space during WM is biased by the meridians of the visual field. In this experiment the animals performed an oculomotor delayed-response (ODR) task. We found that neuronal SWM representations were asymmetrically biased not to extend beyond the meridians of the visual field. This bias is present at multiple levels of examination: in single neuron firing rates, the r_{sc} structure, and neuronal ensemble representations. The

subjects' saccades also exhibited similar biases to those reported in the psychophysical literature. These results could be considered neuronal correlates of known biases in WM-based judgments of visual space. We also found that LPFC neurons are anatomically clustered in a non-retinotopic manner that partially reflects the organization of visual space.

The final experiment addressed Aim 1, determining how ensembles of simultaneously-recorded neurons in LPFC area 8a represent SWM and how it differs from single neuron representations, and Aim 3, quantifying whether functional (i.e. task-related) properties of neurons in LPFC exhibit topography. This experiment utilized the same ODR task as Experiment 2. We found that WM maintenance modulated the r_{sc} structure in manner indicative of increased coupling between similarly-tuned neurons and increased inhibition between dissimilarly-tuned neurons. We then examined the effects of the r_{sc} structure on WM coding, and found that it could facilitate or impair the readout of WM representations, depending on the size of the ensemble and tuning properties of its constituent neurons. Interestingly, ensembles of neurons that contained the most robust SWM representations were more anatomically dispersed than predicted by the statistics of the full recorded population of neurons. Finally, we found neurons that contained

little WM information when examined in isolation, but contributed to coding when part of an ensemble by shaping the ensemble's r_{sc} structure. This final result is a powerful example of a phenomenon that is inaccessible in individually-recorded neurons, and emphasizes the importance of multi-neuron recordings for elucidating the neuronal mechanisms of cognition.

CHAPTER 2

**SPIKE COUNT CORRELATION STRUCTURE REVEALS
INTRINSIC FUNCTIONAL TOPOGRAPHY IN PRIMATE
PREFRONTAL CORTEX AREA 8A**

Anatomical studies indicate that neurons in dlPFC are arranged in recurrent excitatory microcolumns of approximately 0.7mm, while lateral connections between neurons can extend up to 7mm. Prior experiments have shown that functional connectivity in dlPFC, as measured using spike count correlations (r_{sc}), extends up to 1 mm. The goal of this study was to examine the pattern of r_{sc} between neurons across larger distances. The results demonstrate a pattern of r_{sc} indicative of a recurrent excitatory, lateral inhibitory network structure. This chapter is adapted from Leavitt, M., Pieper, F., Sachs, A., Joober, R., Martinez-Trujillo, J.C. (2013) Structure of spike count correlations reveals functional interactions between neurons in dorsolateral prefrontal cortex area 8a of behaving primates, *PLoS ONE*, 8(4): e61503.

2.1 Abstract

Neurons within the primate dorsolateral prefrontal cortex (dlPFC) are clustered in microcolumns according to their visuospatial tuning. One issue that remains poorly investigated is how this anatomical arrangement influences functional interactions between neurons during behavior. To investigate this question we implanted 4 mm × 4 mm multielectrode arrays in two macaques' dlPFC area 8a and measured spike count correlations (r_{sc}) between responses of simultaneously recorded neurons when animals maintained stationary gaze. Positive and negative r_{sc} were significantly higher than predicted by chance across a wide range of inter-neuron distances (from 0.4 to 4 mm). Positive r_{sc} were stronger between neurons with receptive fields (RFs) separated by $\leq 90^\circ$ of angular distance and progressively decreased as a function of inter-neuron physical distance. Negative r_{sc} were stronger between neurons with RFs separated by $> 90^\circ$ and increased as a function of inter-neuron distance. Our results show that short- and long-range functional interactions between dlPFC neurons depend on the physical distance between them and the relationship between their visuospatial tuning preferences. Neurons with similar visuospatial tuning show positive r_{sc} that decay with inter-neuron distance, suggestive of excitatory interactions within and between adjacent microcolumns. Neurons with

dissimilar tuning from spatially segregated microcolumns show negative r_{sc} that increase with inter-neuron distance, suggestive of inhibitory interactions. This pattern of results shows that functional interactions between prefrontal neurons closely follow the pattern of connectivity reported in anatomical studies. Such interactions may be important for the role of the prefrontal cortex in the allocation of attention to targets in the presence of competing distractors.

2.2 Introduction

Neurons in the primate dorsolateral prefrontal cortex (dlPFC) are arranged in microcolumns spanning approximately 0.7 mm, wherein neurons with similar visuospatial tuning are interconnected with one another (Kritzer and Goldman-Rakic, 1995; Rao et al., 1999). This intrinsic connectivity pattern is thought to allow neurons to functionally interact with one another during behavior. However, measuring such interactions is challenging because one must simultaneously record the responses of multiple dlPFC neurons in behaving primates and compute measurements of functional interactions between them in behaving primates.

One commonly used measurement in primate electrophysiological studies is the Pearson's correlation coefficient (r), computed between the firing rates of two simultaneously recorded neurons over many trials (Bair et al., 2001). When r is

computed between neuronal responses over many trials of the same behavioral condition (noise correlations or r_{sc}), it provides a measurement of functional interactions between two units, which can be due to shared sensory inputs. In the absence of sensory inputs, r_{sc} could provide a measurement of functional connectivity between neurons (Palm et al., 1988; Zohary et al., 1994; Shadlen and Newsome, 1998; Averbeck et al., 2006). Previous studies have measured r_{sc} in several cortical areas of the macaque brain and reported a variety of results. In the visual cortex, it seems to be acknowledged that r_{sc} can hinder visual coding, and that cognitive variables such as attention can decrease r_{sc} (reviewed in (Cohen and Kohn, 2011)). In prefrontal neurons, however, this picture is less clear. One study in the frontal eye fields (FEF), a prefrontal area involved in the coding of gaze commands, has proposed that r_{sc} may reflect cooperation and competition between neurons within a network during cognitive tasks such as target selection (Cohen et al., 2010). However, the topography of such interactions and their dependence on sensory input into neurons have not been well documented.

A previous study used single electrodes to record the responses of dlPFC neurons and reported positively correlated firing during different periods of a delayed-match-to-sample task (Constantinidis and Goldman-Rakic, 2002). The r_{sc}

decreased with the distance between neurons and did not change during the different task periods. However, they did not explore inter-neuron distances larger than 1 mm. Two other studies recorded from the same region have reported similar results (Sakurai and Takahashi, 2006; Tsujimoto et al., 2008). Although these results have revealed that dlPFC neurons functionally interact with one another during behavior, they have also generated important questions. First, over what distances do interactions between dlPFC neurons occur? For example, it is known that microcortical columns in the dlPFC span over a range of ~0.7 mm but the pattern of collateral connections between dlPFC neurons could extend as far as 7 mm (Kritzer and Goldman-Rakic, 1995). The results reported by previous studies exploring distances below 1 mm could likely account for interactions between neurons within a microcolumn, but not between microcolumns. According to the pattern of collateral connections, one would anticipate that correlated firing would extend beyond 1 mm. Second, if neurons within the dlPFC interact through both excitatory and inhibitory connections, as suggested by models of the dlPFC circuitry (Wang et al., 2004), one would anticipate finding both positive and negative correlations. With the exception of one report (Constantinidis et al., 2002), which also limited their exploration of functional interactions to neurons located 0.2 to 0.3

mm apart, most studies have mainly reported positive r_{sc} . Third, the aforementioned studies used single electrodes and in some cases computed r_{sc} between units recorded from the same electrode and/or over a few number of trials. Thus, it is possible that these factors influenced the measured r_{sc} (Cohen and Kohn, 2011).

In order to investigate these and other related questions, we implanted a microelectrode array in dlPFC area 8a of two macaques and simultaneously recorded the activity of many neurons while the animals kept stationary gaze and waited for the presentation of a visual stimulus. By collecting a large number of trials in the absence of visual inputs into the neurons' RFs, we controlled for common sensory inputs into the units as the source of r_{sc} . We found that both positive and negative correlations were larger than predicted by chance over a wide range of inter-neuron distances (from 0.4 to 4 mm). Positive correlations were stronger between nearby neurons with visual RFs at similar locations and decreased progressively for neurons farther apart—up to the largest recorded distance of 4 mm. Most importantly, negative correlations were stronger between neurons with dissimilar RF locations and increased as the distance between the units increased.

2.3 Methods

We recorded the responses of neurons in the dlPFC of two behaving adult male *Macaca fascicularis* (“JL” and “F”).

2.3.1 Ethics Statement

All the experimental procedures were carried out in accordance with Canadian Council for Animal Care guidelines and were pre-approved by the McGill University Animal Care Committee. Animals were pair-housed in enclosures and interactive environmental stimuli were provided for enrichment. During experimental days, water was restricted to a minimum of 35ml/kg/day, which they could earn through successful performance of the task. Water intake was supplemented to reach this quantity if it was not achieved during the task, and water restriction was lifted during non-experimental days. The animals were also provided fresh fruits and vegetables daily. Body weight, water intake, and mental and physical hygiene were monitored daily. Blood cell count, hematocrit, hemoglobin, and kidney function were tested quarterly. If animals exhibited discomfort or illness, the experiment was stopped and resumed only after successful treatment and recovery. All surgical procedures were performed under general anesthesia. None of the animals were sacrificed for the purpose of this experiment.

2.3.2 *Task*

A custom computer program controlled the stimulus presentation and monitored eye position signals and behavioral responses. The animal initiated a trial by maintaining gaze within 2° on a central fixation spot (0.08 degrees²) and pressing a lever; fixation needed to be maintained until the end of a trial. After 650 ms of fixation, a sine-wave grating (2.5 Hz/deg, 1° diameter, vertical orientation) appeared at one of 40 locations (8 directions in 45°-steps, 5 eccentricities in 3°-steps), selected randomly, for 650 ms. After that period, the fixation spot disappeared and the animal had 600 ms to saccade toward the grating (Figure 2.1A). If the saccade landed on the grating, the animal received a juice reward and could initiate the next trial after 1 second. Fixation breaks during the trial or failure to saccade to the target resulted in immediate trial abortion without reward.

2.3.3 *Apparatus and recording procedures*

Prior to the experiments, the animals were implanted with 3 head-posts—one placed posterior to the supra-orbital ridge in the midline and the other two on the petrosal bones superior to the external occipital protuberance behind the left and right ears, respectively. The head posts interfaced with a head holder to fix the monkeys' heads to the chair during the recordings. Eye-positions were monitored

using an infrared video-based eye-tracker (EyeLink 1000, SR Research, Ontario, Canada) (Khayat et al., 2010).

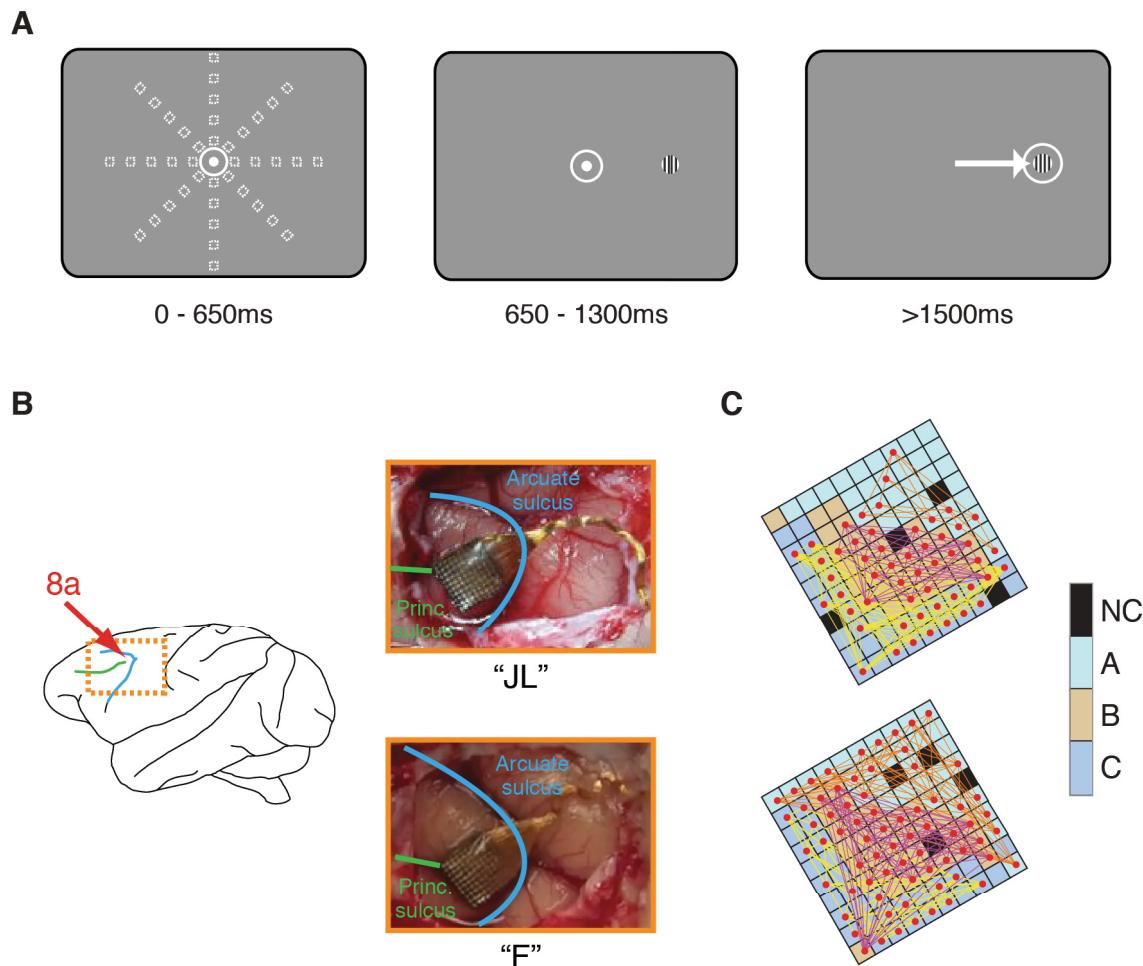


Figure 2.1: Locations of microelectrode array (MEA) implantations, task design, and correlation pairs on the MEA. A) Time-course of events during an example trial. The grey window indicates the screen, the white dot the fixation spot, the squares the positions of the grating, the white circles the fixation-window, and the arrow the subject's saccade. B) Pictures of the MEA implantation positions. Photographs were taken during the implantation procedure in monkeys 'F' and 'JL'. Principal and arcuate sulci are indicated. C) 32-channel recording blocks ('A', 'B', 'C') in each of the two implanted MEAs. Each square represents an electrode. Black squares are not

connected ('NC'). Red dots indicate single units and colored lines represent correlation pairings.

2.3.4 Microelectrode array (MEA) implant

We chronically implanted a 10×10 MEA (Blackrock Microsystems LLC, Utah, USA) (Maynard et al., 1997; Normann et al., 1999) in each monkey's left dlPFC—anterior to the knee of the arcuate sulcus and caudal to the posterior end of the principal sulcus (area 8a) (Figure 2.1B). Shortly, the surgical operation was carried out under general anesthesia with endotracheal intubation. An incision was made on the scalp with a scalpel and electrocautery. The scalp was retracted and the pericranium excised to limit biological reaction around the implant. A craniotomy was then fashioned using a high power drill (Anspach, FL, USA) over the desired implant location. Wet gelfoam was applied to the epidural edges for hemostasis. The array connector was fixed to the skull with cranial screws and the wires to the arrays were bent appropriately. The wire was secured in place with Silastic (silicon polymer, World Precision Instruments, FL, USA). The dura was opened with a #11 blade and Reynold scissors, and then tacked back with 4-0 vicryl sutures.

After exposure of the brain, the array was placed on the cortical surface and an array gun (Blackrock Microsystems LLC, Utah, USA) was set up in a stereotactic

frame. Once properly aligned, the gun was fired, pushing the array ~1 mm into the cortex. A duroplasty was then done using synthetic dura (Durepair, Medtronic, Inc. Minneapolis, MN, USA), and the bone flap (preserved in saline after the craniotomy) was replaced and secured back using cranial fixation plates and screws (Synthes, Inc. PA, USA). Gaps in the bone were filled with Silastic and the scalp was released from retraction. A small incision was made in the scalp over the connector, to allow the connector to be percutaneous. The scalp was closed in layers with buried vicryl sutures in the galea and staples to the skin. The animal fully recovered from the surgery within one week.

2.3.5 Recordings and spike detection

Data were recorded using a Cerebus Neuronal Signal Processor (Blackrock Microsystems LLC, Utah, USA) via a Cereport adapter. After 1x amplification in the headstage (ICS-96), the neuronal signal was band-pass filtered (0.3 Hz/1-pole, 7.5 kHz/3-pole, analog) and digitized (16 bit, 1 μ V per bit) at a sample rate of 30 kHz. Spike waveforms were detected by thresholding (manually adjusted to ~ -4 to -4.5x noise amplitude) the digitally high-pass filtered (250 Hz/4-pole or 750 Hz/4-pole) raw data.

The extracted spikes (48 samples at 30 kHz, ~1.6 ms) were re-sorted in OfflineSorter (Plexon, USA) using a T-Distribution E-M Algorithm in 3-dimensional feature space (Plexon Inc, TX). Only re-sorted spikes of single-neurons distinguishable from the multiunit cluster were included in the analysis. The electrodes on each MEA were separated by at least 0.4 mm and were organized into three blocks of 32 electrodes (A, B, C). Data were collected from one block during each recording session.

In order to obtain an estimate of the signal-to-noise ratio (separation of the spikes from the background noise) in our recordings, we computed a d' index for the distribution of spikes corresponding to each neuron and the associated noise (Equation 1).

$$d' = \frac{\text{mean}(\text{peak}) - \text{mean}(\text{baseline})}{\sqrt{\text{var}(\text{peak}) + \text{var}(\text{baseline})}} \quad (1)$$

The peak is the minimum/maximum voltage of a given spike distribution, and the baseline is the voltage at the spike onset (noise)—before the first deflection of the waveform (Figure 2.2A). Two d' measurements were computed for the first (negative) and second (positive) peaks of the spike waveform. The two were added to

obtain a global d' measurement of the distance between the peaks and baseline in standard deviation units of the latter (Figure 2.2B).

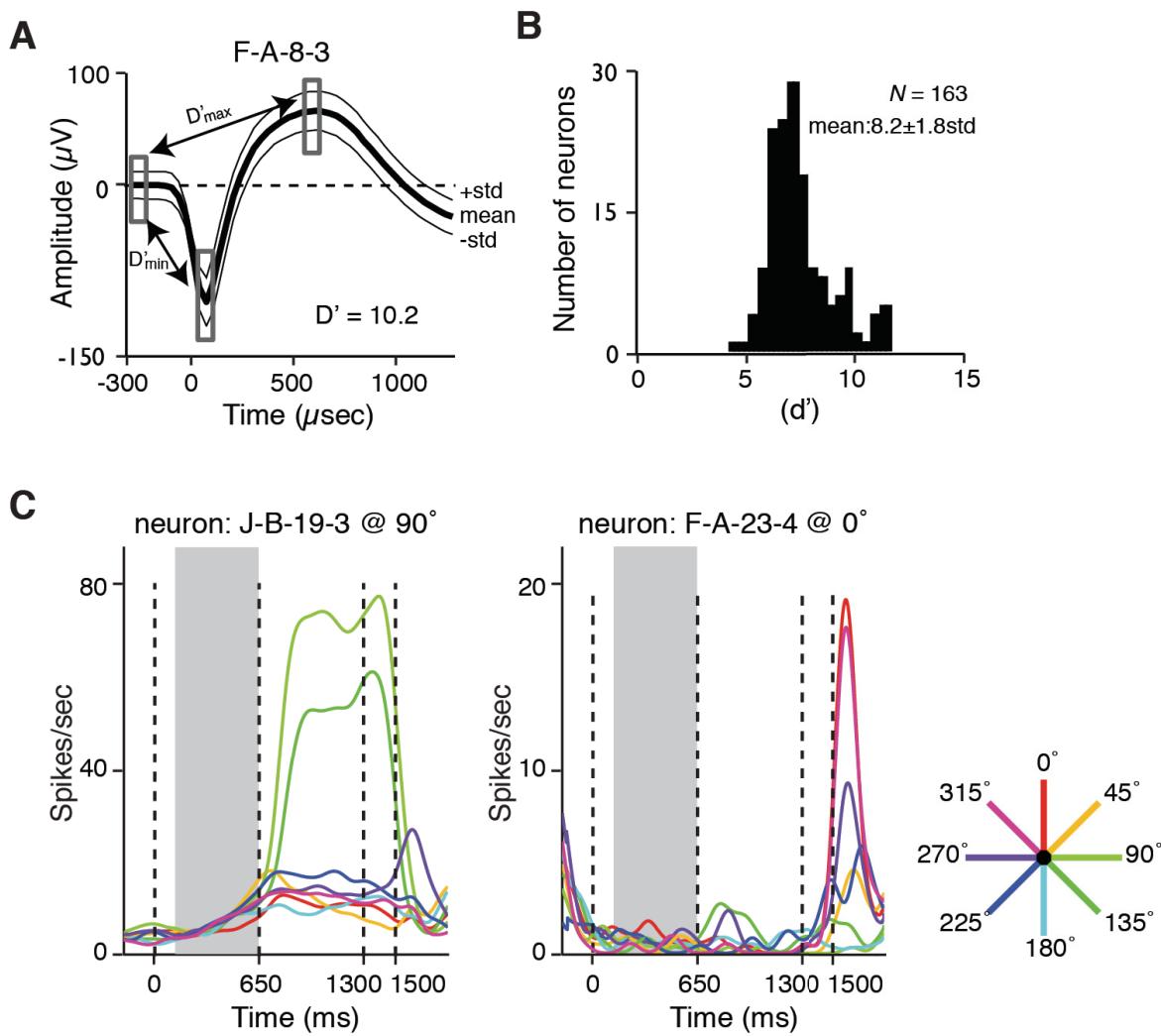


Figure 2.2: Example neuron and population signal-to-noise and selectivity. A) Example average waveform from a recorded neuron (solid line). The abscissa represents time from when the threshold was crossed in microseconds (μs), the ordinate signal amplitude in μV , and the thin lines the standard deviation across waveforms. Rectangles represent the time windows used to measure the standard deviations and to compute d' (see Methods). B) Histogram of d' distribution across all single-units. C) Examples of units' response (spike density function; $\sigma = 50 \text{ ms}$)

across the different time periods of the task and with different stimulus angular positions (color scale). The gray background indicates the analyzed trial period. The stimulus was presented at 650 ms, the “go” cue was presented at 1300 ms, and saccades were initiated at or after 1500 ms.

2.3.6 Estimation of inter-electrode distance

For electrodes located in the same column or row the interelectrode distance was computed by multiplying the length of a grid segment (0.4 mm) by the number of segments in between the electrodes. For electrodes located in different rows or columns, the Euclidean distance was computed (Figure 2.1C).

2.3.7 Data analysis

We collected spike data from a total of 163 single neurons (60 in JL, 103 in F) across 6 recording sessions (3 in JL, 3 in F)—one session in each of the MEA’s 3 electrode blocks. Neurons with a firing rate of less than 0.1 spikes/second during the 500 ms analyzed period were excluded from the analysis ($n = 39$), yielding a total of 124 single neurons (34 in JL, 90 in F). We then combined data in the 500 ms period preceding stimulus onset from all correct trials. During this period the animal had no information about the upcoming stimulus position and gaze was stationary. We assume that the behavioral state was identical across all trials and the animal was expecting the target onset. Furthermore, the absence of visual stimulation (except

for the fixation point, which was the same across trials) and measurable eye movements during that period rule out common visual input into the neurons as the source of spike count correlations. Next, we grouped units into simultaneously recorded pairs ($n=2003$) and computed Pearson's correlation coefficients (r_{sc}) between the z-scores of the units' spike counts (Ecker et al., 2010). In addition, we minimized the risk of falsely inflating the correlation values by excluding correlations between units on the same electrode ($n = 51$) from analysis (Ecker et al., 2010; Cohen and Kohn, 2011). These exclusion criteria yielded 1952 correlation pairs for analysis. Fisher's r -to- z transformation was applied to the correlation coefficients in order to stabilize the variance for hypothesis testing.

2.3.8 Visuospatial tuning

In order to determine whether a unit was visuospatially tuned for stimulus location, we first computed the mean firing rate for each target location during the 500 ms following the visual stimulus onset. We collected an average of 1010 trials per session (lower limit: 380; upper limit: 1300). We fitted the response data (i.e., mean firing rates) with a circular Gaussian function (**Equation 2**) and computed the goodness of fit (r^2) for each eccentricity (Treue and Martinez-Trujillo, 1999).

$$R(\theta) = B + A \sum_{i=-2}^2 \exp\left(\frac{-(\theta - \phi - 360i)^2}{\sigma^2}\right) \quad (2)$$

$R(\theta)$ is firing rate at target angular position θ . The parameters B , A , ϕ , and σ represent the baseline, height, preferred target position and tuning width, respectively. We also computed a selectivity index (**Equation 3**) for each neuron.

$$SI = \frac{fr_{preferred} - fr_{antipreferred}}{fr_{preferred} + fr_{antipreferred}} \quad (3)$$

$fr_{preferred}$ is the maximum firing rate evoked by a target and $fr_{antipreferred}$ is the minimum firing rate evoked by a target. Neurons were classified as visuospatially tuned for the target position if the adjusted r^2 of the fit was ≥ 0.75 and the selectivity index was ≥ 0.5 .

We collapsed across eccentricities, as an ANOVA did not show an effect of eccentricity between neurons with similar ($\leq 15^\circ$) angular tuning preferences ($p = 0.740$).

2.3.9 Control simulations

Variables such as firing rate may affect r_{sc} (Cohen and Kohn, 2011). As such, we obtained estimates of r_{sc} between simulated (independently homogeneously firing) Poisson neurons using the method described in (Heeger, 2000) to serve as a control

comparison. Briefly, the instantaneous firing rate of a neuron during a 500 ms duration trial was modeled as follows: a) time was subdivided into intervals of $\Delta t = 1$ ms, b) each interval duration was multiplied by the desired mean firing rate R ($\Delta t \times R$) to generate 500 identical values (for each Poisson neuron, R was estimated according to the firing rate of a matching recorded neuron), c) a sequence of 500 random values ($X(i)$) from a uniform distribution between 0 and 1 was generated, and d) if $X(i) \leq \Delta t \times R$ the instantaneous firing rate was set to 1, otherwise it was set to 0. For each neuron, we generated 500 spike trains of 500 ms duration each. This simulation produced a Poisson “mirror” neuron with the same firing frequency as the corresponding recorded neuron. When performing analysis on visuospatially-tuned neurons, the corresponding Poisson neurons were used for comparison. Correlations between pairs of Poisson neurons were also computed to obtain a measurement of correlations due to chance ($r_{sc,p}$).

2.4 Results

We analyzed neuronal activity during the 500 ms period preceding the visual stimulus onset (Figure 2.1A). This time period was chosen because: a) it allows measuring of spiking activity during periods of low firing in most of our units during many trials, b) it ensures a common visual input into the neurons RF is not

the source of correlated firing, and c) we could compare our results with the ones reported by previous studies in the same area (Constantinidis and Goldman-Rakic, 2002). In animal F spikes were isolated in 70 out of 96 active electrodes (82%), and in animal JL in 52 out of 96 active electrodes (54%) (Figure 2.1C). The estimated signal-to-noise ratio (d') in the included neurons ($n = 163$) was 8.2 ± 1.8 SD (see methods and Figure 2.2A, B).

The majority of the recorded neurons showed low spike rates preceding the target onset (Figure 2.2C). Some units showed an increase in firing rate when the visual target appeared at certain locations (left panel), while others did not (right panel). Similar response profiles have been previously described in prefrontal neurons (Takeda and Funahashi, 2002).

A representative example of r_{sc} ($n = 1024$ trials) between the firing rate of two units is illustrated in Figure 2.3A. Both units exhibit spike rates between 5 and 20 spikes per second. The bottom abscissa and left ordinate illustrate the actual firing rates, and the top abscissa and right ordinate show the corresponding z-transformed rates. For these two units, the estimated r_{sc} was 0.08 and was significantly different from zero (see Methods, $p = 0.009$, t -test). We followed the same procedure for each one of the recorded units. The mean of all Fisher-

transformed correlations (Figure 2.3B, red histogram) is significantly larger than zero (mean $r_{sc,all} = 0.031$, $p < 0.001$, t -test). The mean chance correlation (blue histogram) was not significantly different from zero ($r_{sc,p} = 3.0 \times 10^{-5}$, $p = 0.958$, t -test; Figure 2.3B, bottom panel) and was significantly smaller than $r_{sc,all}$ ($p < 0.001$, t -test, Bonferroni corrected). This shows that $r_{sc,all}$ in our sample was higher than expected by chance.

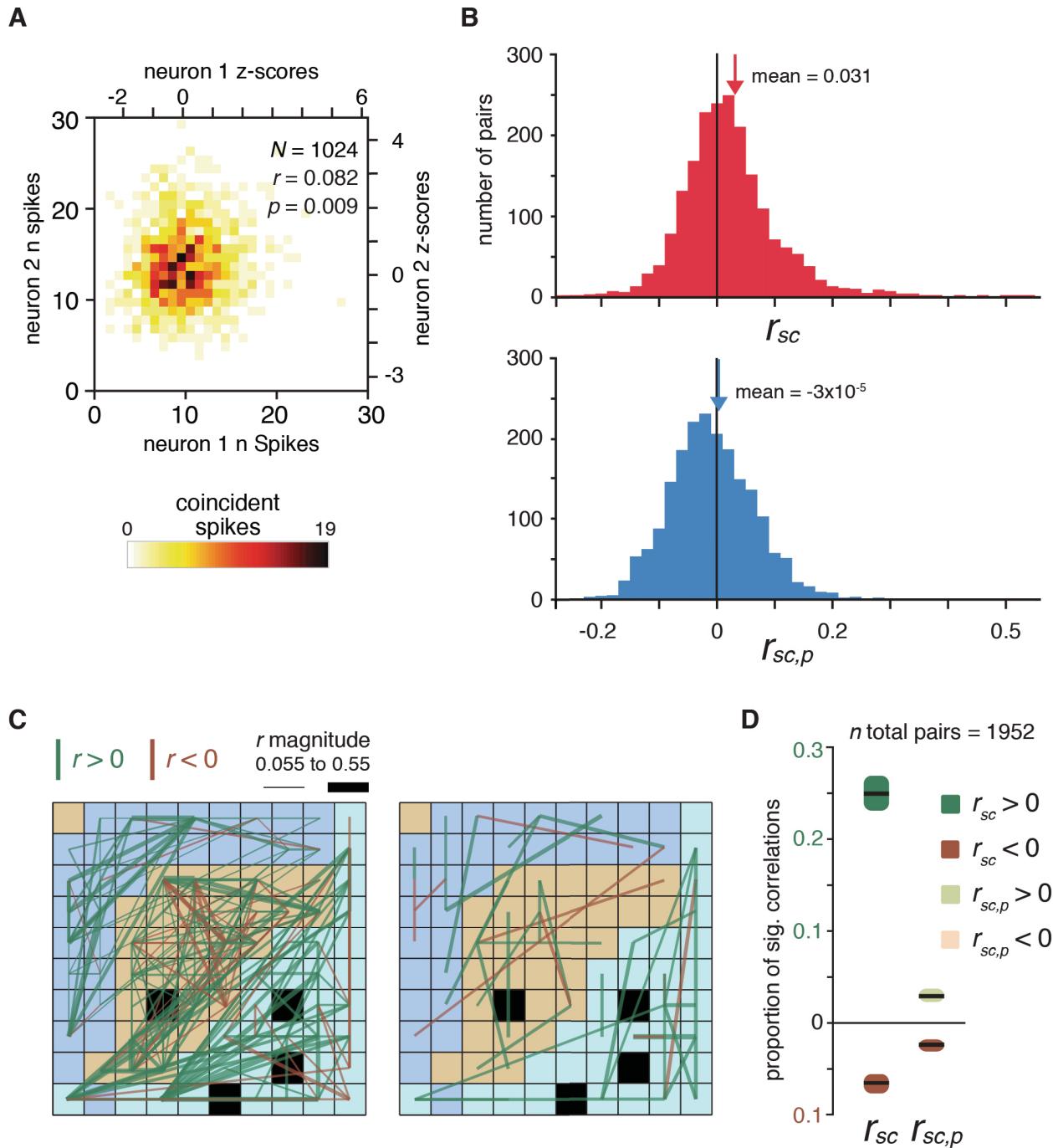


Figure 2.3: Spike count correlation (r_{sc}) example and population statistics. A) Example correlation between the responses of two neurons. The left x-axis and bottom y-axis represent the absolute number of spikes fired by each neuron in simultaneously recorded trials. The right x-axis and top y-axis show the z-transformed data. The color scale represents coincidences of spike counts. B)

Histograms displaying the distribution of Fisher-transformed spike count correlations in the recorded data (red) and in the Poisson simulated neurons (blue). The arrows indicate the means of the distributions. C) Distribution of correlations and neurons on the MEA from animal ‘F’. Each square represents an electrode, and each line represents a correlation between two neurons recorded on the constituent electrodes. The thickness of a line is proportional to the magnitude of the correlation (from 0.055 to 0.55) and the color represents the sign (brown = negative, green = positive). The panel on the left illustrates significant correlations from the recorded data (r_{sc}), while the right panel illustrates significant correlations occurring by chance ($r_{sc,p}$). D) The proportion of significant positive and negative correlations for the data shown in C. Observe that the ordinate of the negative correlations is inverted for a better comparison of the correlation magnitudes.

In order to better visualize the r_{sc} between different neurons in the MEA, we plotted the position of each unit as it was recorded on the MEA and joined significantly correlated pairs of neurons with a line. We then grouped the r_{sc} by sign—into positive and negative. Figure 2.3C illustrates $r_{sc,all}$ (left) and $r_{sc,p}$ (right) in three example networks recorded during different sessions (beige, yellow, and light brown backgrounds in the MEA layout). The proportion of significant positive r_{sc} was significantly larger than that observed by chance (0.250 vs. 0.029; $p < 0.001$, chi-square test, Bonferroni corrected). Interestingly, the proportion of significant negative r_{sc} was also significantly larger than predicted by chance (0.066 vs. 0.020; $p < 0.001$, chi-square test, Bonferroni corrected) (Figure 2.3D, bar graph). The proportion of positive r_{sc} was also significantly larger than the proportion of negative

r_{sc} ($p < 0.001$, difference of proportions from the same survey, Bonferroni corrected (Scott and Seber, 1983)).

2.4.1 Influence of distance between neurons and tuning properties on r_{sc}

Previous studies have found that for prefrontal neurons separated by less than 1 mm, the distance between them (inter-neuron distance) influences r_{sc} during periods of fixation (Constantinidis and Goldman-Rakic, 2002). Here, we examined whether this result can be generalized to longer distances. Indeed, after discretizing the distances into 0.5 mm bins, we found a significant decrease in mean r_{sc} with increasing inter-neuron distance ($p < 0.001$, F-test, see equation of the line fit to the data in Figure 2.4A). Additionally, we examined whether the proportion of significantly correlated pairs changed as a function of distance (Figure 2.4B). For this analysis, we again divided the significantly correlated units by the sign of the correlation. We observed a significant decrease in the proportion of significant positively correlated pairs as distance increased ($p = 0.019$, F-test, see linear equation and green dashed line in Figure 2.4B, top plot). For negative r_{sc} , the proportion of significantly correlated pairs did not change as a function of distance between neurons ($p = 0.876$, F-test see linear equation and brown dashed line in Figure 2.4B, bottom plot).

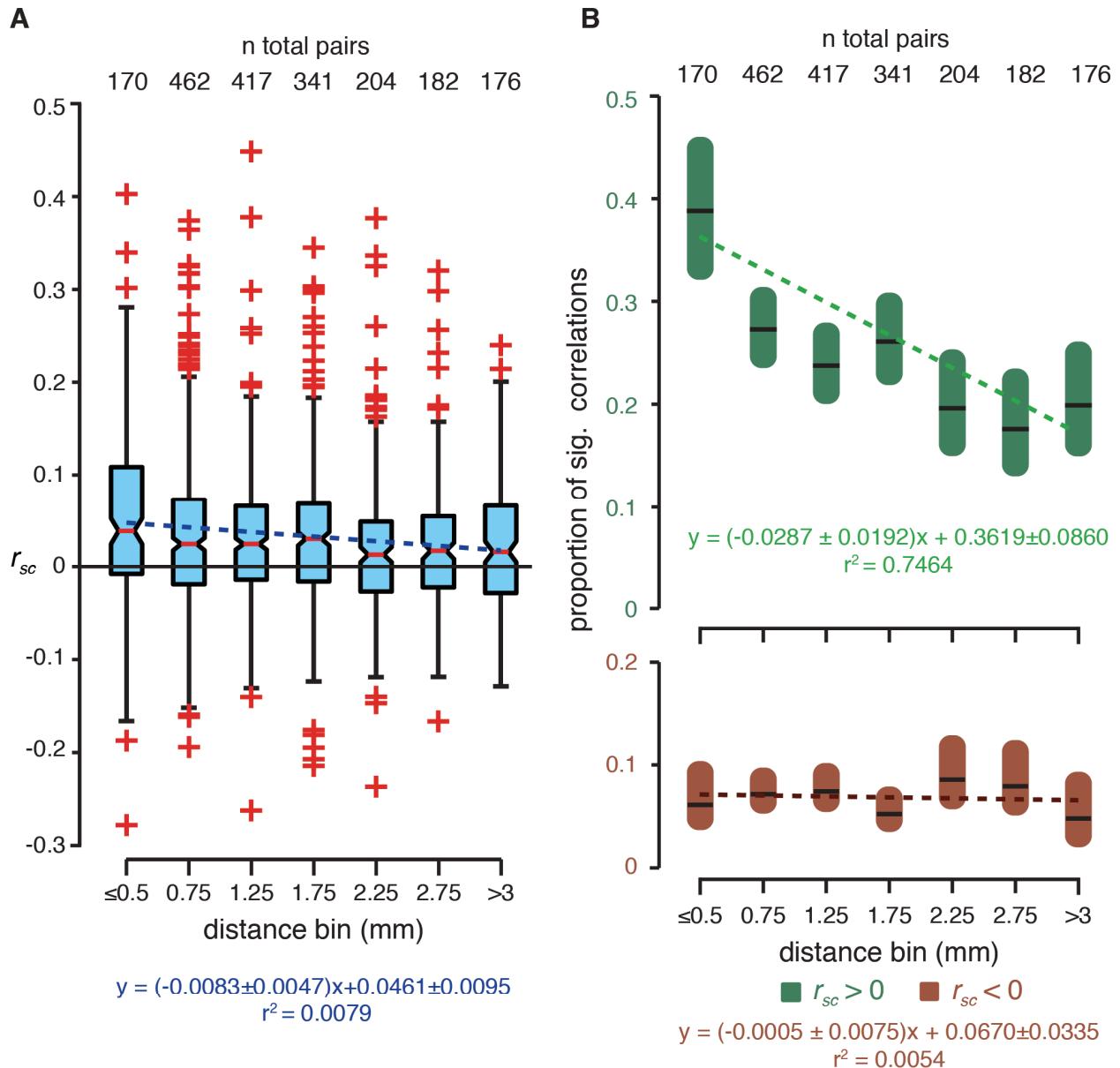


Figure 2.4: Spike count correlations between all neurons as a function of distance between neurons. A) Distribution of Fisher-transformed r_{sc} for all correlations (ordinate) as a function of distance between neurons (abscissa). Each bar indicates the median (red line), the box indicates the center 50%, and the whiskers extend to approximately $\pm 2.7\sigma$. Crosses indicate outliers beyond this interval. The green line and equation represent the best line fit through the data. The 95% confidence intervals for the slope are indicated. The total number of pairs is shown on the top. B) Proportion of significant positive and negative correlations as a function of

distance between units for r_{sc} (see legend). The green and red lines represent best-fit lines for these proportions. The equations for the two lines and 95% confidence intervals for the slopes are shown. There is a significant negative effect of distance on the proportion of significant positive correlations, and no significant effect of distance on the proportion of significant negative correlations ($p = 0.019$ and $p = 0.876$, respectively, F-test for both). The colored regions indicate the 95% confidence intervals for the proportions. The total number of pairs is shown on the top.

Another variable that may affect spike count correlations between neurons is their tuning for the stimulus position (Constantinidis and Goldman-Rakic, 2002). To examine this issue we divided our sample into neurons tuned for the position at which the stimulus would later appear, and untuned neurons (see **Methods**). We then organized the neuronal pairs into three categories: a) both neurons were tuned, b) one neuron was tuned, and c) neither neuron was tuned. Correlations between tuned pairs were further subdivided by angular difference in preferred stimulus location (RF location): a) neurons with a difference in RF location of either $\leq 90^\circ$ ('similar' preference group) or b) $> 90^\circ$ ('dissimilar' preference group). Figure 2.5A-E depict the same example networks from Figure 2.3 decomposed using these two criteria (large panels). At least three main observations can be made in Figure 2.5: First, the amount of significant r_{sc} is larger when at least one neuron of a correlation pair is tuned relative to when both neurons are not visuospatially tuned. Second, the amount of significant r_{sc} in tuned neurons decreases for neurons with

distant RF locations (large panels in first column). Third, there are fewer correlations expected by chance than the quantity observed in the sample (small panels nearby the large ones).

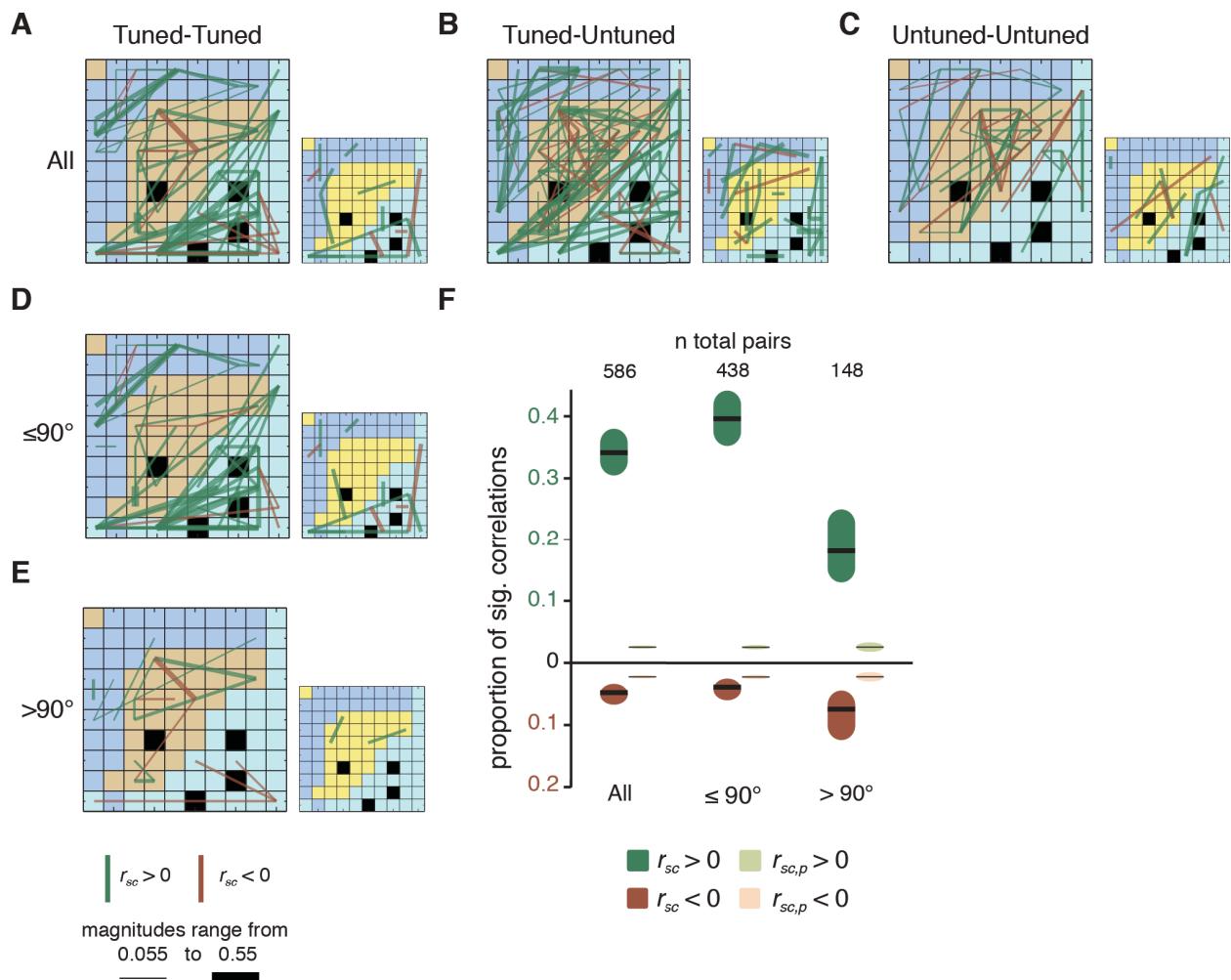


Figure 2.5: Spike count correlations sorted by neurons' visuospatial tuning preferences. Like Figure 2.3, correlations are visualized on the MEA. Each square represents an electrode, and lines between the electrodes represent correlations between neurons recorded on those electrodes. The thickness of the line is proportional to the magnitude of the correlation (from 0.055 to 0.55) and the color represents the sign (brown = negative, green = positive). The large panels represent r_{sc} and the small panels $r_{sc,p}$. The row displays data from A) tuned-tuned, B) tuned-

untuned, and C) untuned-untuned pairs. The column displays data from A) all tuned pairs, D) tuned pairs with RFs separated by less than or 90° , and E) RFs separated by more than 90° (bottom). F) The bar graph displays the proportion of significant correlations for all the tuned-tuned pairs (see legend).

The mean spike count correlation between tuned units ($r_{sc,tuned}$) was significantly larger than in the entire sample ($r_{sc,all} = 0.031$, $r_{sc,tuned} = 0.050$, $p < 0.001$, t -test, Bonferroni corrected). This finding was not due to significant differences in firing rate between tuned neurons and the whole population (la Rocha et al., 2007) during the analyzed period ($p = 0.706$, t -test). Thus, we concentrated on the subset of correlations in which both neurons were tuned ($r_{sc,tuned}$). In this group the mean $r_{sc,tuned}$ was significantly larger than that predicted by chance ($r_{sc,tuned} = 0.050$, $r_{sc,tuned,p} = 0.0002$, $p < 0.001$, t -test, Bonferroni corrected).

When separating the tuned pairs based on difference between angular RF location, we found that r_{sc} in the ‘similarly tuned’ group ($r_{sc,similar} = 0.062$) were significantly larger than those in the ‘dissimililarly tuned’ group ($r_{sc,dissimilar} = 0.012$, $p < 0.001$, t -test, Bonferroni corrected). More importantly, both the ‘similar’ and ‘dissimilar’ groups had a significantly larger proportion of significant, positive r_{sc} than predicted by chance ($p_{positive,similar} < 0.001$ and $p_{positive,dissimilar} < 0.001$, chi-square test and Bonferroni corrected for both groups), but only the ‘dissimilar’ group had a

significantly larger proportion of significant, negative r_{sc} than predicted by chance ($p_{negative,similar} = 0.058$ and $p_{negative,dissimilar} < 0.001$, chi-square test and Bonferroni corrected for both groups) (bar graphs in Figure 2.5F). Furthermore, the ‘similar’ group had a larger proportion of significant positive r_{sc} than the ‘dissimilar’ group ($p < 0.001$, chi-square test, Bonferroni corrected). While there was no significant difference between the proportion of significant negative correlations in the ‘similar’ and ‘dissimilar’ groups ($p = 0.487$, chi-square test, Bonferroni corrected), overall the proportion of significant positive r_{sc} decreased and the proportion of significant negative r_{sc} increased in the ‘dissimilar’ relative to the ‘similar’ group.

We examined the relationship between inter-neuron distance and r_{sc} in tuned pairs. In this group the linear decrease in correlation as a function of distance has a larger y-intercept than that for the whole population (compare Figure 2.4A and Figure 2.6A, $p < .01$, comparison of Bonferroni corrected 95% confidence intervals). When dividing the selective correlations into the ‘similar’ and ‘dissimilar’ RF groups, we did not find a significant difference between the slope or y-intercept of the fitted lines (Figure 2.6B and C, $p > .05$, comparison of Bonferroni-corrected 95% confidence intervals). However, the line fitted to the ‘dissimilar’ group indicates negative values of r_{sc} for the most physically distant pairs, suggesting that the firing

of physically distant neurons with dissimilar RF locations is negatively correlated. While the medians of the two largest distance bins were negative, they were not significantly negative, likely due to the small sample sizes ($n = 14$ for both distances, $p = .99$, $p = .079$ for the second largest and largest distances, respectively, Wilcoxon signed rank test).

We also computed the proportion of significantly correlated pairs as a function of distance. Tuned neurons showed a decrease in the proportion of significant positive r_{sc} as a function of inter-neuron distance ($p = 0.001$, F-test, see equation in Figure 2.6D, top panel). Interestingly, the proportion of significant negative r_{sc} increased as the distance between units increased ($p < 0.037$, F-test, see equation in Figure 2.6D, bottom panel), a trend not observed in the sample that includes non-selective neurons.

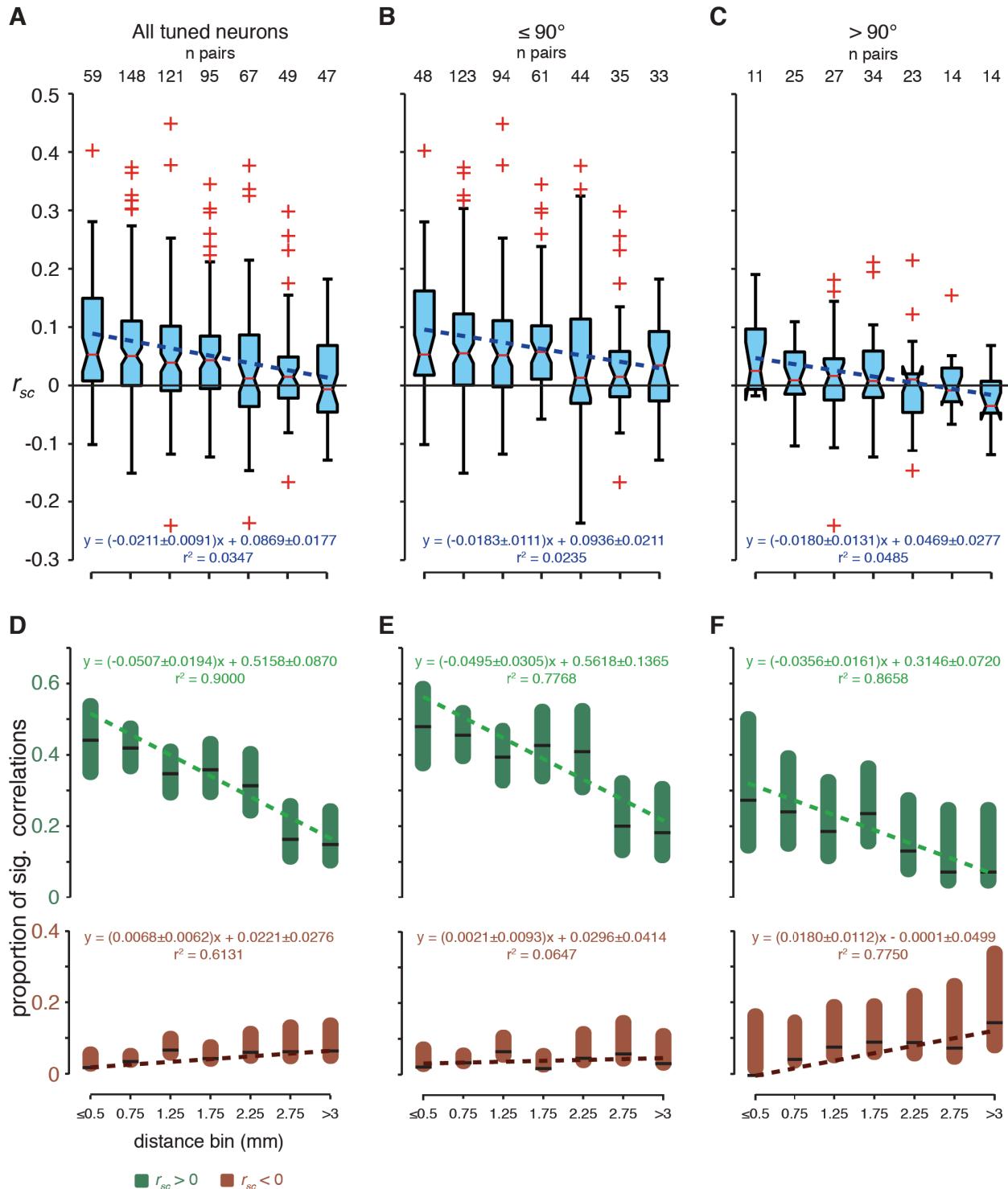


Figure 2.6: Spike count correlations between visuospatially-tuned neurons as a function of distance between neurons. Distributions of r_{sc} (ordinate) as a function of distance between neurons (abscissa). A) Displays all neurons, B) neurons with RFs

$\leq 90^\circ$ apart, and C) neurons with RFs $> 90^\circ$ apart. D-F) Proportion of significant positive and negative r_{sc} and $r_{sc,p}$ (ordinate, see color legend) as a function of distance between neurons (abscissa), organized by distance between RF as in A-C. All symbols are the same as in Figure 2.4.

The proportion of significantly correlated pairs was again divided into ‘similar’ and ‘dissimilar’ groups based on the distance between the two neurons’ RFs. We found that for the ‘similar’ group the proportion of significant positive correlations decreased as neurons became farther apart ($p = 0.009$, F-test see equation in Figure 2.6E, top panel). However, the proportion of negative correlations showed no significant trend ($p = 0.580$, F-test, see equation in Figure 2.6E, bottom panel). For pairs with RFs located more than 90° apart we observed that the proportion of significant positive correlations decreased as a function of distance ($p = 0.002$ F-test, see equation in Figure 2.6F, top panel). What may be the most surprising result regards the proportion of significant negative correlations, which increases as a function of distance ($p = 0.009$, F-test, see equation in Figure 2.6F, bottom panel).

To our knowledge, this is the first report in this cortical region of significant negative correlations that increase in frequency as a function of inter-neuron distance over distances larger than 1mm.

2.4.2 Dissociating proximity and tuning similarity

In order to dissociate the effects of inter-neuron distance and tuning similarity on r_{sc} we applied a multiple linear regression procedure using inter-neuron distance (in mm), RF location difference (in degrees), and the interaction of inter-neuron distance and RF location difference as the predictor variables, and r_{sc} as the response variable. We found that inter-neuron distance ($\beta_1 = (1.87 \pm 1.22) \times 10^{-2}$) and RF location difference ($\beta_2 = (4.97 \pm 2.97) \times 10^{-4}$) were both significant predictors of r_{sc} but that the interaction of the two variables was not ($\beta_3 = (1.48 \pm 16.12) \times 10^{-5}$). The y-intercept was significantly greater than zero ($\beta_0 = 0.107 \pm 0.022$) and the overall model fit was $R^2 = 0.097$ ($p < 0.001$). This result indicates that each variable alone is a predictor of r_{sc} , however their multiplicative interaction is not.

2.5 Discussion

Our study measured spike count correlations between the firing of macaque area 8a neurons using a chronically implanted MEA. We found that in the absence of visual stimulation: a) there were significant positive and negative correlations between neurons that extend over distances ranging from 0.4 to 4 mm; b) such correlations were significantly stronger between visuospatially-selective neurons; c) neurons with RFs $\leq 90^\circ$ apart show a significantly larger proportion of positive correlations

than units with RFs $> 90^\circ$ apart, and such correlations decreased in magnitude as a function of inter-neuron distance; and d) neurons with RFs $> 90^\circ$ apart showed a relatively larger proportion of negative correlations than those with RFs $\leq 90^\circ$, and such correlations increased in magnitude as a function of inter-neuron distance.

2.5.1 Previous correlation studies in dlPFC

Although several studies have examined spike count correlations in different brain areas (reviewed in (Averbeck et al., 2006) and (Cohen and Kohn, 2011)), only a few studies have measured spike count correlations in dlPFC area 8a of behaving monkeys (Constantinidis et al., 2001a; Constantinidis and Goldman-Rakic, 2002; Sakurai and Takahashi, 2006; Tsujimoto et al., 2008). So far, we have not found any study in this brain area that has recorded from multiple neurons using MEAs chronically implanted in the cortex. MEAs have several advantages relative to single electrode recordings. First, they allow precise identification of the targeted brain area because sub-dural implantation is done while visualizing the anatomical landmarks that identify the area (arcuate and principal sulci, (Petrides and Pandya, 2007)). Second, they allow a precise mapping of the recording sites relative to such landmarks that does not change within or between sessions, and thus facilitates the anatomical reconstruction of the explored area (Figure 2.1B and C). Third, they

provide more stable isolation of single units during a session since the arrays are fixed on the cortex rather than attached to the skull, as in single electrode recordings (Normann et al., 1999). This guarantees that small movements of the brain relative to the skull do not produce movement of the electrodes.

Unlike previous studies, we explored correlations between neurons located as far as 3-4 mm apart. For example, (Constantinidis and Goldman-Rakic, 2002) and (Constantinidis et al., 2002) explored inter-neuron distances of up to 1 mm, and (Sakurai and Takahashi, 2006) distances up to 0.5 mm. We measured correlations between neurons separated by distances from 0.4 to 4 mm. Considering that the width of a prefrontal microcolumn is about 0.7-0.9 mm (Kritzer and Goldman-Rakic, 1995; Hirata and Sawaguchi, 2008), these studies could not isolate functional interactions between neurons located in adjacent microcolumns but likely limited their findings to neurons located within the same or in nearby microcolumns. Our results agree with the results reported by these authors for distances smaller than 1 mm. Moreover, we found that the proportion of both positive and negative correlations were greater than zero for distances up to 3 mm and higher. Interestingly, negative correlations only became apparent between tuned neurons

at distances greater than 1.5-2 mm; hence, these correlations could not be observed in the aforementioned studies.

A novel finding of our study is the pattern of negative correlations shown in Figure 2.5 and Figure 2.6. Proportions of negative correlations significantly higher than predicted by chance were mainly found between pairs of neurons with RFs located more than 90° apart (Figure 2.5). Remarkably, such a proportion increased with the distance between neurons (Figure 2.6). A previous study has reported negative correlations between FEF neurons (Cohen et al., 2010) with non-overlapping RFs but did not quantify the trend that correlations increase as a function of interneuron distance. One factor that may have contributed to the novelty of our results is that previous studies used a smaller number of trials, thus reducing the power of statistical tests. For example Constantinidis and Goldman-Rakic (Constantinidis and Goldman-Rakic, 2002) used an average of ~10 trials; Sakurai and Takahashi (Sakurai and Takahashi, 2006) used a larger number of trials (240 to 270), but they did not explore distances beyond 1 mm. We used an average of 1010 trials and explored a wider range of distances. The increase in the proportion of significant negatively correlated pairs with increasing distance

between units cannot be an artifact associated with trial number since this variable was similar for all unit pairs across distances.

Because negative correlations occurred in the absence of visual inputs into the neurons RFs when the animals had no information about the upcoming stimulus location, they likely reflect direct or indirect interactions between neurons rather than a common sensory input to their RFs. These interactions may reflect an anatomical and functional organization of area 8a (i.e., units closer together with similar coding preferences share excitatory connections, and units far apart with dissimilar preferences share inhibitory connections; (Rao et al., 1999)). It may be possible that the correlations were due to inputs from other non-sensory brain areas reflecting preparation for the task and/or the attentional state of the animal. Although this is a possibility, such inputs must be organized according to the neurons' preferences within area 8a (RF similarity) and according to the distance between them to produce the pattern of correlations isolated in our data. Thus, we consider it more likely that the isolated pattern reflects the intrinsic functional connectivity between neurons within the area.

One previous study used crosscorrelation techniques to explore functional interactions between neurons located 0.2 to 0.3 mm apart and reported a trough in

the crosscorrelation histogram for pairs of neurons with similar tuning (Constantinidis et al., 2002). They interpreted this finding as evidence of inhibitory interactions between units that may play a role in functions such as memory maintenance. In our study we found that negative correlations between tuned neurons located closer together were near chance values (see Figure 2.6). However, we did not explore inter-neuron distances as short as 0.2 mm. It is possible that the short-range inhibitory interactions reported by these authors are limited to such short distances. This issue may require further investigation.

More importantly, our results indicate that functional interactions between dlPFC area 8a neurons extend beyond the previously reported distance of 0.9-1 mm (Constantinidis and Goldman-Rakic, 2002). Since the width of a dlPFC microcolumn has been estimated as approximately 0.7-0.9 mm (Levitt et al., 1993; Kritzer and Goldman-Rakic, 1995), our results may reflect functional interactions between different microcolumns, and agree with reports of dlPFC modules (Bugbee and Goldman-Rakic, 1983; Selemon and Goldman-Rakic, 1988; Constantinidis et al., 2001a; Wang et al., 2004). Moreover, the pattern of positive and negative correlations shown in Figure 2.6 suggests that neurons in nearby microcolumns mainly interact through excitatory connections while neurons in clusters far away

interact through inhibitory connections. These may underlie the competitive interactions between neuronal representations that are thought to play an important role in target selection and the allocation of attention (Desimone and Duncan, 1995; Everling et al., 2002; Szabo et al., 2004; Messinger et al., 2009; Lennert and Martinez-Trujillo, 2011).

2.5.2 Comparison with studies of r_{sc} in other brain areas

The majority of studies examining spike count correlations in primates have been conducted in visual areas ((Gawne et al., 1996; Reich et al., 2001; Kohn and Smith, 2005; Cohen and Newsome, 2008; Gutnisky and Dragoi, 2008; Smith and Kohn, 2008); see (Cohen and Kohn, 2011) for a review). Our results share similarities and differences with the results reported by these studies. For example, the decrease in positive correlations as a function of distance between neurons seems ubiquitous, as well as the changes in correlations as a function of neuronal tuning properties. Our results also agree with those of a previous study in area V1 that reported correlations extending over several millimeters (Smith and Kohn, 2008). However, different from our study, they mainly reported positive correlations. This apparent discrepancy with our results may be explained by differences in the intrinsic connectivity pattern of areas 8a and V1. For example, the granular structure and

connections with thalamic nuclei differ between visual and prefrontal cortices (Fuster, 2008; Xiao et al., 2009).

In visual areas, correlated firing between neurons with similar coding preferences may hinder stimulus coding (Zohary et al., 1994). A solution to this problem may be to decrease the correlated firing during behavioral tasks through mechanisms such as attention and learning (Cohen and Maunsell, 2009; Mitchell et al., 2009; Gu et al., 2011). However, in the dlPFC, correlation studies have not reported such changes in correlation across behavioral states (Constantinidis et al., 2001a; Constantinidis and Goldman-Rakic, 2002; Sakurai and Takahashi, 2006; Tsujimoto et al., 2008). Interestingly, a study in the FEF (located posterior to our recording sites) reported a pattern of negative and positive correlations as a function of RF distance between units and suggested this pattern reflects cooperation and competition between neurons during target selection (Cohen et al., 2010). Further investigation is needed to clarify whether correlations in the dlPFC change as a function of behavioral states.

Interestingly, it has been proposed that in the visual cortex the structure of the correlations rather than their absolute value determines choice probability during certain tasks (Nienborg and Cumming, 2010). Our results identify a

correlation structure in area 8a that seems to reflect the pattern of direct or indirect connectivity between neurons. This basic correlation structure could be modified during behavior depending on the input signals and computations performed by individual neurons and the entire area network.

2.5.3 Factors that affect r_{sc}

Several factors may affect the computation of spike count correlations (Cohen and Kohn, 2011). Amongst them are: response strength, time period for counting spikes, spike sorting, and fluctuation in behavioral state. Differences in response strength between neurons cannot explain the pattern of correlations reported in our study since neuronal firing rates did not differ across distances and RF locations. The time period used to count the spikes also cannot explain our results, because it was identical for all pairs. Spike sorting errors are unlikely the explanation to our results. First, we have provided a measure of signal-to-noise ratio in our sample of recorded neurons. Second, we excluded neurons recorded from the same electrode from the analysis. Finally, one may argue that we may have measured multiunit activity and therefore inflated the correlations. Although this cannot be fully ruled out, we consider it unlikely because we only included units with waveforms that were clearly classified as single neurons by the sorting algorithm. Nevertheless,

even if such a factor may have influenced our results, the pattern of positive and negative correlations across distances cannot be explained by spike sorting errors. Fluctuations in the animals' behavioral state also cannot explain our findings since our task was the same across all recording sessions.

2.6 Conclusion

In sum, our results demonstrate a pattern of spike count correlations as a function of physical distance between units and distance between their RFs in dlPFC area 8a. They suggest that functional interactions between neurons in this area extend over multiple millimeter distances, likely reflecting the interactions between cortical micro-columns. These may facilitate, under certain behavioral conditions, the competition between neurons holding neural representations of different objects or locations within the area topographic map.

2.7 Acknowledgements

The authors would like to acknowledge Ms. Megan Schneiderman, Mr. Walter Kucharski, and Mr. Stephen Nuara for their assistance.

CHAPTER 3

A QUADRANTIC BIAS IN PREFRONTAL REPRESENTATION OF VISUAL-MNEMONIC SPACE

The previous study reported a pattern of r_{sc} in dlPFC area 8a indicative of a recurrent excitatory, lateral inhibitory network structure. It is thought that this network structure is necessary for sustaining neuronal activity in the absence of stimulus input, which is a neuronal correlate of working memory. WM-based estimations of visual space are known to be biased by meridians of the visual field, so we sought to determine whether area 8a WM representations are also subject to such biases. We found that single neuron firing rates, the ensemble r_{sc} structure, and the simultaneous ensemble activity were all non-linearly biased by the meridians of the visual field. We also found that WM selectivity is anatomically clustered in a non-retinotopic manner. This chapter is adapted from Leavitt, M., Pieper, F., Sachs, A., Martinez-Trujillo, J.C. A quadratic bias in prefrontal representation of visual-mnemonic space (2017), *Cerebral Cortex*.

3.1 Abstract

Single neurons in primate dorsolateral prefrontal cortex (dLPFC) are known to encode working memory (WM) representations of visual space. Psychophysical studies have shown that the horizontal and vertical meridians of the visual field can bias spatial information maintained in WM. However, most studies and models have tacitly assumed that dLPFC neurons represent mnemonic space homogenously. The anatomical organization of these representations has also eluded clear parametric description. We investigated these issues by recording from neuronal ensembles in macaque dLPFC with microelectrode arrays while subjects performed an oculomotor delayed-response task. We found that spatial WM representations in macaque dLPFC are biased by the vertical and horizontal meridians of the visual field, dividing mnemonic space into quadrants. This bias is reflected in single neuron firing rates, neuronal ensemble representations, the spike count correlation structure, and eye movement patterns. We also found that dLPFC representations of mnemonic space cluster anatomically in a non-retinotopic manner that partially reflects the organization of visual space. These results provide an explanation for known WM biases, and reveal novel principles of WM representation in prefrontal neuronal ensembles and across the cortical surface, as

well as the need to re-conceptualize models of WM to accommodate the observed representational biases.

3.2 Introduction

Working memory (WM) is the ability to transiently maintain and manipulate information that is no longer available in the environment (Baddeley and Hitch, 1974). It is strongly correlated with measures of human intelligence, and a critical foundation for complex behaviors (Fuster, 1973; Engle et al., 1999; Miller and Cohen, 2001). Sustained neuronal activity in the absence of stimulus input is considered a neural mechanism for WM (Hebb, 2005). Indeed, single neurons in dorsolateral prefrontal cortex (dLPFC) and other regions of the macaque brain exhibit spatially-selective sustained activity during WM maintenance (Fuster and Alexander, 1971; Niki, 1974a; Batuev, 1986; Gnadt and Andersen, 1988; Funahashi et al., 1989; Constantinidis and Procyk, 2004).

Psychophysical studies have shown that maintaining visuospatial information in WM subjects it to stereotyped distortions, or biases. Saccades to remembered target locations show biases in their endpoint distributions that vanish when saccade targets remain visible (White et al., 1994). The horizontal and vertical meridians of the visual field also appear to exert biases on the contents of spatial

WM: remembered locations are repelled away from the meridians, towards the center of a quadrant (Huttenlocher et al., 1991; 2004; Merchant et al., 2004; Haun et al., 2005). These results suggest inhomogeneities in the representation of remembered locations across the visual field. However, little is known about how mnemonic representations vary across the visual field. The preponderance of previous studies have parameterized visual space as either binary (e.g. left/right) or unidimensional (e.g. degrees of angle across the same eccentricity) (Funahashi and Kubota, 1994; Goldman-Rakic, 1995). One study provided examples of dLPFC neurons with non-Gaussian spatial WM fields, but did not further elaborate on the receptive fields' structures (Rainer et al., 1998). Although these studies have substantially advanced our understanding of WM, they have also led to models that assume a continuous and/or homogenous representation of the visual-mnemonic space (Camperi and Wang, 1998; Compte et al., 2000; Constantinidis and Wang, 2004; Wimmer et al., 2014). This assumption, however, has not been systematically tested.

Recent behavioral and physiological studies examining WM capacity have demonstrated varying degrees of independence between the left and right visual hemifields (Vogel and Machizawa, 2004; Delvenne, 2005; Buschman et al., 2011;

Delvenne et al., 2011). However, these studies treated visual space as a binary variable, thus restricting their ability to make conclusions about visual mnemonic space beyond that it is represented separately for each hemifield.

Spatial attention is also subject to biases by the meridians of the visual field, which is relevant given the known overlap in neural substrates between attention and WM (LaBar et al., 1999; Awh and Jonides, 2001; Constantinidis et al., 2001b; Miller and Cohen, 2001; Lebedev et al., 2004; Awh et al., 2006; Postle, 2006; Theeuwes et al., 2009; Ikkai and Curtis, 2011; Gazzaley and Nobre, 2012). Psychophysical research has shown that attentional capabilities seem to be somewhat independent for different visual hemifields (Alvarez et al., 2012) and/or quadrants (Carlson et al., 2007; Liu et al., 2009), and that shifting the focus of attention across a meridian incurs a substantial reaction time penalty (Rizzolatti et al., 1987). It is possible that WM and attentional representations share similar constraints, and therefore WM representations of visual space exhibit hemifield or quadrantic biases.

It has also remained ambiguous whether dLPFC contains a topographically organized representation of visual-mnemonic space. There is some evidence that dLPFC is organized in a microcolumnar manner, such that groups of cells within

the same ~.7mm region share recurrent excitatory connections, while inhibitory connections to other microcolumns extend laterally up to 7mm (Kritzer and Goldman-Rakic, 1995; Rao et al., 1999). Such an organization could result in clustering of spatial mnemonic selectivity, such that during WM maintenance neurons within a cluster encoding the same representation share mutual excitation while inhibiting neurons in other clusters encoding different representations. This, however, has yet to be documented.

In order to address these questions, we recorded from ensembles of single neurons in dLPFC area 8a while subjects performed an oculomotor delayed-response task. We found that spatial WM representations are biased in a quadrantic manner: activity underlying WM for stimuli on the opposite side of a meridian from a neuron's memory field is substantially decreased relative to representations of stimuli on the same side of a meridian. This bias is also present in the structure of correlated variability (i.e. spike rate or noise correlations) during WM maintenance, and evident in the subjects' behavior, as saccades to remembered locations exhibit a tendency to repel away from horizontal and vertical meridians and attract towards quadrant centers. We also found that dLPFC neurons encoding similar remembered locations tend to cluster anatomically, and that representation

of the contralateral hemifield on the cortical surface partially reflects the relative distances between points on the retina, though not in a retinotopic manner.

3.3 Materials and methods

3.3.1 Ethics Statement

The animal care and ethics are identical to those in (Leavitt et al., 2013; 2017c) and were in agreement with Canadian rules and regulations and were pre-approved by the McGill University Animal Care Committee. Animals were pair-housed in enclosures according to Canadian Council for Animal Care guidelines. Interactive environmental stimuli were provided for enrichment. During experimental days, water was restricted to a minimum of 35ml/kg/day, which they could earn through successful performance of the task. Water intake was supplemented to reach this quantity if it was not achieved during the task and water restriction was lifted during non-experimental days. The animals were also provided fresh fruits and vegetables daily. Body weight, water intake, as well as mental and physical hygiene were monitored daily. Blood cell count, hematocrit, hemoglobin, and kidney function were tested quarterly. If animals exhibited discomfort or illness, the experiment was stopped and resumed only after successful treatment and recovery. All surgical

procedures were performed under general anesthesia. None of the animals were sacrificed for the purpose of this experiment.

3.3.2 Task

The task was identical to (Leavitt et al., 2017c). Trials were separated into four epochs: fixation, stimulus presentation (stimulus), delay, and response (Figure 3.1A). The animal initiated a trial by maintaining gaze on a central fixation spot (0.08 degrees²) and pressing a lever; the subject needed to maintain fixation within 1.4° of the spot until cued to respond. The fixation period lasted either 482, 636, or 789ms, determined randomly at the beginning of each trial. After fixation, a sine-wave grating (2.5 Hz/deg, 1° diameter, vertical orientation) appeared at one of 16 randomly selected locations for 505ms. The potential stimulus locations were arranged in a 4x4 grid, spaced 4.7° apart, centered around the fixation point. The stimulus period was followed by a randomly variable delay period of 494-1500ms. The delay period ended and the response period commenced when the fixation point was extinguished, cuing the animal to make a saccade to the location of the previously presented stimulus and then to release the lever. The animal had 650ms to respond. Successful completion of the trial yielded a juice reward. The minimum duration between trials was 300ms. Fixation breaks during the trial or failure to

saccade to the target in the allotted time resulted in immediate trial abortion without reward and a delay of 3.5 seconds before the next trial could be initiated.

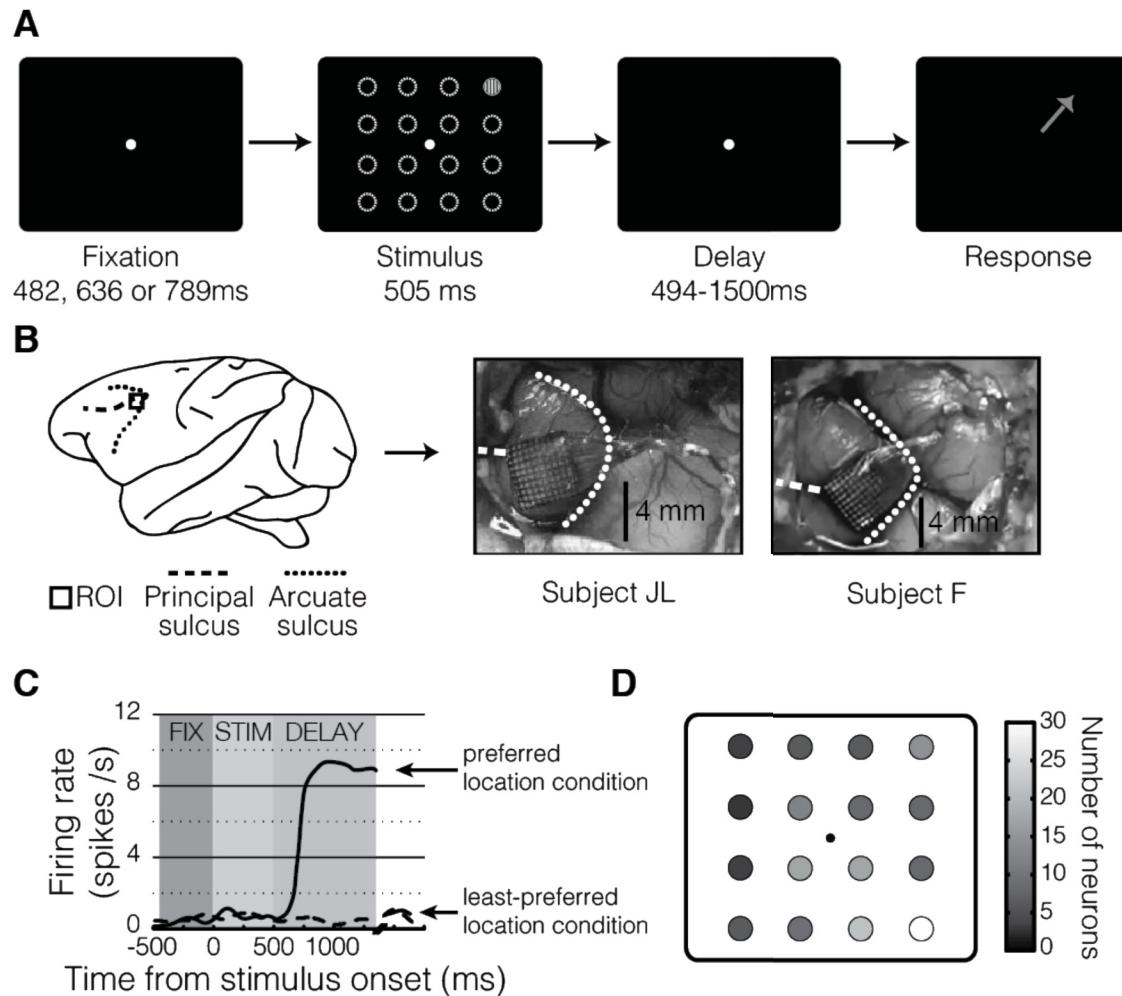


Figure 3.1: Task, method, and single-cell data. (A) Overview of oculomotor delayed-response task, described in detail in the methods. The dashed circles indicating potential cue locations are shown for illustrative purposes and are not present in the task. (B) Array implantation sites and anatomical landmarks in both subjects. (C) Example delay-selective neurons. (D) Distributions of neurons' preferred locations during the delay epoch.

3.3.3 *Experimental Setup*

The experimental setup is identical to (Leavitt et al., 2013; Tremblay et al., 2014; Leavitt et al., 2017c). The stimuli were back-projected onto a screen located 1 meter from the subjects' eyes using a DLP video projector (NEC WT610, 1024x768 pixel resolution, 85 Hz refresh rate). Subjects performed the experiment in an isolated room with no illumination other than the projector, which still provides some illumination even when projecting black. Eye positions were monitored using an infrared optical eye-tracker (EyeLink 1000, SR Research, Ontario, Canada) and endpoint centroids were adjusted to match the target location for each session. A custom computer program controlled stimulus presentation and reward dispensation, and recorded eye position signals and behavioral responses. Subjects performed the experiment while seated in a standard primate chair, and were delivered reward via a tube attached to the chair and an electronic reward dispenser (Crist Instruments, TX, USA) that interfaced with the computer. Prior to the experiments, subjects were implanted with head posts. The head post(s) interfaced with a head holder to fix the monkeys' heads to the chair during experiment sessions.

3.3.4 Microelectrode Array (MEA) Implant

As in (Leavitt et al., 2013; Tremblay et al., 2014; Leavitt et al., 2017c), we chronically implanted a 10x10, 1.5mm MEA (Blackrock Microsystems LLC, Utah, USA; (Maynard et al., 1997; Normann et al., 1999) in each monkey's left dLPFC—anterior to the knee of the arcuate sulcus and caudal to the posterior end of the principal sulcus (area 8a) (Figure 3.1B). Detailed surgical procedures can be found in (Leavitt et al., 2013).

3.3.5 Recordings and Spike Detection

Data were recorded using a ‘Cerebus Neuronal Signal Processor’ (Blackrock Microsystems LLC, Utah, USA) via a Cereport adapter. Spike waveforms were detected online by thresholding. The extracted spikes (48 samples at 30 kHz) were re-sorted manually in ‘OfflineSorter’ (Plexon Inc, TX). The electrodes on each MEA were separated by at least 0.4 mm and were organized into three blocks of 32 electrodes (A, B, C). We collected data from one block during each recording session. Detailed recording procedures can be found in (Leavitt et al., 2013).

3.3.6 Analysis Epochs

We analyzed the final 483ms of the fixation epoch and the entirety of the stimulus epoch; we analyzed the entire delay epoch after the first 150ms in order to minimize

the potential impact of signal latency and stimulus aftereffects (Mendoza-Halliday et al., 2014). We only analyzed successfully completed trials. Data analysis was performed using MATLAB and SPSS.

3.3.7 Single Unit Yield and Epoch Selectivity

We collected spike data from a total of 201 single neurons (99 in JL, 102 in F) from 70 unique recording sites (24 in JL, 46 in F) across 15 recording sessions (7 in JL, 8 in F). To determine whether a neuron was spatially tuned for the stimulus location during the stimulus or delay epochs, we computed a Kruskal-Wallis one-way analysis of variance on the average firing rates with location as the factor. Tuned neurons showed at least one location with a significantly different firing rate ($p < .05$). We found 143 of 201 (71%) neurons exhibited stimulus-selectivity and 157 (78%) exhibited delay-selectivity (Figure 3.1C&D, Supplementary Figure 3.1), yielding 902 correlation pairs between delay-selective units. A neuron's preferred location was defined as the location that elicited the largest response during the epoch of interest.

3.3.8 Spatial Autocorrelation Analysis

To determine whether delay epoch selectivity is anatomically clustered, we first determined the preferred memory location of the parcel of cortex around each

electrode on the microelectrode array, which we defined as the remembered location that generated the greatest response of all thresholded activity on that electrode, across all recording sessions. This yielded a single preferred location for each electrode on the array. Next, we computed Moran's I (Moran, 1950; Zuur et al., 2007; Bullock et al., 2017) across the entire array. Moran's I is a measure of spatial autocorrelation—the degree of clustering or similarity among objects in space—defined as:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2} \quad (4)$$

where N is the number spatial units indexed by i and j ; X is the variable of interest; \bar{X} is the mean X ; and w_{ij} is an element of a matrix of spatial weights. Values of I range from -1 to 1. Positive values of Moran's I indicate that similar feature values are spatially clustered, while negative values of Moran's I indicate that similar feature values are spatially repellent or dispersed. Moran's I was computed iteratively, extending the radius of included locations (the spatial radius) each time, until the whole array was included. This allowed us to determine how preferred location similarity clusters across different spatial scales. For example, computing Moran's I for the smallest cluster radius (400 mm) only included adjacent units,

while computing it for the largest cluster radius included all units on the array. This was performed separately for the horizontal and vertical components of the preferred location, and the results were averaged. Significance was assessed using permutation tests.

3.3.9 Single Unit Firing Rate Meridian Effects

In order to test whether single neurons' firing rates were significantly biased by meridians, we first computed the mean response of each selective neuron to each stimulus location for the epoch of interest. Next, we z-scored each neuron's 16 mean responses to yield standardized firing rates that could be compared across neurons. Finally, we calculated whether a neuron's firing rates were significantly lower for locations that lie across a meridian from that neuron's preferred location, relative to equidistant neurons that fall within the same quadrant as the preferred location. The comparison intervals between group medians in boxplots (e.g. Figure 3.3) are defined as the median $\pm 1.57(q_3 - q_1)/\sqrt{n}$, where q_3 is the 75th percentile, q_1 is the 25th percentile.

In order to control for the difference in the proportion of intraquadrant vs. extraquadrant locations relative to the preferred locations (Figure 3.3A&C; there are two extraquadrant diagonal locations vs. only one intraquadrant diagonal

location), we randomly subsampled half of the diagonal extraquadrant locations such that the number of diagonal intraquadrant and extraquadrant locations were matched. This procedure was repeated 5000 times to obtain a bootstrapped distribution of the median extraquadrant response. This distribution of median values was then compared to the median intraquadrant response.

3.3.10 Stepwise Regression

The distance between stimulus locations covaries with other factors, such as eccentricity and angle. In order to test whether the observed quadrantic biases in single neuron firing rates could be ascribed to these covarying factors, we performed a stepwise linear regression ($P_{entry} = 0.05$, $P_{removal} = 0.1$) to determine which factors significantly affect single neuron firing rates. The regression equation is of the form:

$$y = \beta_0 + \beta_D D + \beta_\theta \theta + \beta_E E + \beta_H H + \beta_V V + \beta_{D\theta} D\theta + \beta_{DE} DE \\ + \beta_{DH} DH + \beta_{DV} DV + \beta_{E\theta} E\theta + \beta_{H\theta} H\theta + \beta_{V\theta} V\theta + \beta_{EH} EH \quad (5) \\ + \beta_{EV} EV + \beta_{HV} HV$$

y is the delay epoch firing rate, β_0 is the constant (intercept term), D is the Euclidean distance between the remembered location and preferred location, θ is the angle between the remembered location and the preferred location, E is the difference in eccentricity between the remembered location and the preferred

location, H is whether the remembered location is across a horizontal meridian from the preferred location, and V is whether the remembered location is across a vertical meridian from the preferred location. The full model is thus composed of a constant, each of the primary factors listed above, and all the first-order interaction terms. In order to control for collinearity, we determined whether the variance inflation factor (VIF) of any coefficients in the final model were greater than 10. If so, we removed the coefficient with largest VIF and repeated the stepwise regression. This procedure was repeated until all coefficients in the final model had a VIF less than 10. The coefficients removed due to collinearity were $\beta_{D\theta}$, $\beta_{\theta H}$, β_{DH} , $\beta_{\theta V}$, β_{DE} , and β_{DV} .

3.3.11 Quadrantic Bias Visualization

In order to visualize the quadrantic bias and obtain a continuous estimate of each neuron's response to the entire region of the visual field covered by the stimulus array, we fit a surface to each neuron's delay-epoch activity for the 16 stimulus locations (Figure 3.4). Specifically, we computed the mean firing rate for each of the 16 locations, then fit a 2-dimensional, 2nd order polynomial of the form

$$f(x, y) = p_{0,0} + p_{1,0}x + p_{0,1}y + p_{2,0}x^2 + p_{0,2}y^2 + p_{1,1}xy \quad (6)$$

to the x- and y-coordinates of each stimulus location. This yielded a function we refer to as the “response surface”. Other than the location of the function’s peak, the firing rate variability represented across the response surfaces was only used for visualization purposes and not for quantitative analysis.

3.3.12 Correlation (r_{sc}) Analysis

In order to compute r_{sc} , we first calculated the z-scores of each unit’s spike counts for each condition (i.e. stimulus location). This removes the spike rate variability across conditions due simply to variability in firing rate responses to different stimuli (i.e. stimulus selectivity) and differences in baseline firing rates for different neurons. We then grouped units into simultaneously recorded pairs ($n=1319$) and computed Pearson’s correlation coefficients ($r_{sc,raw}$) between the z-scored spike counts during each task epoch (Cohen and Kohn, 2011; Leavitt et al., 2013; 2017c). In addition, we minimized the risk of falsely inflating the correlation values by excluding correlations between units on the same electrode from analysis. Fisher’s r -to- z transformation was applied to the correlation coefficients in order to stabilize the variance for hypothesis testing. We also calculated correlations after shuffling the spike rates for all trials (Averbeck and Lee, 2006; Cohen and Kohn, 2011; Tremblay et al., 2014; Leavitt et al., 2017c). The shuffling procedure consisted of

randomizing the trial order within each location condition for each neuron, then computing the spike count correlation ($r_{sc,shuff}$). This procedure destroys the simultaneity in the recordings, thereby providing a measure of the magnitude of correlations expected by chance. The shuffling was repeated 1000 times. The mean of the 1000 shuffles was subtracted from the corresponding $r_{sc,raw}$ to yield a corrected value, henceforth referred to as r_{sc} .

3.3.13 Population Decoding

We used a support vector machine (SVM; Libsvm 3.14; (Chang and Lin, 2011)), a linear classifier, to extract task-related activity from the population-level representations of simultaneously-recorded neural ensembles (Cortes and Vapnik, 1995; Chang and Lin, 2011; Moreno-Bote et al., 2014; Tremblay et al., 2014; Leavitt et al., 2017c). The SVM was given firing rate data from an ensemble in order to predict at which of the 16 locations the stimulus was presented for a given trial, during each of the fixation, stimulus, and delay epochs. The classification was performed separately for each session, using the epoch-averaged firing rates (see **Analysis Epochs**) of each simultaneously recorded neuron. We normalized each unit's firing rates across all trials by subtracting its midrange rate value and dividing by its range (max-min), in order to prevent units with larger absolute

changes in firing rate from dominating the classification boundaries. These two parameters were determined from the training set and applied to both the training and testing sets. We assessed the classifier's performance using cross-validation: a technique in which some proportion of the trials are used to train the decoder, and the decoder attempts to classify the remaining trials. We trained the decoder on 80% of the trials and tested on the remaining 20%. This procedure was repeated such that every trial would be represented once in the testing set. In order to determine whether ensemble representations are biased by meridian effects, we then computed the probability of the decoder mistakenly decoding the remembered location as falling within the same quadrant as the true location and compared it to the probability that the decoder mistakenly decodes the remembered location as being at an equidistant but extraquadrant location from the true location.

3.3.14 Saccade Endpoint Distribution Variability

The variability of saccade endpoint distributions for each target location was computed using an elliptic bivariate normal distribution as in (Merchant et al., 2004) (Figure 3.7). The ellipse was centered at the x - y mean of the saccade endpoints for a given target location. We obtained the two axes of the ellipse via eigendecomposition of the covariance matrix of the x - y eye positions (i.e. a matrix in

which rows = trials, column 1 = x -position, and column 2 = y -position). The resultant orthogonal eigenvectors are the major and minor axes of the ellipse, and are scaled by the square root of their eigenvalues (the variance) and thus lie along the axes of greatest variability of the data. We scaled the axes by the upper 95th percentile of the χ^2 distribution in order to create an ellipse that contains the central 95% of the saccade endpoint distribution. The orientation of the ellipse is the arctangent of the x and y components of the eigenvector from the major axis (i.e. the larger eigenvector).

3.4 Results

Two adult *Macaca fascicularis* performed an oculomotor delayed-response task (Figure 3.1A) while we recorded from neural ensembles in dLPFC area 8a using chronically-implanted 96-channel microelectrode arrays (Figure 3.1B). The neural correlates of WM for spatial locations have been extensively documented in this brain region (Funahashi, 2006; Riley and Constantinidis, 2015). The target stimulus could appear at any one of 16 possible locations, arranged in a uniformly spaced 4×4 grid around a central fixation point. We collected spike data from a total of 201 single neurons across 15 recording sessions, out of which 157 (78%) exhibited delay-epoch spatial selectivity ($P < .05$, Kruskal-Wallis; firing rate×location; Figure

3.1C, Supplementary Figure 3.1). A neuron's preferred location was defined as the location that elicited the largest response averaged over the delay epoch (Figure 3.1D). Both subjects made incorrect choices about the stimulus location in <1% of completed trials.

3.4.1 Anatomical Topography of Mnemonic Representations in dLPFC

One outstanding question in studies of spatial WM is whether the dLPFC contains a topographically organized representation of mnemonic space and whether such a representation follows a retinotopic scheme. In order to answer this question, we first determined the memory location that elicited the largest delay-epoch activity in the cortex surrounding each electrode (a “cortical parcel”—see Methods), and defined this as the preferred location for that electrode (Figure 3.2A). We then computed Moran's I , a measure of spatial autocorrelation (see Methods), across the range of all distances between electrodes on the array, allowing us to determine how similarity in preferred location clusters across different spatial scales for each subject (Figure 3.2B). Positive values of Moran's I indicate that similar feature values are spatially clustered; a given location in space is more likely to be in a local neighborhood with other similar values. Negative values of Moran's I indicate that

similar feature values are spatially repellent or dispersed; a given location is more likely to be in a neighborhood with dissimilar values.

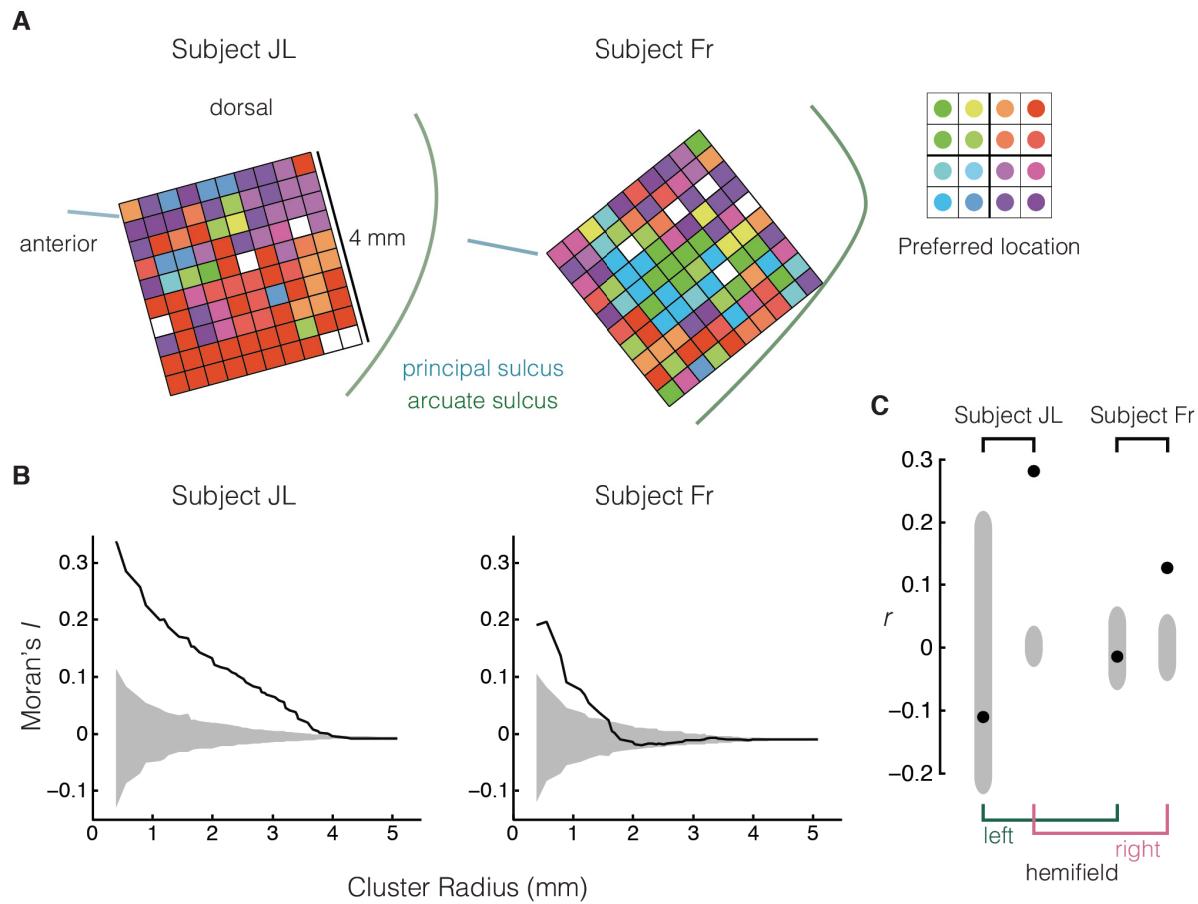


Figure 3.2: Delay Selectivity is Anatomically Clustered in dLPFC. (A) Preferred memory location of the cortical parcel around each electrode of the microelectrode arrays for subjects JL (left) and FR (right). Preferred memory location was defined as the lococation eliciting the maximum response of all thresholded activity on an electrode. (B) Moran's I (y-axis) across spatial scales (x-axis) for each subject. Moran's I is a measure of spatial autocorrelation (i.e. clustering) that ranges between -1 and 1. Positive values of Moran's I indicate that similar preferred locations are spatially clustered. Negative values indicate that similar preferred locations are spatially repellent. Moran's I is computed across the range of all distances between electrodes on the array. The 'cluster radius' is maximum distance

between units included in the computation. For example, computing Moran's I for the smallest cluster radius (400 mm) only includes adjacent electrodes. Shaded region indicates central 95% of null distribution generated by permutation test, thus any point outside the shaded region is considered significant. Preferred locations are significantly spatially autocorrelated at distances $\leq 1.5\text{mm}$ in both subjects. (C) Correlation between anatomical distance between cortical parcels, and Euclidean distance between parcels' preferred locations (i.e. Mantel test), computed separately for each visual hemifield and subject. The black point represents the observed value, while the shaded region indicates the central 95% of the null distribution generated by a permutation test. Thus, the correlation is significant for both subjects in the right (i.e. contralateral) hemifield but not the left hemifield.

Our analysis showed that preferred locations are significantly spatially autocorrelated at distances $\leq 1.5\text{mm}$ in both subjects ($P < .001$ in both subjects, permutation test; Figure 3.2B); a given cortical parcel is more likely to be surrounded by other parcels that have similar delay epoch selectivity than by parcels that have dissimilar delay epoch activity. We also found a correlation between anatomical distance between parcels and the Euclidean distance between the preferred memory locations of those parcels, but only for parcels that have preferred memory locations in the contralateral (right) hemifield to the recording sites ($r = .28, P < .001$, subject JL; $r = .13, P < .001$, subject FR; Mantel Test; Figure 3.2C). The correlation was not significant for parcels with selectivity in the ipsilateral/left hemifield ($r = -.11, P = .17$, subject JL; $r = -.02, P = .33$, subject FR; Mantel Test). These results indicate that the relative spatial relationships in the

retina are partially preserved in the cortical surface. However, retinal coordinates do not strictly map onto cortex (e.g. we found no single foveal region), nor is the effect uniform. Indeed, simple inspection of the data in Figure 3.2A shows that clusters of electrodes with different spatial selectivities (e.g. neurons selective for different hemifields, or for opposite locations in the same hemifield) can sometimes be close, or even adjacent to one another. Thus memory fields in dLPFC exhibit topography, but in a different form than the retinotopic organization of visual areas such as V1 and the FEF.

3.4.2 Spatial Bias in Single-Neuron Firing Rates

To determine whether dLPFC neurons represent visual space homogenously, we first examined how delay activity changed when remembering locations *within* versus *between* quadrants of the visual field. For each neuron with a preferred location adjacent to both a horizontal and vertical meridian (Figure 3.3A & B, grey circles, see Methods), we examined delay activity in trials in which stimuli were remembered at locations equidistant from the preferred location. These locations could fall within the same quadrant as a neuron's preferred location (intraquadrant; Figure 3.3A, green circles), or across a meridian (extraquadrant; Figure 3.3A, red circles). We found that delay epoch activity was significantly lower for remembered

extraquadrant stimuli as compared to equidistant intraquadrant remembered stimuli for neurons that preferred one of the central four stimulus locations ($P < 10^{-10}$, bootstrap test—see methods, Figure 3.3A). We then examined each pairing of intra- and extraquadrant locations. The quadrantic bias was significant for both the vertical and horizontal meridians, and was not due to differences in eccentricity between remembered locations, or dominated by a single meridian (Figure 3.3B). Quadrantic biases for neurons with more eccentric preferred locations showed a similar trend, though not significantly for the horizontal meridian (Supplementary Figure 3.2). In order to control for the possibility that the same neurons were recorded from repeatedly across multiple sessions, the same analyses were repeated including only one session per block of recording electrodes (See Methods), yielding a similar trend (Supplementary Figure 3.3).

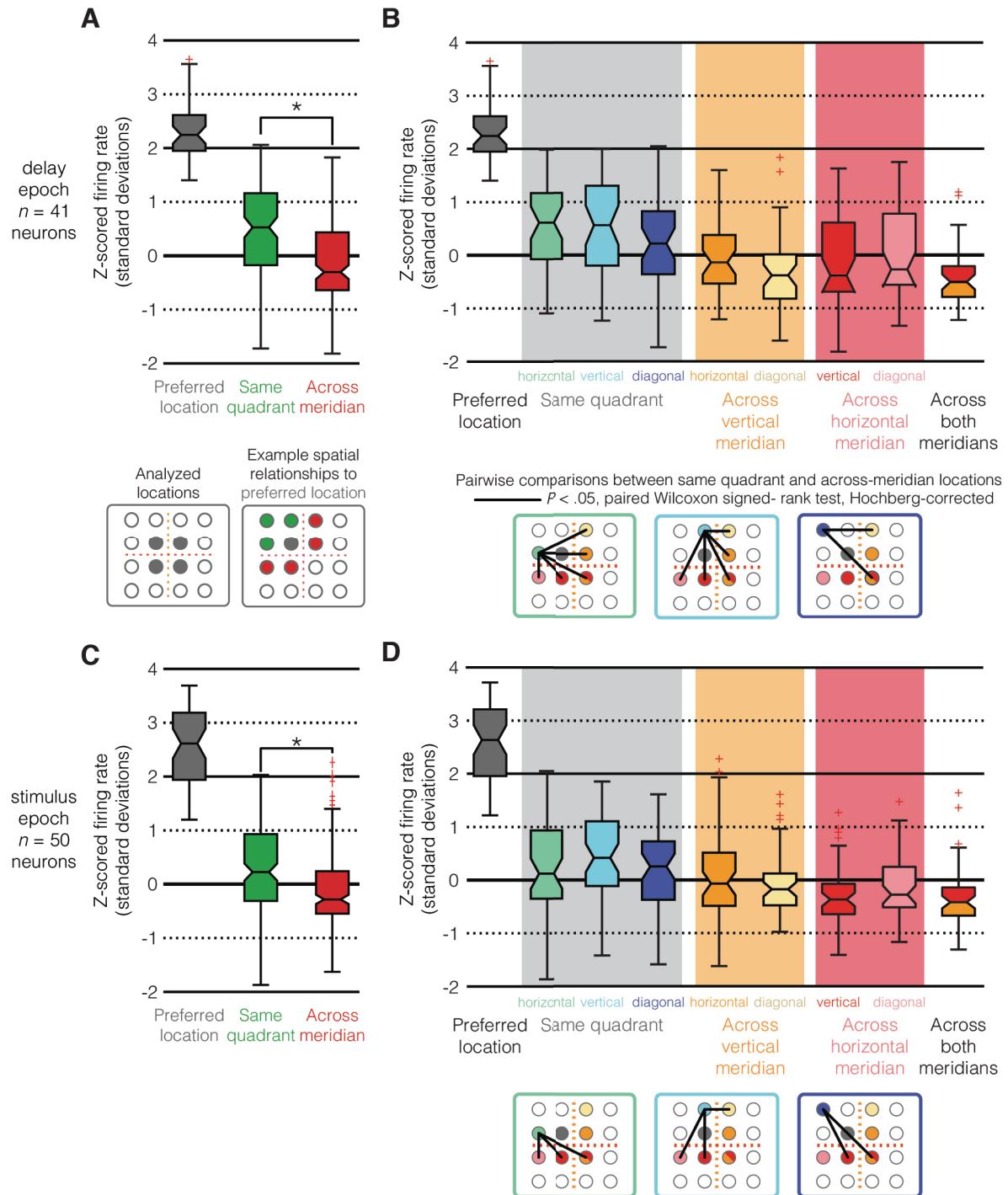


Figure 3.3: Quadrantic Bias in Single Neuron Firing Rates. (A) Quadrantic bias in single neuron firing rates (y-axis) pooled across preferred (grey), intraquadrant (green), and extraquadrant (red) locations during the delay epoch. Firing rates are

z-scored across all 16 locations. $n = 41$ neurons. $*P < 10^{-10}$, bootstrap test—see methods. Notches indicate 95% comparison intervals of the median (see Methods). Edges of boxes extend one quartile from median. Whiskers extend to ~99.3% distribution coverage. Red crosses indicate outlying values. Only the delay epoch and neurons with preferred locations in the central four locations are analyzed in this figure. (B) Similar to (A), but each location lying adjacent to a neuron's preferred location is presented individually. The spatial relationships to the preferred location and significance of pairwise comparisons are depicted in the legends below the figure. Note that the spatial relationships depicted in the legends are relative; the legends use one of the four analyzed preferred locations as an example and neurons with preferred locations at each of the four central locations are analyzed. (C) Identical to (A), but during the stimulus epoch. $n = 50$ neurons. (D) Identical to (B), but during the stimulus epoch. Note that there are fewer significant differences between responses to intrad quadrant vs extraquadrant locations.

In order to determine whether the bias is exclusively present during the delay period or also exists during visual stimulus input, we applied the same analysis to the firing rates during the stimulus epoch. We found that the quadrantic bias during the stimulus epoch was present, though weaker than during the delay epoch; there were fewer significant differences in firing rate between intra- vs extraquadrant stimulus locations during stimulus presentation (Figure 3.3C & D; Supplementary Figure 3.4). This difference was most pronounced for the central four preferred locations: 12 of the potential 15 pairs of locations were significantly different during the delay epoch, as compared to 8 out of 15 during the stimulus

epoch ($P < 0.05$, paired Wilcoxon signed-rank test, Hochberg-corrected; Figure 3.3**B** & **D**).

Examining the stimulus array, one can see that a number of additional factors, such as eccentricity and angle, may covary with the Euclidean distance between stimuli. Thus, it is possible that these factors are responsible for the observed quadrantic biases in single neuron firing rates during the delay epoch. We assessed this possibility by performing a stepwise linear regression to determine which factors significantly affect single neuron firing rates. Our model attempted to predict the delay epoch firing rates using the following factors: Euclidean distance of remembered location from preferred location; angular distance from preferred location; eccentricity difference from preferred location; crossing of the vertical meridian; crossing of the horizontal meridian; and all the first order interaction terms; collinear terms were also removed (see Methods). The horizontal meridian crossing and vertical meridian crossing terms (as well as the rest of the primary factors and multiple interaction terms – see Supplementary Table 3.1 for the full results of the analysis), were significant in the final model ($P < 0.001$ for all primary factors), indicating that crossing a meridian significantly influences firing rates,

and that the observed quadrantic biases cannot be ascribed to alternative covarying factors.

To better convey the quadrantic bias in firing rates, we visualized how the mnemonic activity of single neurons varies across all 16 remembered locations. We did this by fitting a 2-dimensional, 2nd order polynomial to the firing rates for each of the 16 locations (see Methods; Figure 3.4A). The resulting surface approximates the location of maximum activity (the “response peak”) for neurons that respond with similar intensity to multiple adjacent locations, and also provides a continuous estimation of the neuron’s response to the portion of the visual field covered by the stimulus array. The neuron in Figure 3.4A has a preferred location in the lower right quadrant. The epicenter of the preferred field is within the quadrant, far from the horizontal meridian. We superimposed the response surfaces of multiple example neurons in Figure 3.4B. The quadrantic bias in firing rates is clearly visible in the restriction of neural activity to within-quadrant areas and relative lack of activity that extends across the meridians.

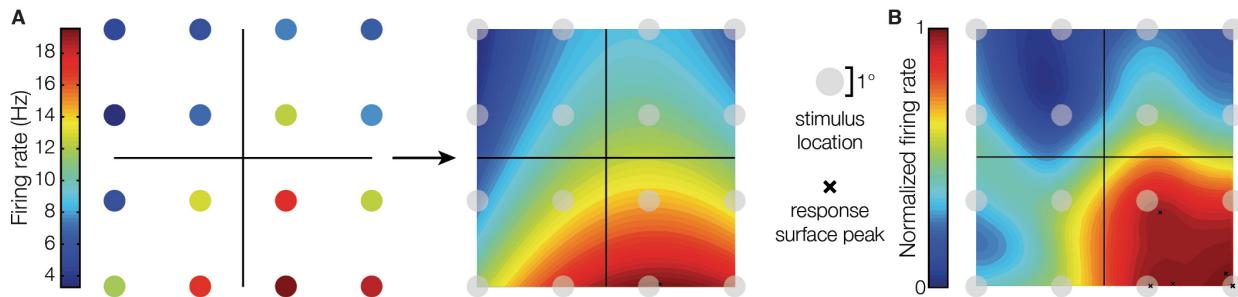


Figure 3.4: Visualizing the Quadrantic Bias. (A) “Response surfaces” were computed by fitting a 2-dimensional, 2nd order polynomial to the mean firing rate for each of the 16 locations (see Methods). The resulting surface provides a continuous estimation of the neuron’s response to the portion of the visual field covered by the stimulus array. (B) By superimposing the response surfaces of five single neurons, the quadrantic bias in firing rates becomes clearly visible. Notice the restriction of neural activity to within-quadrant areas and relative lack of activity that extends beyond quadrants. X’s indicate the peaks of each of the five neurons included in this panel.

3.4.3 Correlated Variability During WM Maintenance

Although the previous results reveal that WM representations of visual space are non-linearly biased by meridians, they do not inform us about the mechanisms underlying this bias. It is thought that the sustained activity encoding visuo-spatial WM is maintained by a neural circuit structure characterized by recurrent excitatory connections between similarly-tuned neurons and lateral inhibitory connections between dissimilarly-tuned neurons (Zipser et al., 1993; Batuev, 1994; Goldman-Rakic, 1995; Camperi and Wang, 1998; Compte et al., 2000; Durstewitz et al., 2000; Constantinidis and Wang, 2004; Compte, 2006). One prediction of this

connection scheme is that correlated variability in firing rate should be greater between similarly-tuned neurons than dissimilarly-tuned neurons. Accordingly, the correlated variability between pairs of neurons encoding visuo-spatial WM representations in the same quadrant (intraquadrant pairs) should be greater than between neurons encoding representations in different quadrants (extraquadrant pairs).

We found that the relationship between tuning similarity and spike count correlation (r_{sc} —a measure of correlated variability—see Methods) varies depending on the task epoch. During fixation, the magnitude of r_{sc} roughly followed the neuron pairs' tuning similarity (Figure 3.5A). Median r_{sc} was significantly greater than zero for intraquadrant pairs (Figure 3.5A, red; $P < 0.05$, sign test, Hochberg-corrected), and for pairs with response peaks in the same left-right hemifield but different top-bottom hemifield (Figure 3.5A, purple; $P < 0.05$, sign test, Hochberg-corrected). Median r_{sc} between intraquadrant neurons was also significantly higher than between neurons with response peaks across both meridians (i.e. the diagonally opposite quadrant; Figure 3.5A; $P < 0.05$, Wilcoxon rank-sum test, Hochberg-corrected). During the stimulus epoch, median r_{sc} was not significantly different from zero for most tuning similarity groups (Figure 3.5C; $P > 0.05$, sign test,

Hochberg-corrected), and no groups were significantly different from each other (Figure 3.5C; $P > 0.05$, Wilcoxon rank-sum test; Hochberg-corrected). We found that median r_{sc} during the delay epoch between intraquadrant neuron pairs was significantly greater than median r_{sc} between neurons with response peaks on the same side of the vertical meridian but opposite sides of the horizontal meridian (Figure 3.5E; $P = 0.012$, Wilcoxon rank-sum test, Hochberg-corrected). This difference was absent for neurons with response peaks on opposite sides of the vertical meridian.

It is noteworthy that the predicted relationship between r_{sc} and the distance between response peaks was only visible within a left/right hemifield but not between left/right hemifields, which may reflect the independence of WM resources for the left and right hemifields (Buschman et al., 2011). Furthermore, these effects are not ascribable to differential responses to stimulus inputs, nor to differences in baseline firing rate of constituent neurons in a correlation pair, because r_{sc} are computed in a manner that control for these factors (see Methods). Thus, we consider these effects to result from the underlying network architecture and not from firing rate or stimulus-driven effects.

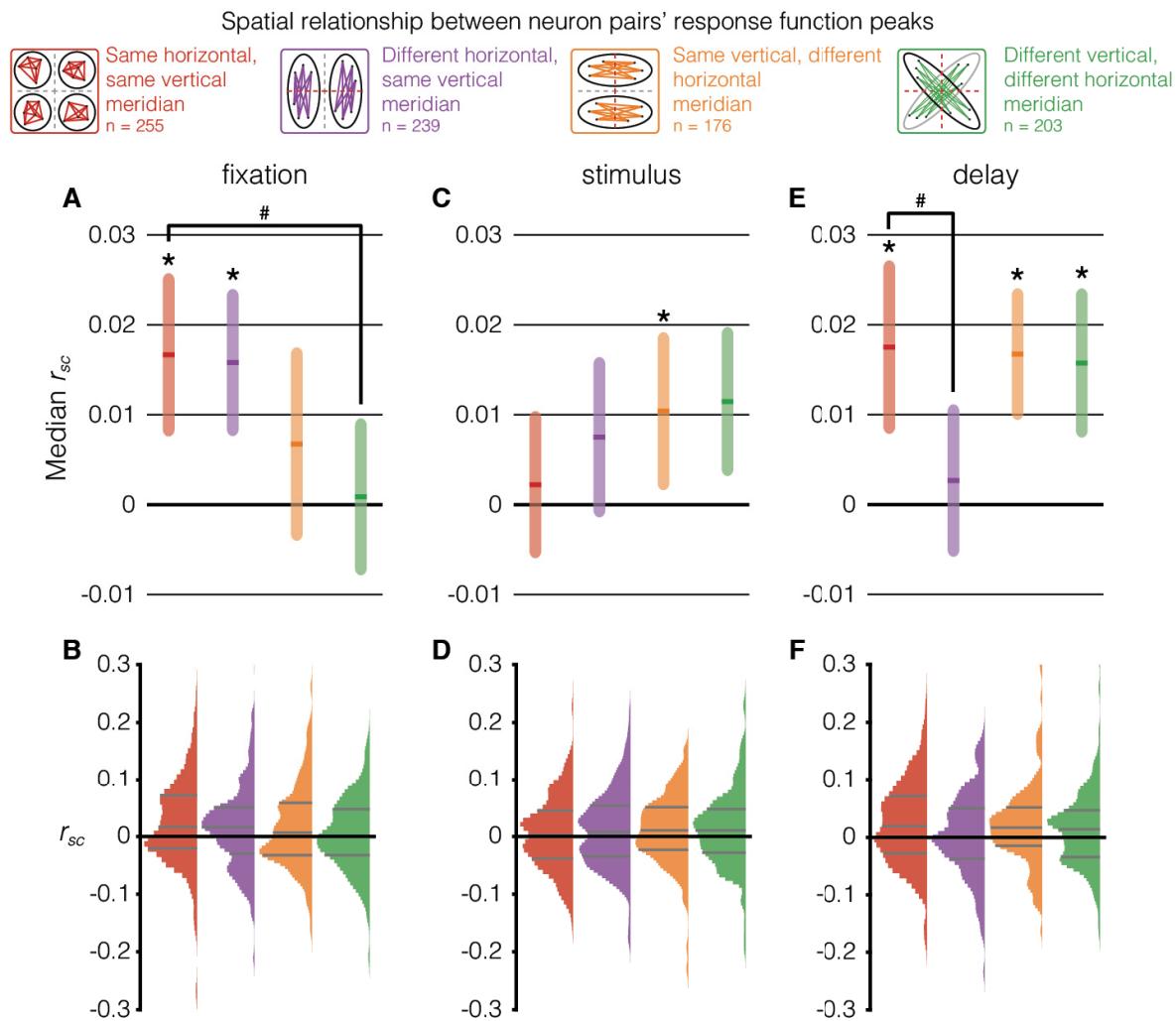


Figure 3.5: Correlated Variability across Task Epochs. (A) Median r_{sc} (y-axis) between delay-selective neurons during the fixation epoch, grouped based on tuning similarity (x-axis). Tuning similarity is determined based on the relative spatial relationship between the quadrants that contain the constituent neurons' response peaks. The legend at the top of the figure depicts each spatial relationship category, showing response peak locations for example correlation pairs in that category. The shaded region is the 95% confidence intervals of the median. *median different from 0, $P < 0.05$, Sign test, Hochberg-corrected. # $P < 0.05$, Wilcoxon rank-sum test, Hochberg-corrected. (B) r_{sc} distributions for each tuning similarity group in (A). Grey lines denote 25th, 50th (i.e median), and 75th percentiles. Values of $r_{sc} > 0.3$ or < -0.3, which constitute less than 5% of the distributions, are omitted from the plot. (C) Same as (A), but for the stimulus epoch. (D) Same as (B), but for the stimulus

epoch. (E) Same as (A), but for the delay epoch. (F) Same as (B), but for the delay epoch.

3.4.4 Quadrantic Bias in Single-Trial Ensemble Representations

Given that single neuron firing rates for different remembered locations within quadrants were more similar to each other than locations between quadrants, it follows that dLPFC ensemble representations of within-quadrant locations should be more similar than across-quadrant representations. To test this hypothesis, we decoded the remembered stimulus location from ensembles of simultaneously recorded neurons on a single-trial basis using a machine-learning algorithm (support vector machine, or SVM, see Methods (Cortes and Vapnik, 1995)) as in (Leavitt et al., 2017c). This method is well-suited to decoding the high-dimensional representations of large groups of neurons (Rigotti et al., 2013; Moreno-Bote et al., 2014).

If population representations of visual mnemonic space have lower resolution within a quadrant than across quadrants, the decoder should commit intraquadrant classification errors with greater probability than extraquadrant classification errors. Indeed, this is exactly what our data show ($P < 0.001$, χ^2 test, Hochberg-corrected; Figure 3.6A). The probability of committing an erroneous intraquadrant

classification is approximately twice that of committing an erroneous extraquadrant classification. We found a similar quadrantic bias in the ensemble representation during the stimulus epoch (Figure 3.6B), however the effect was significantly stronger during the delay epoch. The odds ratio of intraquadrant : extraquadrant decoding errors was significantly greater during the delay epoch compared to the stimulus epoch (Figure 3.6C; $P < 0.01$, z-test). As with the single neuron firing rate data, we also analyzed each meridian and stimulus eccentricity configuration separately, and found that the effect was present in all combinations during memory and some combinations during stimulus presentation (Supplementary Figure 3.5). These results indicate that ensemble-level representations of a given location are more similar to the representations of other intraquadrant locations than equidistant extraquadrant locations, and that this bias is stronger during memory maintenance than during sensory input.

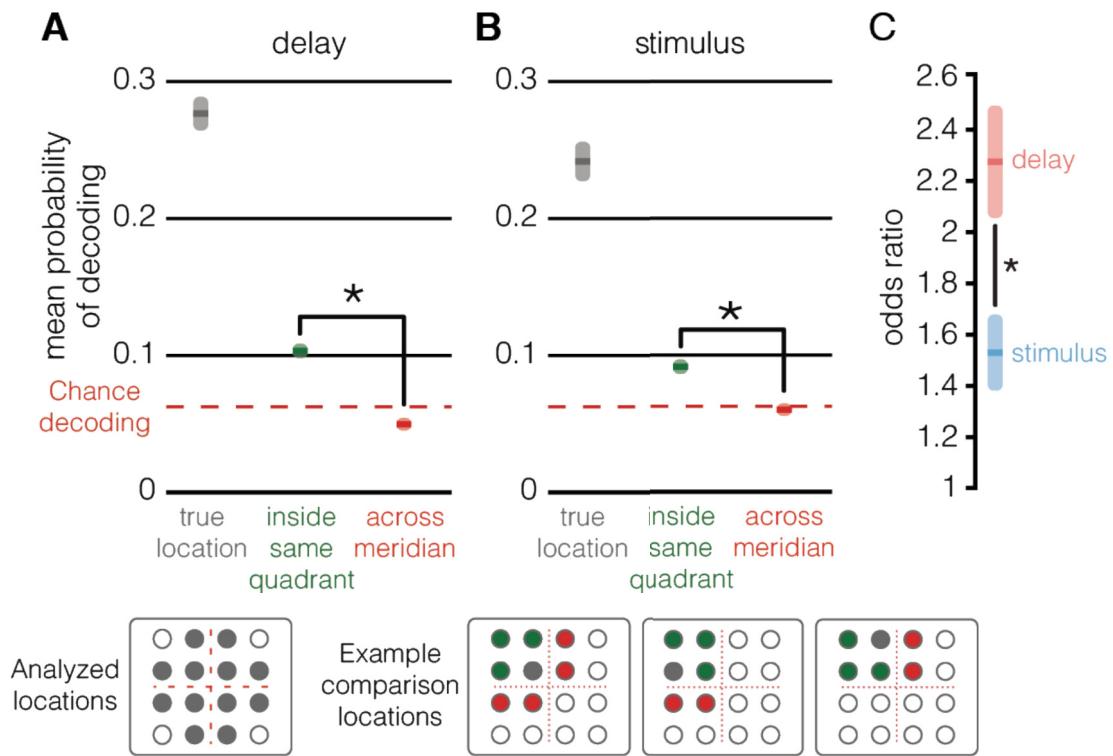
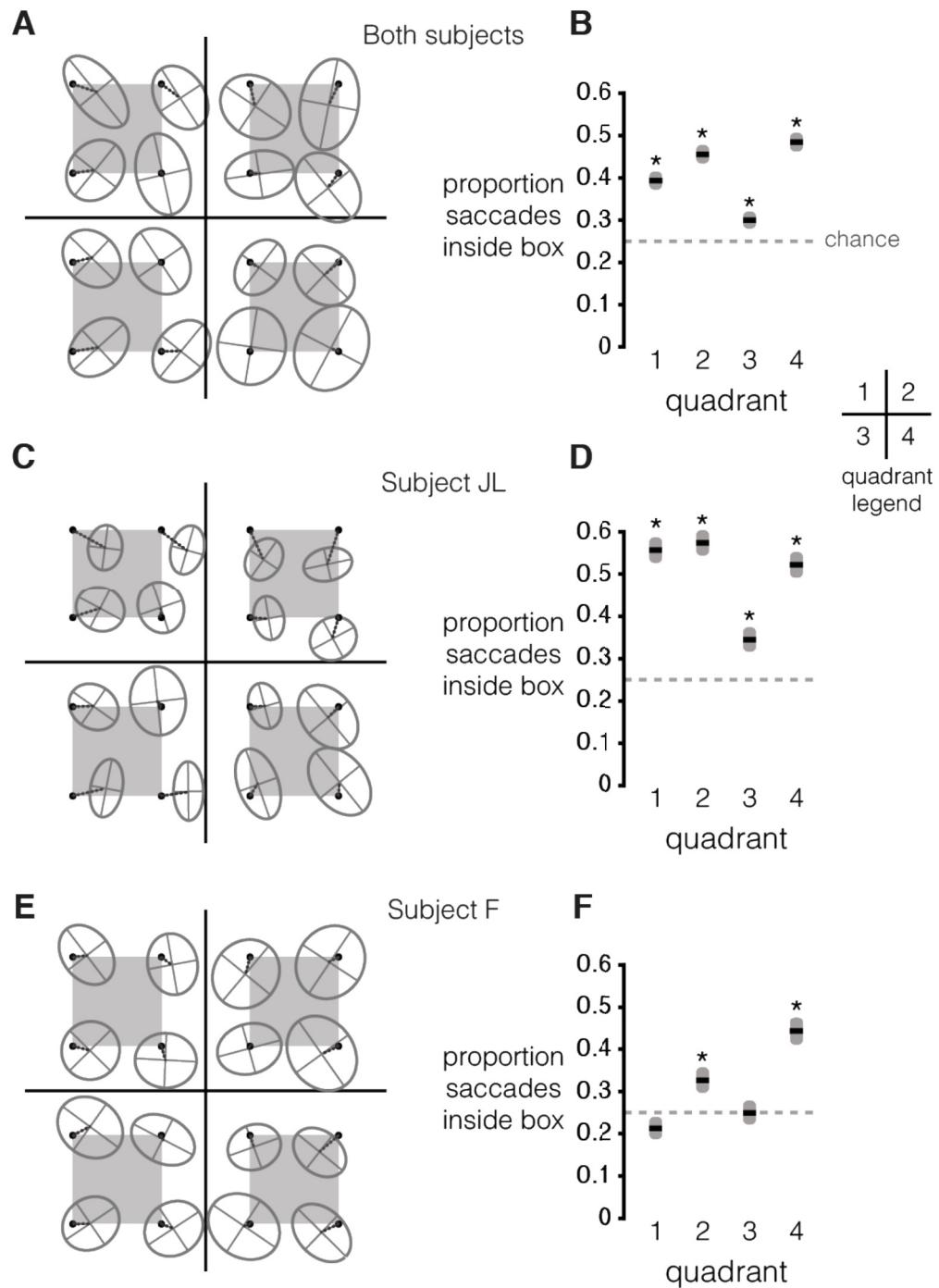


Figure 3.6: Quadrantic Bias in Ensemble Representation. (A) Mnemonic representations were decoded from ensembles of simultaneously-recorded neurons during the delay epoch. The probability of correctly decoding the remembered location during the delay epoch (grey), erroneously decoding it as an intraquadrant location (green), and erroneously decoding it as an extraquadrant location (red), pooled across all locations that lie adjacent to a meridian. Note that this analysis controls for the different proportion of intraquadrant vs. the extraquadrant locations. Shaded regions indicate 95% confidence intervals of the proportion. * $P < 0.001$, χ^2 test, Hochberg-corrected. (B) Same as (A), but for the stimulus epoch. (C) The odds ratio of intraquadrant : extraquadrant decoding errors is plotted for the delay (red) and stimulus (blue) epochs. Shaded regions indicate 99% confidence intervals. * $P < 0.01$, z-test, Hochberg-corrected.

3.4.5 *Quadrantic Biases in Behavior*

Given that we observed significant effects of intraquadrant vs. extraquadrant visuo-spatial mnemonic representations on firing rates, ensemble coding, and spike-rate correlations, we wanted to know whether these effects also manifest in the animals' behavior. We hypothesized that because intraquadrant representations have lower resolution than extraquadrant representations, this should systematically bias memory-guided saccade endpoints toward quadrants and away from meridians, an effect previously reported in human and monkey psychophysical studies (Huttenlocher et al., 1991; 2004; Merchant et al., 2004; Haun et al., 2005). We tested this hypothesis by using the four target locations as outer boundaries to delineate a square region within a quadrant, and calculated the proportion of saccades that fell within that square region (Figure 3.7A). If saccades are not systematically drawn toward quadrant centers, only 25% of saccade endpoints should fall in this square region, whereas if saccades are biased toward quadrant centers, more than 25% of saccades should fall within this region.



* $P < 0.05$, z-test, Hochberg-corrected

Figure 3.7: Saccades Attract to Quadrant Centers. (A) Distributions of saccade endpoints for both subjects. The black dots denote the target locations, and the vertical and horizontal black lines represent the vertical and horizontal meridians,

respectively. We calculated the elliptic bivariate normal distribution of the saccadic endpoints for each target (see Methods). The ellipse is centered at the x-y mean of the endpoints, and the length of the major and minor axes scaled to 95% of the distribution. A dotted line connects the target location to the center of the saccade endpoint distribution in order to visualize the difference between the two points. If significantly more than 25% of saccade endpoints fall within the grey box, we consider the quadrant center to act as an attractor for saccades (see Results). (B) Proportion of both subjects' saccades falling inside the grey box (y-axis) for each quadrant (x-axis). Shaded regions indicate 95% confidence intervals of the proportion. * $P < 0.01$, z-test. $n = 1442, 1457, 1431$, and 1412 , for quadrants 1, 2, 3, and 4, respectively. (C & D) Same as (A & B), but for subject JL. $n = 756, 763, 757$, and 755 , for quadrants 1, 2, 3, and 4, respectively. (E & F) Same as (A & B), but for subject F. $n = 686, 694, 674$, and 657 , for quadrants 1, 2, 3, and 4, respectively.

We found that saccades were systematically attracted toward quadrant centers in all four quadrants when the data from both subjects were pooled (Figure 3.7B; $P < 0.05$, z-test, Hochberg-corrected). However, the strength of the bias was heterogeneous across individual subjects. One subject robustly exhibited a quadrantic saccade bias in all four quadrants (Figure 3.7D), but the other exhibited the bias in only two of four quadrants (Figure 3.7F). Thus the quadrantic biases that we observed in measurements of neural activity are also reflected in the animals' behavior similarly to previous findings in human and non-human primates, though we observed substantial variability across individuals.

Prior studies have found that biases in memory-based estimates of spatial location are more pronounced for longer memory delays (White et al., 1994;

Merchant et al., 2004). In order to determine whether this effect was present in our data, we split the trials into two groups based on the duration of the memory delay: short trials, in which the memory delay was \leq 1000ms, and long trials, in which the memory delay was $>$ 1000ms. We compared the strength of the quadrantic bias in saccades in the short vs. long memory delay trials, and did not find a significant difference between the two groups in any of the four quadrants (Supplementary Figure 3.6; $P > 0.05$, χ^2 test, Hochberg-corrected).

It is possible that even if we could not detect an effect of memory delay duration on the quadrantic bias at the behavioral level, it could still be present at the neuronal level. However, a comparison the early delay epoch—defined as 151–450ms after the beginning of delay epoch—and the late delay epoch—defined as the final 200ms of the delay epoch—did not reveal any major differences in the strength of the quadrantic bias in the single neuron firing rates (Supplementary Figure 3.7 and Supplementary Figure 3.8). The number of significant differences between responses to intraquadrant and extraquadrant locations and the pattern of significant differences were similar during the early and late delay epoch. It is possible that we failed to find a significant effect of delay epoch duration on the

strength of the quadrantic bias because our maximum delay duration was 1500ms, while prior studies used delay durations of up to 5600ms.

3.5 Discussion

We systematically varied the position of a remembered location across multiple dimensions of visual space in an oculomotor delayed-response task while simultaneously recording from ensembles of single neurons in dLPFC area 8a. We found a quadrant-centric bias of visual mnemonic space representations, evident in single neuron firing rates, pairwise correlated variability, ensemble encoding of remembered location, and a bias in saccade endpoint towards quadrant centers. We also found that mnemonic activity is anatomically organized and clustered across dLPFC in a manner that partially reflects the geometric properties of visual space, but is not retinotopic.

3.5.1 Clustering of Mnemonic Representations in dLPFC

While there are abundant examples of topographic organization in brain regions more directly involved in sensory and motor processing, evidence for topography in dLPFC has historically been limited. This is likely because the basic sensory quantities under investigation in mapping studies do not have a straightforward relationship with the structure and function of dLPFC, a region known to be

involved in comparatively abstract components of sophisticated behavior (Miller and Cohen, 2001). One study found that receptive fields for visual stimuli tend to become larger and more eccentric as one moves dorsally, away from the ventral portion of the arcuate sulcus (Suzuki and Azuma, 1983). Given the heterogeneity of individual samples and subjects in the trend they observed, the distribution of visual and mnemonic preferred location we report does not appear at odds with their findings.

A recent study by Kiani et al. (Kiani et al., 2015) took a novel approach to investigating electrophysiological topography in dLPFC. Also using microelectrode arrays implanted in area 8a, they applied techniques similar to those used in determining resting state networks in fMRI experiments, and grouped neurons into modules based on shared variability in firing rate across entire sessions of experimental recordings. They found that the modules were anatomically distinct, and organized more on the basis of “common noise” than on task-related activity, even across different tasks. Given the difference in analytical techniques between their study and ours, it does not seem that the two sets of findings necessitate reconciliation. Indeed, considering our results together with theirs leads to a potential conclusion that task-related properties of neurons cluster independently

or are embedded within the modules that emerge as a result of intrinsic or task-independent variability.

Spatial representations are retinotopically organized across many primate visual areas (Van Essen et al., 1984; Maunsell and Van Essen, 1987). However, our data do not show such a strict organization within the area covered by the microelectrode arrays, despite the fact that we found non-random representation of the entire stimulus array distributed across the area. This strongly suggests that such a retinotopic organization is absent in the dLPFC. One possible explanation for this finding is that interactions between neurons representing different locations across the visual space through lateral connections are facilitated by the heterogeneity within a relative retinotopic arrangement. Supporting this claim, lateral connections between neurons in the dLPFC are limited to a few millimeters (Kritzer and Goldman-Rakic, 1995), and such connections may be critical to the implementation of delay activity dynamics during WM maintenance by ensembles of neurons (e.g., recurrent excitation and mutual inhibition) (Goldman-Rakic, 1995). Prior work from our laboratory has also found that networks of dLPFC neurons that maximize WM-related information span a larger anatomical area than predicted by the statistics of a randomly-sampled neuronal population (Leavitt et al., 2017c). Our

findings are also concordant with previous work indicating that the structure of spike count correlations seem to reflect a proposed coding scheme for WM networks in which narrow range excitation and wider range inhibition are critical to the maintenance of the representations (Camperi and Wang, 1998; Compte et al., 2000; Leavitt et al., 2013).

3.5.2 Potential Origins of Quadrantic Biases in Visual and Mnemonic Space

Previous studies investigating meridian effects in WM have typically focused on WM capacity independence between the left and right hemifields, and all found some degree of hemifield independence at behavioral and/or physiological levels (Vogel and Machizawa, 2004; Delvenne, 2005; Buschman et al., 2011; Delvenne et al., 2011; Matsushima and Tanaka, 2014). Although these experiments were designed to address WM capacity, their results can be interpreted as demonstrating a vertical meridian effect in visual mnemonic space, albeit within a context constrained by a binary parameterization of the space. Our results demonstrate the existence of a vertical meridian effect in a more sophisticated model of visual mnemonic space, and specify the spatial structure and variability of this phenomenon.

We found that mnemonic representations of a given location are more similar to other locations within the same quadrant than to equidistant locations that lie across a meridian, at both the single neuron and population levels. One interpretation of this finding is that visual and memory fields adjacent to a meridian extend further in the same quadrant in which the field's epicenter is located than across the meridian. The literature on multi-dimensional visual and memory field characteristics of neurons in area 8a is limited (Rainer et al., 1998).

It is possible that the biases in the representations of visuo-mnemonic space reported here result from biases in the structure of receptive fields in areas upstream from dLPFC area 8a. For example, receptive fields in visual striate and extrastriate areas are retinotopically organized, are constrained to the contralateral visual hemifield, and their size is smallest in the fovea and increases proportionally to eccentricity (Virsu and Rovamo, 1979; Van Essen et al., 1984; Kandel et al., 2000). However, in areas downstream from unimodal visual cortices, such as the dLPFC, this organization changes. Neurons mainly respond to visual stimuli that are behaviorally relevant, and receptive fields are located in both visual hemifields (Suzuki and Azuma, 1983; Boch and Goldberg, 1989). Interestingly, bilateral representation of the visual field does not start de-novo in dLPFC; areas upstream

along both the dorsal and ventral pathways, such as medial superior temporal (MST) and Inferior temporal (IT), also show bilateral receptive fields (Gross et al., 1969; Desimone et al., 1984; Komatsu and Wurtz, 1988; Raiguel et al., 1997).

One possible explanation for the vertical meridian bias observed in our data is that the contralateral hemifield representation bias present in visual areas is “passed on” to neurons in area 8a. This explanation could be extended to the observed quadrantic bias, as vertical asymmetries are known to exist along the visual system; a greater area of the LGN, V1, and MT are devoted to representing the inferior half of the visual field (Connolly and Van Essen, 1984; Van Essen et al., 1984; Maunsell and Van Essen, 1987). These anatomical properties have been proposed as the reason why spatial frequency perception is superior along the inferior portion of the vertical meridian relative to the superior portion, and the origin of the BOLD signal asymmetries in human V1 and V2 that mirror the behavioral phenomenon (Carrasco et al., 2001; Liu et al., 2006; Abrams et al., 2012).

It is surprising that the strength of the horizontal and vertical meridian effects are similar for neurons with central preferred locations, given the contralateral representation bias that is ubiquitous across the brain. It is still possible that this contralateral representation bias underlies the lack of a significant horizontal

meridian effect for neurons with peripheral preferred locations. However, the lack of a significant effect could also be due to an insufficiently large sample of neurons; we obtained only 21 neurons with peripheral preferred locations along the horizontal meridian, compared to 42 neurons for the vertical meridian.

One may speculate that the bias in WM representations results directly from a bias in visual representations, perhaps due to the overlap in populations of neurons representing visual and mnemonic information (Supplementary Figure 3.1) (Constantinidis et al., 2001b). This is plausible, but the increased strength of the quadrantic bias during the delay epoch relative to during stimulus presentation indicates that this explanation is incomplete, and that WM maintenance amplifies existing representational biases and/or creates novel ones entirely.

Another series of studies in humans and non-human primates posits that WM-based estimates of spatial location rely on two distinct processes: an unbiased fine-grain representation of visual space, and a categorical representation of a larger region bounded by landmarks or natural divisions in visual space (e.g. meridians) that encompasses the fine-grain values (Huttenlocher et al., 1991; 2004; Merchant et al., 2004; Haun et al., 2005). The fine-grain information is subject to temporal decay, and thus over longer memory delays the representation becomes biased

toward category (the quadrant centers, in the present study). While such a system introduces bias, it can also reduce trial-to-trial variability to a degree that yields a net accuracy benefit. One experimentally-verified prediction of this model is that the categorical bias grows stronger as the memory delay increases (Merchant et al., 2004). We did not observe this phenomenon in the behavior or the neuronal activity in the present experiment, which we ascribe to the length of the memory delay in our task. Our memory delay ranged from 494-1500ms, while the memory delay in the prior study ranged from 500-5000ms. It is likely that the strength of the bias did not change sufficiently to be detectable in our shorter time window. The limited time windows for analysis of single neuron firing rates could also have yielded noisier firing rate estimations that obscured an underlying difference. Nevertheless, given the PFC's involvement in WM and categorical representation of continuous quantities (Freedman and Miller, 2008; Merchant et al., 2011; Goodwin et al., 2012), further studies with longer memory delays may reveal neurophysiological correlates of the behavioral phenomena we were unable to find in the present study.

3.5.3 Alternative Factors Affecting Memory-Guided Saccades

The amplitude of memory-guided saccades is known to be influenced by a number of factors, including illumination (Goffart et al., 2006), training (Visscher et al., 2003),

and orbital position of the eyes (Barton and Sparks, 2001), though it is unlikely these factors significantly biased eye movement data in this experiment. Head orientation, the location of the subject relative to the screen, and the location of the fixation point on the screen were all constant within and across sessions, thus there should have been little variation in orbital position and its consequences on saccades (Barton and Sparks, 2001) should be minimal. Subjects were tested on the same 16 locations throughout training and recording, thus the interactions between trained and novel remembered saccade targets described in (Visscher et al., 2003) are not present in this experiment. While previous studies have reported an upward bias in memory-guided saccades that decreases with the vertical position of the target (Gnadt et al., 1991; White et al., 1994; Goffart et al., 2006), this effect is largely eliminated in the presence of dim illumination (6.5×10^{-3} - 0.05 cd/m²) (Gnadt et al., 1991; Goffart et al., 2006) comparable to that generated by the projector in this task. As such, our experimental design seems to control for factors known to affect the amplitude of memory-guided saccades.

3.5.4 Meridian Effects Elucidate the Relationships Between WM, Attention, and Motor Activity

Substantial literature exists on the overlap and interaction between the behavioral effects and neural substrates of attention and WM (LaBar et al., 1999; Awh and Jonides, 2001; de Fockert, 2001; Miller and Cohen, 2001; Lebedev et al., 2004; Awh et al., 2006; Postle, 2006; Theeuwes et al., 2009; Ikkai and Curtis, 2011; Mendoza et al., 2011; Gazzaley and Nobre, 2012). Indeed, there is good reason to believe that much of the neural activity in dLPFC that is traditionally considered WM maintenance-related can instead be attributed to the attentional component of WM tasks (Owen et al., 1996; Lebedev et al., 2004). As such, our finding of quadrantic divisions of visual mnemonic space could share neural origins with the quadrant-independent capacity for multi-object attention (Carlson et al., 2007) and the mitigating effects of meridians on visual crowding (Liu et al., 2009).

Perceptual biases relative meridians have been reported in previous behavioral studies. For example, it is well known that subjects overestimate the angle/direction relative to a meridian during orientation and motion direction discrimination tasks, a phenomenon known as motion or orientation repulsion (Loffler and Orbach, 2001; Changizi et al., 2008; Dakin et al., 2010). This effect may be related to our findings

that perceptual and mnemonic representations are “repulsed” away from main meridians.

Three decades ago, Rizzolatti and colleagues found that shifts of attention across a meridian cause substantially larger reaction time penalties than equidistant shifts of attention within a quadrant (Rizzolatti et al., 1987). This finding contributed to the basis of the premotor theory of attention (Rizzolatti et al., 1987; Sheliga et al., 1994). Many of the theory’s claims about the relationship between eye movements and attention have been called into question by later studies, yielding a refined version of the theory that is best summarized as “saccade preparation is necessary for exogenous attentional orienting, whereas endogenous attentional orienting is entirely independent of motor control” (Smith and Schenk, 2012). Interestingly, meridian effects seem to only exist for endogenous attention and not exogenous attention (Reuter-Lorenz and Fendrich, 1992; Botta et al., 2010). We consider the presence of mnemonic spatial meridian effects as evidence that dLPFC delay activity is not principally motor-related because, as mentioned before, meridian effects only exist for endogenous attention, which acts independently of motor control (Smith and Schenk, 2012). Furthermore, because attentional meridian effects are only observed in endogenous attention, our observation of

quadrantic biases in spatial WM representations specifies an overlap between WM and *endogenous* attention, not exogenous attention.

3.5.5 r_{sc} and WM Tasks

Previous studies in dLPFC have reported varying degrees of correlated activity during WM tasks (Funahashi and Inoue, 2000; Constantinidis et al., 2001a; Constantinidis and Goldman-Rakic, 2002; Funahashi, 2006; Qi and Constantinidis, 2012; Katsuki and Constantinidis, 2013; Wimmer et al., 2014; Leavitt et al., 2017c).

One prior study found that cross-correlation strength and significance was positively correlated with tuning similarity during memory (Constantinidis et al., 2001a), another found a non-significant trend that task epoch affects r_{sc} (Constantinidis and Goldman-Rakic, 2002), and a third demonstrated r_{sc} differences between animals naïve to and proficient at a spatial WM task (Qi and Constantinidis, 2012). Recently, analysis by Wimmer and colleagues found that r_{sc} varied during memory in a tuning-dependent manner (Wimmer et al., 2014) indicative of a continuous ‘bump attractor’ representation scheme (Compte et al., 2000) of spatial WM. We found a similar relationship between tuning and r_{sc} during the fixation epoch. However, we found that r_{sc} during the delay epoch only changes as a function of tuning for neurons that have preferred memory locations on the

same side of the vertical meridian (i.e. within the same left-right hemifield). It is possible that the observed effects of tuning on r_{sc} in the aforementioned studies were dominated by the within-hemifield effect, which can be explained by the higher proportion of contralateral-selective neurons in this region (Funahashi and Kubota, 1994; Goldman-Rakic, 1995; Funahashi and Takeda, 2002; Lennert and Martinez-Trujillo, 2013).

3.5.6 A Hemifield-Independent WM Mechanism

We found that r_{sc} during the delay epoch between intraquadrant neuron pairs was significantly larger than r_{sc} between neurons with response peaks in the same left-right hemifield and in different top-bottom hemifields. However, we did not find any significant difference in r_{sc} between neuron pairs with response peaks on opposite sides of the vertical meridian. Our interpretation of this finding is that WM maintenance results in inhibition between neurons with response peaks in the same left-right hemifield and in different top-bottom hemifields. The finding that the horizontal meridian exerts the strongest effect on the r_{sc} structure can synthesize two well characterized, but previously unrelated findings regarding neuronal correlates and behavioral phenomena in WM. The first is a model network architecture that stabilizes WM representations across time (Polk et al., 2012). One

hallmark of this architecture is that correlated activity between neurons that maintain similar WM representations (e.g. locations within the same visual quadrant) is stronger than correlated activity between neurons that store different kinds of WMs (e.g. locations in different visual quadrants). The second finding is that WM resources are independent for the left and right hemifields of visual space (Delvenne, 2005; Buschman et al., 2011; Alvarez et al., 2012; Matsushima and Tanaka, 2014). Combining both of these factors yields the prediction that the correlation structure indicative of a WM-stabilizing architecture should be present separately for each left/right hemifield. Indeed, our data match this prediction, advancing a model of WM that integrates a model network architecture with known behavioral and neural biases.

3.6 Conclusion

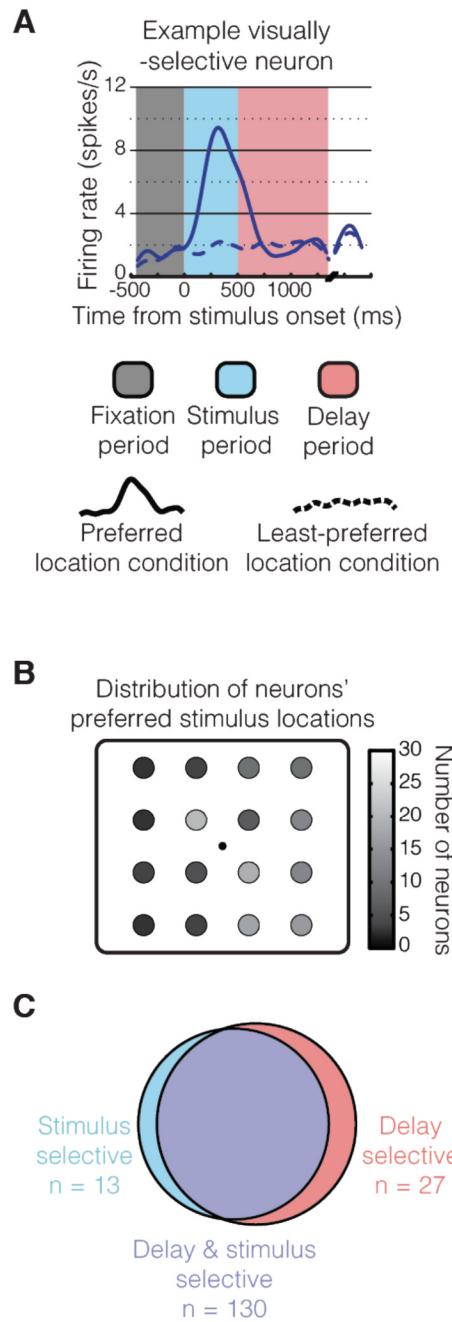
Our results indicate that dLPFC contains a non-retinotopic, topographic organization of spatial WM representations likely shaped to support interactions between neurons that are essential to the origin of delay activity in the absence of visual inputs. This is supported by our observation of a pattern of correlated variability that is thought to be a hallmark of a mechanism for temporally stabilizing WM representations. Our results also provide a neural correlate for

known quadrantic biases in human and monkey visuo-spatial WM. Finally, our results suggest revising current models of WM to accommodate the non-linearities of visual mnemonic space representation in dLPFC.

3.7 Acknowledgements

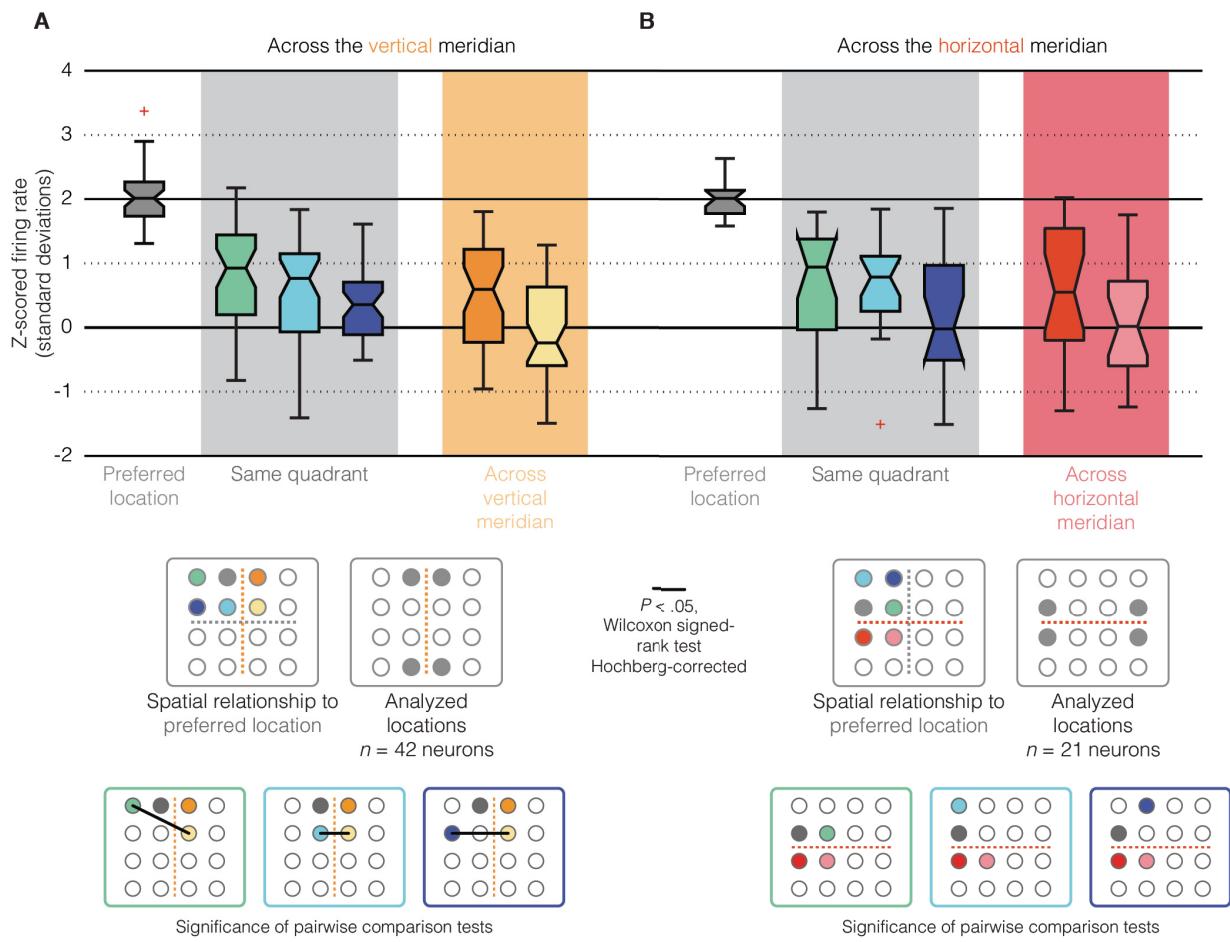
This research was supported by the CIHR and NSERC. The authors would like to thank M. Schneiderman, W. Kucharski, and S. Nuara for technical assistance, the M.T. Lab and T. Quail for their scrutiny and mirth, and R. Gulli for providing a successful heading.

3.8 Supplemental Data

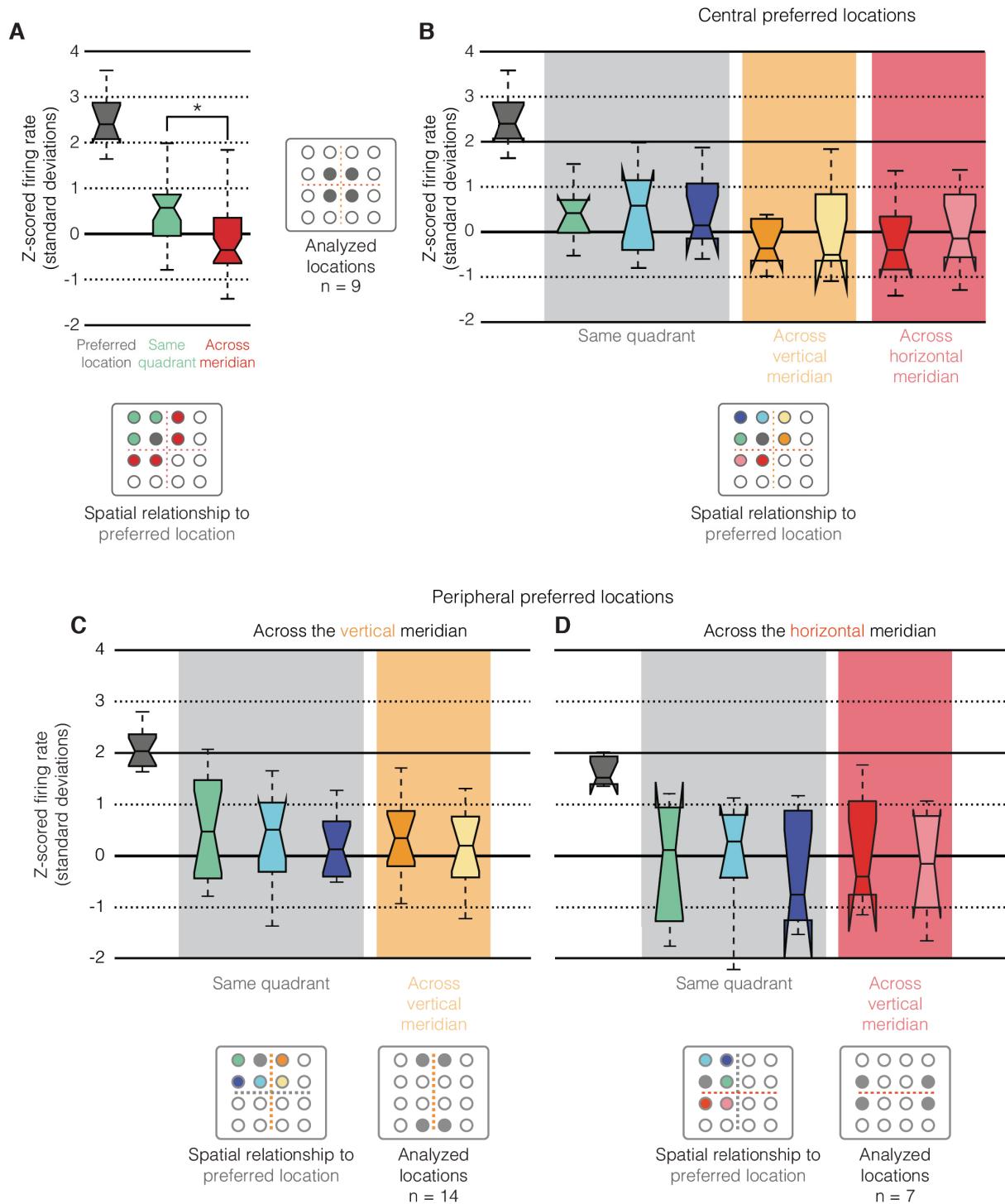


Supplementary Figure 3.1: Stimulus Selectivity. Similar to Figure 3.1, but for the stimulus epoch. (A) Example stimulus-selective neurons. (B) Distributions of

neurons' preferred locations during the stimulus epoch. (C) Quantities of exclusively stimulus-selective, exclusively delay-selective, and delay- & stimulus-selective neurons.

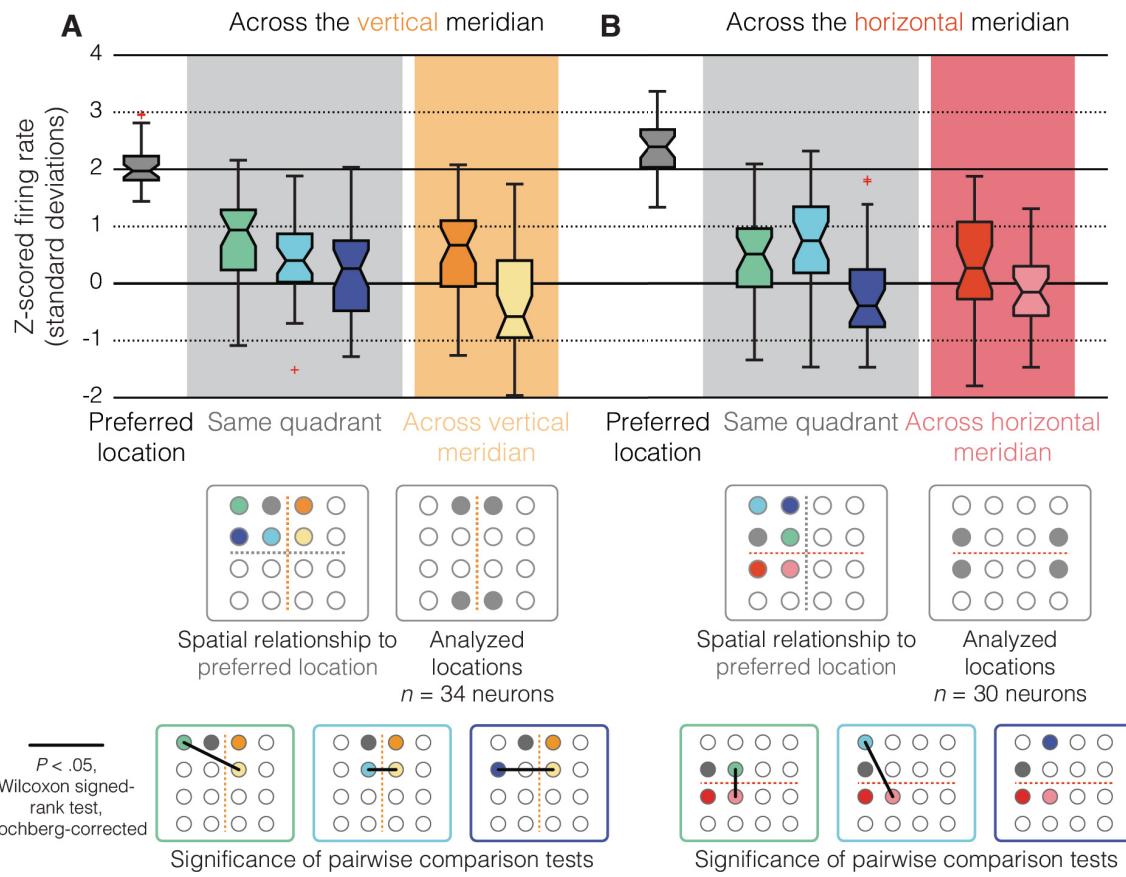


Supplementary Figure 3.2: Quadrantic Bias in Firing Rates for Single Neurons with Peripheral Preferred Locations During the Delay Epoch. Similar to Figure 3.3B, but for neurons with peripheral preferred locations in the delay epoch. Firing rates on the ordinate (y) axis are z-scored across all 16 locations. Notches indicate 95% comparison intervals of the median (see Methods). Edges of boxes extend one quartile from median. Whiskers extend to ~99.3% distribution coverage. Results for the: (A) vertical meridian, and (B) horizontal meridian.

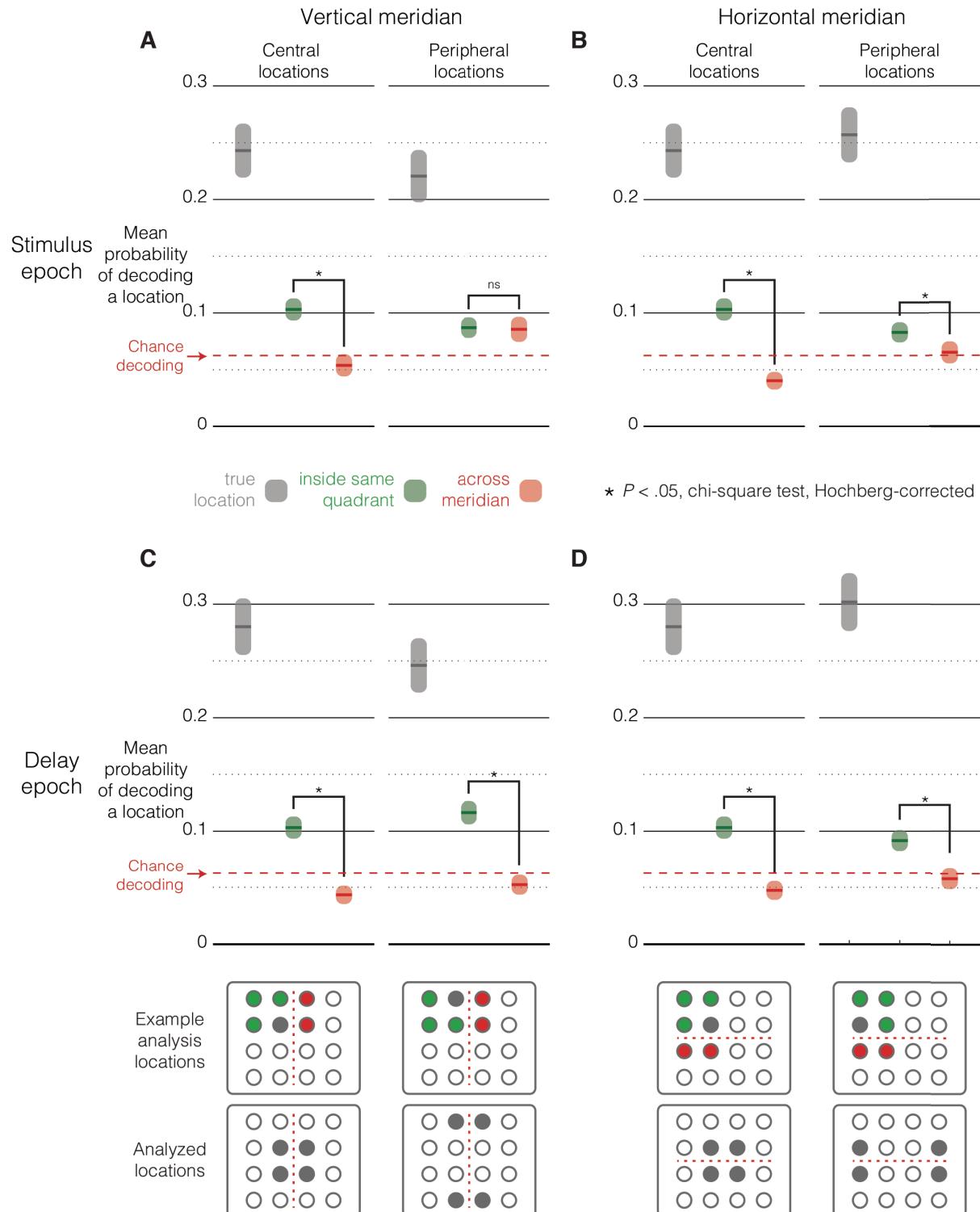


Supplementary Figure 3.3: Quadrantic bias in data without repeated recordings from same electrodes. Repeated sessions on the same array electrodes were

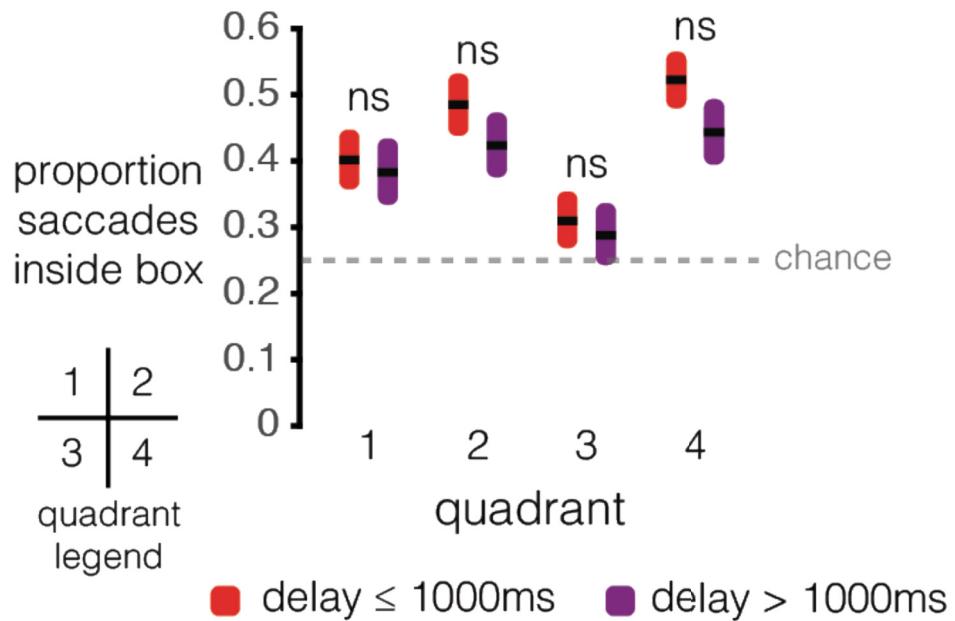
removed, such that there was only one session per electrode. Analyses identical to those in Figure 3.3 and Supplementary Figure 3.2 were performed. (A) Quadrantic bias in single neuron firing rates pooled across preferred (grey), intraquadrant (green), and extraquadrant (red) locations. * $P < 0.05$, Wilcoxon rank-sum test, Hochberg-corrected. (B) Similar as in (A), but the analysis is separate for each location lying adjacent to a neuron's preferred (grey) location. (C) Similar to (B), but for peripheral preferred locations lying adjacent to the vertical meridian, and the (D) horizontal meridian.



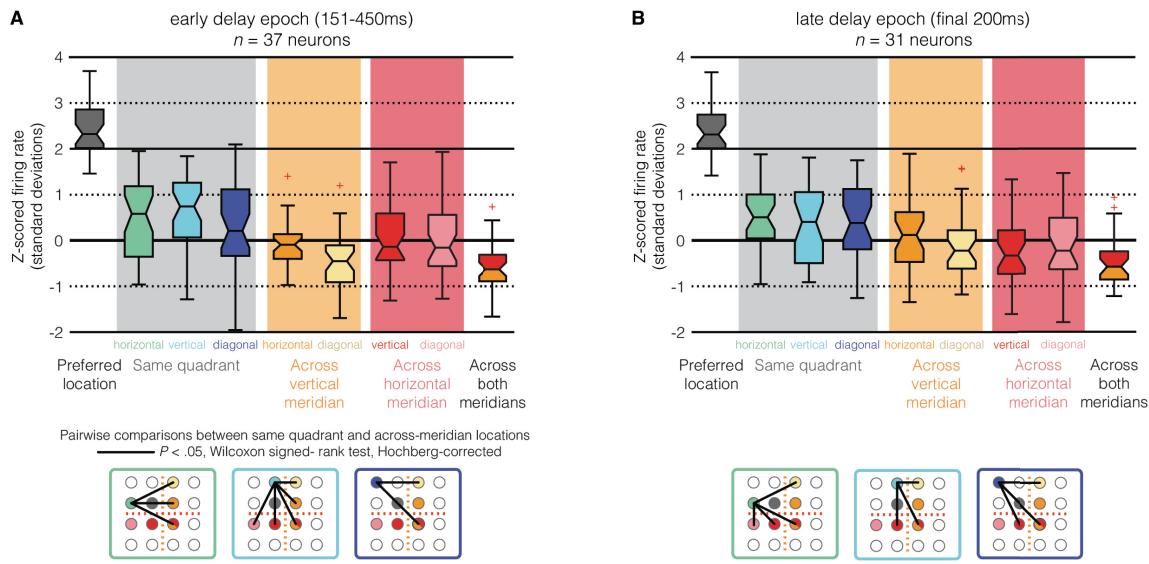
Supplementary Figure 3.4: Quadrantic Bias in Firing Rates for Single Neurons with Preferred Locations in the Periphery During the Stimulus Epoch. As in Figure 3.3 and Supplementary Figure 3.2, firing rates on the ordinate (y) axis are z-scored across all 16 locations. Notches indicate 95% comparison intervals of the median. Edges of boxes extend one quartile from median. Whiskers extend to ~99.3% distribution coverage. (A) Quadrant effect analyses for neurons with peripheral preferred locations along the vertical meridian, and (B) for neurons with peripheral preferred locations along the horizontal meridian.



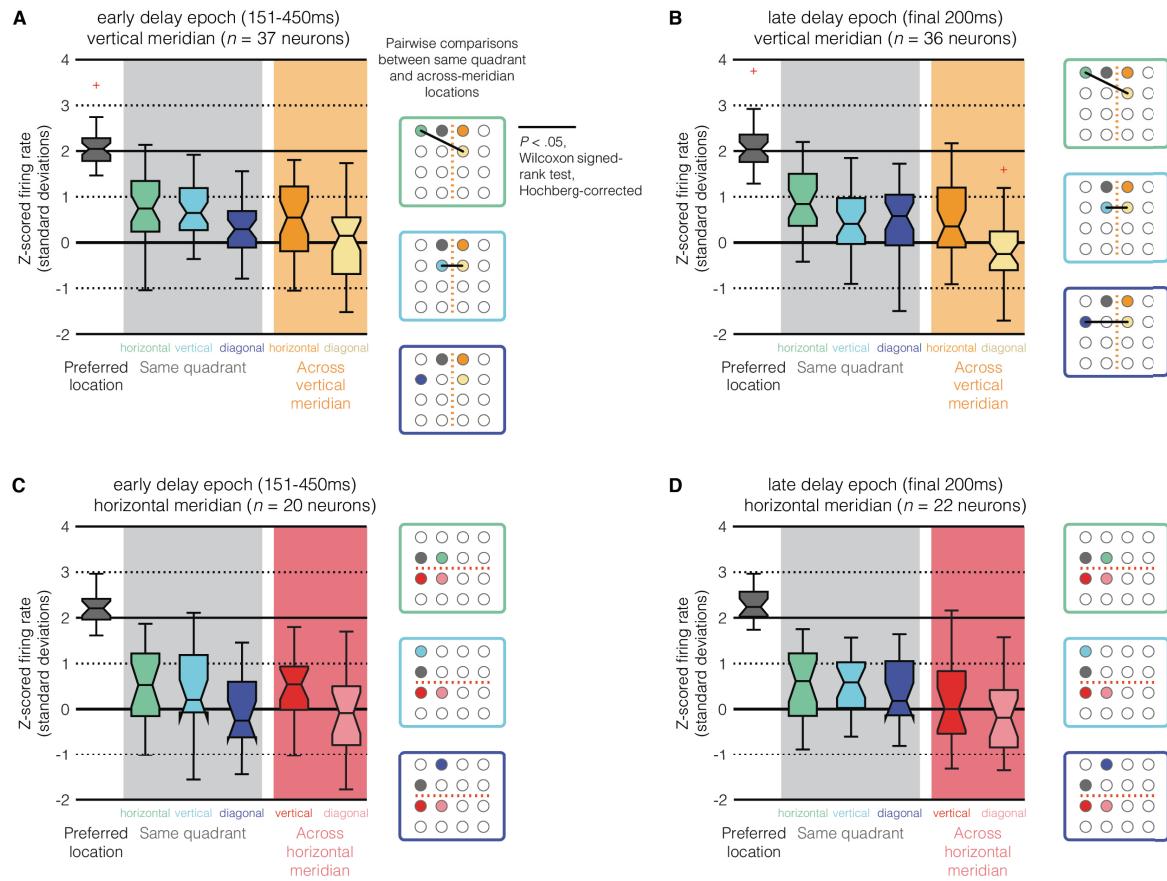
Supplementary Figure 3.5: Quadrantic Bias in Ensemble Representation for Each Eccentricity, Meridian, and Epoch. Similar to Figure 3.6. Shaded regions indicate 95% confidence intervals of the proportion. * $P < 0.05$, χ^2 test, Hochberg-corrected. (A) The vertical meridian biases representations of stimuli at the center locations of the array, but not peripheral locations during the stimulus epoch. (B) The same as (A), but for the horizontal meridian. An effect is visible for both central and peripheral representations. (C) The same as (A), but during the delay epoch. Both the central and peripheral representations are biased by the vertical meridian. (D) The same as (C), but for the horizontal meridian. A meridian bias is present in representations of central and peripheral locations.



Supplementary Figure 3.6: Quadrantic Bias in Saccade Endpoints in Long vs. Short Memory Delays. Similar to Figure 3.7B, except trials are split into short memory delay trials ($\leq 1000\text{ms}$, red), and long memory delay trials ($> 1000\text{ms}$, purple). The proportion of both subjects' saccades falling inside the grey box (y-axis) is shown for each quadrant (x-axis). Shaded regions indicate 95% confidence intervals of the proportion. ns $P > 0.05$, χ^2 test, Hochberg-corrected.



Supplementary Figure 3.7: Quadrantic Bias in Single Neuron Firing Rates for Neurons with Central Preferred Locations in Early vs. Late Delay Epoch. (A) Format is identical to Figure 3.3B, except only the period of 151-450ms after the beginning of the delay epoch is analyzed. (B) Identical to (A), except only the final 200ms of the delay epoch are analyzed.



Supplementary Figure 3.8: Quadrantic Bias in Single Neuron Firing Rates for Neurons with Peripheral Preferred Locations in Early vs. Late Delay Epoch.

Format is identical to Supplementary Figure 3.7, except for neurons with preferred locations in the periphery. (A) Firing rates during the early delay epoch for neurons with preferred locations along the vertical meridian. (B) Firing rates during the late delay epoch for neurons with preferred locations along the vertical meridian. (C) Firing rates during the early delay epoch for neurons with preferred locations along the horizontal meridian. Note that there are no significant differences between intraquadrant vs. extraquadrant locations. (D) Firing rates during the late delay epoch for neurons with preferred locations along the horizontal meridian. Note that there are no significant differences between intraquadrant vs. extraquadrant locations.

Model Summary After Removing Collinear Terms

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.330 ^a	.109	.109	.942816204
2	.341 ^b	.116	.116	.938875728
3	.351 ^c	.123	.123	.935265256
4	.354 ^d	.125	.125	.934261358
5	.360 ^e	.129	.129	.931989960
6	.360 ^f	.129	.129	.931966285
7	.360 ^g	.130	.130	.931775578

a. Predictors: (Constant), euclidean_distance

b. Predictors: (Constant), euclidean_distance, angle_difference

c. Predictors: (Constant), euclidean_distance, angle_difference,
crosses_horizontal_meridian_x_crosses_vertical_meridian

d. Predictors: (Constant), euclidean_distance, angle_difference,
crosses_horizontal_meridian_x_crosses_vertical_meridian, crosses_horizontal_meridian

e. Predictors: (Constant), euclidean_distance, angle_difference,
crosses_horizontal_meridian_x_crosses_vertical_meridian, crosses_horizontal_meridian,
crosses_vertical_meridian

f. Predictors: (Constant), euclidean_distance, angle_difference,
crosses_horizontal_meridian_x_crosses_vertical_meridian, crosses_horizontal_meridian,
crosses_vertical_meridian, eccentricity_difference

g. Predictors: (Constant), euclidean_distance, angle_difference,
crosses_horizontal_meridian_x_crosses_vertical_meridian, crosses_horizontal_meridian,
crosses_vertical_meridian, eccentricity_difference, angle_difference_x_eccentricity_difference

Coefficients After Removing Collinear Terms^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error				Tolerance	VIF
1 (Constant)	.672	.008		79.368	.000		
	euclidean_distance	-.329	.004	-.330	-88.415	.000	1.000
2 (Constant)	.670	.008		79.520	.000		
	euclidean_distance	-.209	.006	-.210	-32.880	.000	.340
	angle_difference	-.003	.000	-.148	-23.213	.000	.340
3 (Constant)	.734	.009		82.743	.000		
	euclidean_distance	-.181	.006	-.182	-28.015	.000	.327
	angle_difference	-.005	.000	-.266	-32.158	.000	.200
crosses_horizontal_meridi an_x_crosses_vertical_me ridian	.293	.013		.127	22.262	.000	.420
							2.379
4 (Constant)	.730	.009		82.395	.000		
	euclidean_distance	-.179	.006	-.179	-27.697	.000	.326
angle_difference crosses_horizontal_meridi an_x_crosses_vertical_me ridian	-.004	.000		-.233	-26.731	.000	.179
							5.576
	.314	.013		.136	23.669	.000	.413
							2.423

Chapter 3. A Quadrantic Bias in PFC WM Representation

	crosses_horizontal_meridian	-.121	.010	-.060	-11.768	.000	.518	1.929
5	(Constant)	.743	.009		83.736	.000		
	euclidean_distance	-.168	.006	-.169	-26.039	.000	.324	3.089
	angle_difference	-.002	.000	-.117	-10.686	.000	.114	8.771
6	crosses_horizontal_meridian_x_crosses_vertical_meridian	.443	.015	.192	29.312	.000	.317	3.152
	crosses_horizontal_meridian	-.357	.017	-.179	-21.218	.000	.192	5.204
	crosses_vertical_meridian	-.279	.016	-.140	-17.693	.000	.219	4.569
7	(Constant)	.743	.009		83.744	.000		
	euclidean_distance	-.169	.006	-.170	-26.121	.000	.322	3.109
	angle_difference	-.002	.000	-.116	-10.588	.000	.114	8.787
	crosses_horizontal_meridian_x_crosses_vertical_meridian	.442	.015	.192	29.276	.000	.317	3.153
	crosses_horizontal_meridian	-.357	.017	-.179	-21.208	.000	.192	5.205
	crosses_vertical_meridian	-.279	.016	-.139	-17.676	.000	.219	4.569
	eccentricity_difference	-.010	.005	-.008	-2.061	.039	.994	1.006

crosses_horizontal_meridi								
an_x_crosses_vertical_me	.441	.015	.192	29.236	.000	.317	3.153	
ridian								
crosses_horizontal_meridi								
an	-.356	.017	-.178	-21.202	.000	.192	5.205	
crosses_vertical_meridian								
	-.278	.016	-.139	-17.677	.000	.219	4.569	
eccentricity_difference								
	-.051	.009	-.038	-5.504	.000	.287	3.484	
angle_difference_x_eccen								
tricity_difference	.000	.000	.036	5.213	.000	.287	3.480	

a. Dependent Variable: firing_rate

Excluded Variables After Removing Collinear Terms^a

Model	Beta In	t	Sig.	Partial Correlatio	Collinearity Statistics			
					Tolerance	VIF	Minimum Tolerance	
1	angle_difference	-.148 ^b	-23.213	.000	-.091	.340	2.944	.340
	eccentricity_difference	-.015 ^b	-4.068	.000	-.016	.999	1.001	.999
	crosses_horizontal_meridi							
	an	-.084 ^b	-18.698	.000	-.074	.687	1.455	.687
	crosses_vertical_meridian							
		-.028 ^b	-6.497	.000	-.026	.756	1.323	.756
	eccentricity_difference_x_c							
	rosses_horizontal_meridia							
	n	.005 ^b	1.274	.203	.005	.997	1.003	.997

	eccentricity_difference_x_crosses_vertical_meridian	-.004 ^b	-1.126	.260	-.004	.997	1.003	.997
	angle_difference_x_eccentricity_difference	-.001 ^b	-.180	.857	-.001	.997	1.003	.997
	crosses_horizontal_meridian_x_crosses_vertical_meridian	.009 ^b	2.109	.035	.008	.715	1.400	.715
2	eccentricity_difference	-.009 ^c	-2.467	.014	-.010	.994	1.006	.338
	crosses_horizontal_meridian	-.044 ^c	-8.620	.000	-.034	.528	1.894	.261
	crosses_vertical_meridian	.017 ^c	3.587	.000	.014	.622	1.609	.279
3	eccentricity_difference_x_crosses_horizontal_meridian	.008 ^c	2.031	.042	.008	.996	1.004	.339
	eccentricity_difference_x_crosses_vertical_meridian	.000 ^c	-.093	.926	.000	.995	1.005	.338
	angle_difference_x_eccentricity_difference	.003 ^c	.832	.405	.003	.995	1.005	.338
	crosses_horizontal_meridian_x_crosses_vertical_meridian	.127 ^c	22.262	.000	.088	.420	2.379	.200
4	eccentricity_difference	-.008 ^d	-2.168	.030	-.009	.994	1.006	.199
	crosses_horizontal_meridian	-.060 ^d	-11.768	.000	-.046	.518	1.929	.179
	crosses_vertical_meridian	-.007 ^d	-1.407	.159	-.006	.591	1.693	.191
5	eccentricity_difference_x_crosses_horizontal_meridian	.008 ^d	2.048	.041	.008	.996	1.004	.200

	eccentricity_difference_x_crosses_vertical_meridian	.001 ^d	.205	.837	.001	.995	1.005	.199
	angle_difference_x_eccentricity_difference	.004 ^d	.984	.325	.004	.995	1.005	.200
4	eccentricity_difference	-.008 ^e	-2.200	.028	-.009	.994	1.006	.179
	crosses_vertical_meridian	-.140 ^e	-17.693	.000	-.070	.219	4.569	.114
	eccentricity_difference_x_crosses_horizontal_meridian	.006 ^e	1.741	.082	.007	.996	1.004	.179
	eccentricity_difference_x_crosses_vertical_meridian	.002 ^e	.485	.627	.002	.995	1.005	.179
	angle_difference_x_eccentricity_difference	.003 ^e	.943	.345	.004	.995	1.005	.179
5	eccentricity_difference	-.008 ^f	-2.061	.039	-.008	.994	1.006	.114
	eccentricity_difference_x_crosses_horizontal_meridian	.007 ^f	1.976	.048	.008	.996	1.005	.114
	eccentricity_difference_x_crosses_vertical_meridian	.001 ^f	.332	.740	.001	.994	1.006	.114
	angle_difference_x_eccentricity_difference	.004 ^f	1.063	.288	.004	.995	1.005	.114
6	eccentricity_difference_x_crosses_horizontal_meridian	.025 ^g	4.836	.000	.019	.501	1.997	.114
	eccentricity_difference_x_crosses_vertical_meridian	.013 ^g	2.518	.012	.010	.500	2.000	.114
	angle_difference_x_eccentricity_difference	.036 ^g	5.213	.000	.021	.287	3.480	.114

7	eccentricity_difference_x_c rosses_horizontal_meridia n	.012 ^h	1.688	.091	.007	.259	3.861	.114
	eccentricity_difference_x_c rosses_vertical_meridian	-.009 ^h	-1.245	.213	-.005	.281	3.560	.114

a. Dependent Variable: firing_rate

b. Predictors in the Model: (Constant), euclidean_distance

c. Predictors in the Model: (Constant), euclidean_distance, angle_difference

d. Predictors in the Model: (Constant), euclidean_distance, angle_difference, crosses_horizontal_meridian_x_crosses_vertical_meridian

e. Predictors in the Model: (Constant), euclidean_distance, angle_difference, crosses_horizontal_meridian_x_crosses_vertical_meridian, crosses_horizontal_meridian

f. Predictors in the Model: (Constant), euclidean_distance, angle_difference, crosses_horizontal_meridian_x_crosses_vertical_meridian, crosses_horizontal_meridian, crosses_vertical_meridian

g. Predictors in the Model: (Constant), euclidean_distance, angle_difference, crosses_horizontal_meridian_x_crosses_vertical_meridian, crosses_horizontal_meridian, crosses_vertical_meridian, eccentricity_difference

h. Predictors in the Model: (Constant), euclidean_distance, angle_difference, crosses_horizontal_meridian_x_crosses_vertical_meridian, crosses_horizontal_meridian, crosses_vertical_meridian, eccentricity_difference, angle_difference_x_eccentricity_difference

Final Model Summary Including Collinear Terms

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
14	.375 ⁿ	.141	.141	.925767433

n. Predictors: (Constant), euclidean_distance, angle_difference, euclidean_distance_x_angle_difference, angle_difference_x_crosses_horizontal_meridian, crosses_vertical_meridian, crosses_horizontal_meridian, euclidean_distance_x_crosses_vertical_meridian, eccentricity_difference, euclidean_distance_x_eccentricity_difference, angle_difference_x_crosses_vertical_meridian

Final Model Coefficients Including Collinear Terms^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error				Toler ance	VIF
14 (Constant)	.990	.012		79.955	.000		
euclidean_distance	-.436	.013	-.437	-33.210	.000	.077	12.913
angle_difference	-.006	.000	-.303	-16.890	.000	.042	24.023
euclidean_distance_x_angle_difference	.002	.000	.387	15.611	.000	.022	45.779
angle_difference_x_crosses_horizontal_meridian	.001	.000	.100	4.751	.000	.031	32.686
crosses_vertical_meridian	-.185	.031	-.092	-5.918	.000	.055	18.165
crosses_horizontal_meridian	-.261	.031	-.131	-8.548	.000	.058	17.361

euclidean_distance_x_c								
rosses_vertical_meridia	.089	.017		.124	5.094	.000	.023	44.102
n								
eccentricity_difference	-.149	.016		-.111	-9.239	.000	.093	10.762
euclidean_distance_x_e				.109	9.032	.000	.093	10.774
ccentricity_difference	.068	.007						
angle_difference_x_cro				-.044	-1.998	.046	.027	36.562
sses_vertical_meridian	-.001	.000						

a. Dependent Variable: firing_rate

Final Model Excluded Variables Including Collinear Terms^a

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics			
					Tolerance	VIF	Minimum Tolerance	
14	eccentricity_difference_x_cr	.001°	.191	.848	.001	.321	3.112	.022
	osses_horizontal_meridian							
	eccentricity_difference_x_cr	-.007°	-1.167	.243	-.005	.343	2.912	.022
	osses_vertical_meridian							
	angle_difference_x_eccen	-.015°	-1.235	.217	-.005	.086	11.667	.022
15	tricity_difference							
	crosses_horizontal_merid	.018°	.833	.405	.003	.028	35.401	.011
16	n_x_crosses_vertical_merid							
	ian	-.052°	-1.759	.079	-.007	.015	65.617	.009

o. Predictors in the Model: (Constant), euclidean_distance, angle_difference, euclidean_distance_x_angle_difference, angle_difference_x_crosses_horizontal_meridian, crosses_vertical_meridian, crosses_horizontal_meridian, euclidean_distance_x_crosses_vertical_meridian, eccentricity_difference, euclidean_distance_x_eccentricity_difference, angle_difference_x_crosses_vertical_meridian

Supplementary Table 3.1: Stepwise regression indicates significance of meridian crossings.

CHAPTER 4

CORRELATED VARIABILITY MODIFIES WORKING MEMORY FIDELITY IN PRIMATE PREFRONTAL NEURONAL ENSEMBLES

The previous two studies demonstrated the presence of r_{sc} in area 8a during different behavioral states, including WM maintenance. r_{sc} can dramatically modify the amount of information that can be decoded from a neuronal ensemble, but it is unknown whether the pattern of r_{sc} associated with WM maintenance also affects WM coding. More generally, we know very little about how ensemble of neurons in 8a represent WM, and how it differs from single neuron representations. We found that an ensemble's r_{sc} structure could facilitate or impair the readout of WM representations, depending on the size of the ensemble and tuning properties of its constituent neurons. We also found that neurons with poor WM selectivity could still improve coding when part of an ensemble by shaping the ensemble's r_{sc} structure.

This chapter is adapted from Leavitt, M., Pieper, F., Sachs, A., Martinez-Trujillo, J.C. (2017) Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles, *PNAS*, 114(12): E294-E205.

4.1 Abstract

Neurons in the primate lateral prefrontal cortex (LPFC) encode working memory (WM) representations via sustained firing, a phenomenon hypothesized to arise from recurrent dynamics within ensembles of interconnected neurons. Here we tested this hypothesis by using microelectrode arrays to examine spike count correlations (r_{sc}) in LPFC neuronal ensembles during a spatial WM task. We found a pattern of pairwise r_{sc} during WM maintenance indicative of stronger coupling between similarly-tuned neurons and increased inhibition between dissimilarly-tuned neurons. We then used a linear decoder to quantify the effects of the high-dimensional r_{sc} structure on information coding in the neuronal ensembles. We found that the r_{sc} structure could facilitate or impair coding, depending on the size of the ensemble and tuning properties of its constituent neurons. A simple optimization procedure demonstrated that near-maximum decoding performance could be achieved using a relatively small number of neurons. These WM-optimized subensembles were more r_{signal} -diverse and anatomically dispersed than predicted by the statistics of the full recorded population of neurons, and they often contained neurons that were poorly WM-selective, yet enhanced coding fidelity by shaping the ensemble's r_{sc} structure. We observed a pattern of r_{sc} between LPFC neurons

indicative of recurrent dynamics as a mechanism for WM-related activity, and the r_{sc} structure can increase the fidelity of WM representations. Thus, WM coding in LPFC neuronal ensembles arises from a complex synergy between single neuron coding properties and multidimensional, ensemble-level phenomena.

4.2 Introduction

To interact with a complex, dynamic environment, organisms must be capable of maintaining and manipulating information that is no longer available to their sensory systems. This capability, when applied transiently (i.e. for milliseconds to seconds), is referred to as working memory (WM) (Baddeley and Hitch, 1974)—a hallmark of intelligence and crucial component of goal-directed behavior (Miller and Cohen, 2001). In 1949, Hebb postulated that sustained neuronal activity in the absence of stimulus input could serve as the neural substrate for WM (Hebb, 2005). Fuster and Alexander later discovered neurons in the lateral prefrontal cortex (LPFC) of monkeys that exhibited sustained firing during WM tasks (Fuster and Alexander, 1971). Subsequent neurophysiological studies have corroborated that neuronal activity in the LPFC and other regions can represent WM for visual-mnemonic space (Battuev, 1986; Gnadt and Andersen, 1988; Funahashi et al., 1989),

as well as non-spatial visual features (Miller et al., 1996; Mendoza-Halliday et al., 2014; Riley and Constantinidis, 2015).

Electrophysiological studies of spatial WM have traditionally relied on recording from one or a few neurons simultaneously (Riley and Constantinidis, 2015). However, the neuronal computations that underlie sophisticated behaviors such as WM require the coordinated activity of many neurons within and across brain networks (Quijan Quiroga and Panzeri, 2009). We currently lack a clear understanding of how single neuron coding properties scale to neuronal ensembles. Can the properties of an ensemble be predicted by aggregating the individually and independently measured properties of its constituent neurons? The answers to this and related questions hinge on how ensembles are affected by phenomena that emerge from interactions *between* neurons.

The sustained activity presumed to underlie WM maintenance is thought to be achieved by increasing the strength of recurrent excitation and lateral inhibition between neurons within an ensemble (Amit and Brunel, 1997; Camperi and Wang, 1998; Compte et al., 2000; Durstewitz et al., 2000; Wang, 2001; Constantinidis and Wang, 2004; Wimmer et al., 2014). These dynamics should modify patterns of correlated firing between neurons in a manner dependent on differences in their

tuning properties. Such a pattern can be quantitatively characterized by two measurements: The first is signal correlation (r_{signal}), the similarity of two neurons' responses to a set of different stimuli or experimental conditions. The second is spike count (or noise) correlations (r_{sc}), the similarity in the variability of two neurons' responses to the same stimulus or experimental condition (Averbeck et al., 2006).

Given a fixed ensemble of neurons (and thus a constant r_{signal} structure), changes in r_{sc} can have profound effects on information coding (Shadlen and Newsome, 1994; Zohary et al., 1994; Abbott and Dayan, 1999; Averbeck et al., 2006; Cohen and Kohn, 2011; Moreno-Bote et al., 2014). For example, spatial attention improves neural coding in visual cortex primarily by reducing r_{sc} (Cohen and Maunsell, 2009; Mitchell et al., 2009; Tremblay et al., 2014). Another study reported that increased r_{sc} improved perceptual discrimination in macaque area S2 (Romo et al., 2003). These results are difficult to extend to WM coding in LPFC. Furthermore, there are relatively few studies investigating r_{sc} in the LPFC (Constantinidis et al., 2001a; Constantinidis and Goldman-Rakic, 2002; Cohen and Kohn, 2011; Qi and Constantinidis, 2012; Katsuki et al., 2014; Tremblay et al., 2014; Markowitz et al., 2015); and only one of these studies directly examined the effects of r_{sc} on

information coding (Tremblay et al., 2014). Prior results examining pairwise correlations are also difficult to extrapolate to larger neuronal ensembles, which have a complex, multidimensional r_{sc} structure that cannot be characterized by pairwise measurements alone (Moreno-Bote et al., 2014). Currently it remains unknown whether and how r_{sc} structure modulates the fidelity of WM coding in LPFC neuronal ensembles.

We used microelectrode arrays to record from neuronal ensembles in LPFC of two monkeys while they performed an oculomotor delayed-response task, and assessed ensemble information content using a linear decoder. We found that r_{sc} varied as a function of r_{signal} during WM maintenance in a manner predicted by a recurrent-excitation and lateral-inhibition scheme. Using all simultaneously recorded neurons, the decoder could reliably predict which of 16 locations was being remembered. We also devised procedures to systematically investigate how WM coding varies across the “configuration space” of potential neuronal ensembles. Removing the r_{sc} structure could increase or decrease the information content of neuronal ensembles across the configuration space. However, the intrinsic r_{sc} structure improved WM coding in smaller neuronal sub-ensembles of neurons optimized for WM representation. These optimized ensembles had a stereotyped

r_{signal} distribution with peaks at zero and extreme negative values, and spanned farther across the cortical surface than predicted by the statistics of the full population of recorded LPFC units. Finally, we observed individual units that did not encode WM in isolation (“non-selective” neurons), but still contributed to WM coding when part of an ensemble by altering the r_{sc} structure.

4.3 Results

Two adult male *Macaca fascicularis* (subjects ‘JL’ and ‘F’) performed an oculomotor delayed-response task (Figure 4.1a) while we recorded from neuronal ensembles in left LPFC area 8A, anterior to the arcuate sulcus and posterior to the principal, using chronically-implanted 96-channel microelectrode arrays (Figure 4.1b). The neural correlates of WM for spatial locations have been extensively documented in this brain region(Riley and Constantinidis, 2015). The target stimulus could appear at any one of 16 possible locations, arranged in a uniformly spaced 4×4 grid around a central fixation point. We collected spike data from a total of 545 single- and multiunits across 12 recording sessions, out of which 417 (76%) exhibited sustained activity and selectivity during the delay epoch ($P < 0.05$, Kruskal-Wallis; firing rate × location—see Materials and Methods). We included both multi- and single units in our analyses, as in similar previous studies (Cohen and Maunsell, 2009; Cohen and

Kohn, 2011; Ruff and Cohen, 2014; Tremblay et al., 2014). A unit's preferred location during a given epoch was defined as the location that elicited the largest response averaged over that epoch (Figure 4.1c & d). Subjects made incorrect choices about the stimulus location in <1% of completed trials. Only correct, completed trials were included for analysis.

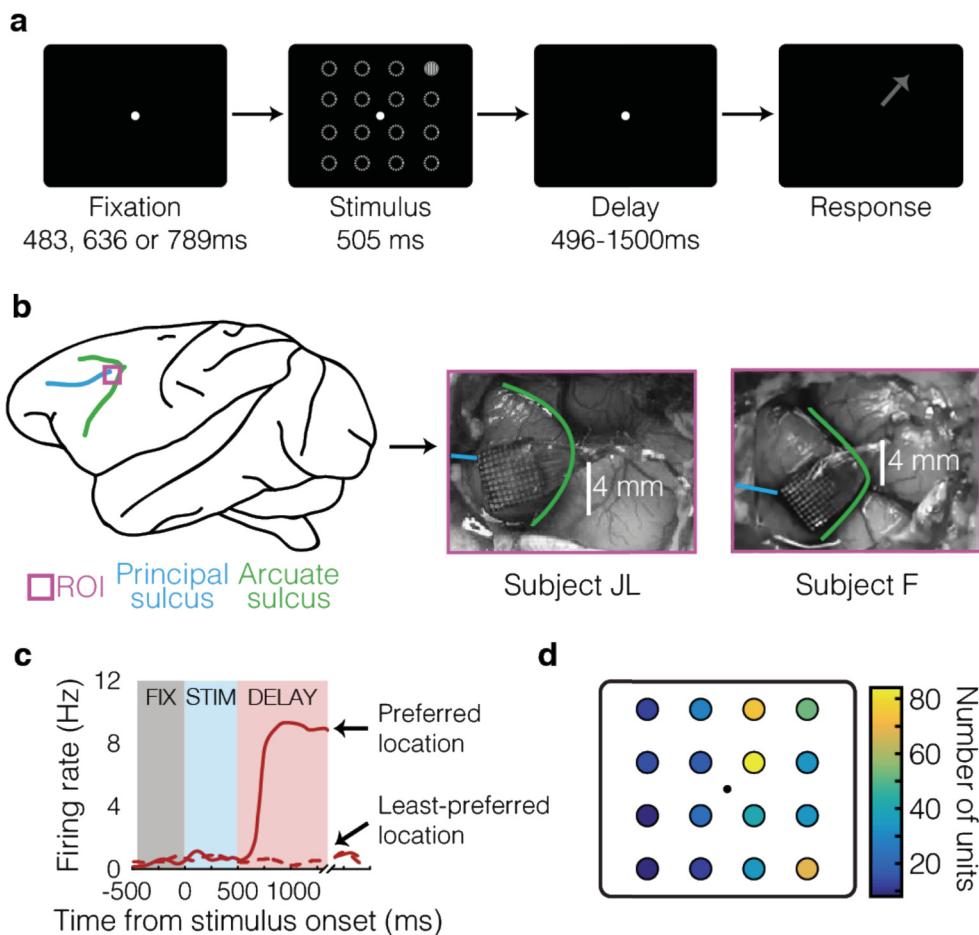


Figure 4.1: Task, method, and single-cell data. (a) Overview of oculomotor delayed-response task. The dashed circles indicate potential cue locations, and are shown for illustrative purposes only and are not present in the task. The arrow represents the

correct saccade direction. (b) Array implantation sites and anatomical landmarks in both subjects. (c) Example delay-selective neuron. (d) Distribution of delay-selective units' preferred locations. FIX, fixation; ROI, region of interest; STIM, stimulus.

4.3.1 Task-related modulation of spike count correlations (r_{sc})

We computed r_{sc} between pairs of neurons (pairwise r_{sc} —Materials and Methods) during the fixation, stimulus, and delay epochs. r_{sc} can covary with firing rate (la Rocha et al., 2007; Cohen and Kohn, 2011), so to ensure that differences in r_{sc} across epochs were not confounded by differences in firing rates, we implemented a distribution matching procedure (Materials and Methods). We replicated two findings from previous studies: mean pairwise r_{sc} was significantly above zero in each task epoch (Figure 4.2a; $P < 0.005$ for all epochs, bootstrap test—Materials and Methods); and r_{sc} varied as a function of tuning similarity, which we quantified as signal correlation between pairs of neurons during the delay epoch (r_{signal} —Materials and Methods; Figure 4.2b) (Constantinidis et al., 2001a; Constantinidis and Goldman-Rakic, 2002; Qi and Constantinidis, 2012; Katsuki et al., 2014; Wimmer et al., 2014; Markowitz et al., 2015). Specifically, we found that the median pairwise r_{sc} was consistently larger for similarly-tuned neuron pairs (defined as $r_{signal} > 0.25$) as compared to dissimilarly-tuned pairs (defined as $r_{signal} < -0.25$) ($P < 0.001$ for all epochs, bootstrap test—Materials and Methods). We also found that mean pairwise

r_{sc} was greater during the fixation and delay epochs as compared to the stimulus epoch (Figure 4.2a; $P < 0.001$ for both fixation vs. stimulus and delay vs. stimulus, bootstrap test). Most importantly, we found that the relationship between r_{sc} and r_{signal} changed across task epochs (Figure 4.2b); specifically, median pairwise r_{sc} for similarly-tuned neurons was larger during the delay epoch than during the fixation and stimulus epochs ($P < 0.001$ for both comparisons, bootstrap test; Figure 4.2c), and median pairwise r_{sc} between dissimilarly-tuned neurons was lower during the stimulus and delay epochs than during the fixation epoch ($P < 0.001$ for both comparisons, bootstrap test; Figure 4.2c). These results indicate that WM maintenance modifies pairwise r_{sc} in the LPFC in a manner consistent with a recurrent excitation, lateral inhibition scheme (Amit and Brunel, 1997; Camperi and Wang, 1998; Compte et al., 2000; Durstewitz et al., 2000; Wang, 2001; Laing et al., 2002; Constantinidis and Wang, 2004).

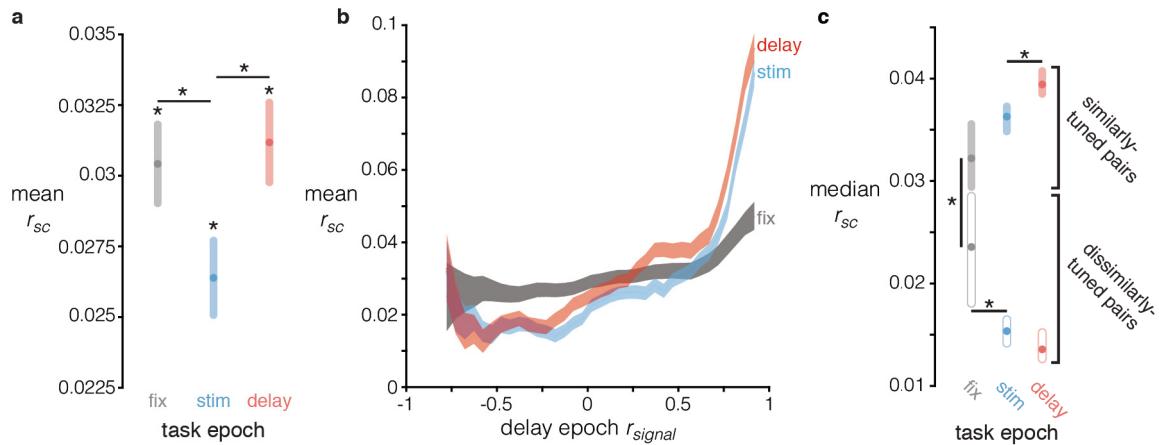


Figure 4.2: Measures of correlated variability and its effects on WM information in full ensembles. (a) Mean pairwise r_{sc} (y-axis) across task epochs (x-axis), controlling for firing rate (Online Methods). The mean is computed across all 2000 subsampled distributions, and shaded regions are s.e.m. calculated using the sample size of a single subsampled distribution ($n = 10,535$ pairs). * $P < 0.001$, bootstrap test. (b) Mean r_{sc} for each task epoch (y-axis) as a function of delay epoch r_{signal} (x-axis). The same subsampling procedure as in (a) was applied, then the r_{sc} of each neuron pair was binned based on its corresponding r_{signal} , and the mean r_{sc} computed in each bin. r_{signal} bins are size = 0.2, stepped by increments of 0.05. The shaded regions are s.e.m., calculated using the sample size of the corresponding r_{signal} bin. (c) Median r_{sc} for similarly-tuned neuron pairs ($r_{signal} > 0.25$) and dissimilarity-tuned neuron pairs ($r_{signal} < -0.25$) in each task epoch. The colored region around each point represents the bootstrapped 99.9% confidence interval of the median, derived from 2000 bootstrap iterations. Non-overlapping colored regions indicate $P < 0.001$, bootstrap test; however pairwise comparisons that are visually ambiguous have explicitly-marked (*) significant differences. FIX, fixation; STIM, stimulus.

4.3.2 Quantifying information content in neuronal ensembles using linear decoders

Pairwise measurements of r_{sc} are insufficient for predicting the effects of r_{sc} structure on ensemble information in large, multidimensional ensembles with heterogeneous tuning (Moreno-Bote et al., 2014). Furthermore, analytical methods

for determining the effects of r_{sc} structure on information content can be complicated to calculate for large stimulus sets, and can also be inaccurate unless applied to data consisting of hundreds of trials per stimulus (Moreno-Bote et al., 2014; Kanitscheider et al., 2015). Linear decoders are demonstrably well-suited for extracting low-dimensional representations from high-dimensional neuronal ensemble data and for directly assessing the impact of r_{sc} structure on ensemble information content, and thus offer a pragmatic solution to the issues of dimensionality and correlated variability (Rigotti et al., 2013; Moreno-Bote et al., 2014).

Previous studies have decoded the identity of stimuli maintained in spatial (Markowitz et al., 2011; Meyers et al., 2012) and non-spatial WM (Meyers et al., 2012; Mendoza-Halliday et al., 2014) in pseudopopulations of LPFC neurons, typically using sets of 2-8 unique stimuli. We were able to reliably decode which of 16 target locations was being held in WM during the delay epoch by applying a linear support vector machine (SVM-Materials and Methods) to simultaneously-recorded ensemble data (Supplementary Figure 4.1a; max = 77%; mean across sessions = 52%).

Examining ensembles consisting of every simultaneously-recorded neuron and/or only tuned neurons is a standard practice in neurophysiology. However, this practice assumes that all of the examined neurons contribute to coding, an assumption difficult to verify. It is possible that a subset of the recorded neurons can represent nearly as much information as the entire ensemble, and that such a subset could form a “unit” of information coding that is read out by a downstream mechanism. Furthermore, the information-modifying effects of the r_{sc} structure have been proposed to increase with ensemble size, but most of our knowledge about these scaling effects is drawn from extrapolations of pairwise recordings, which do not necessarily predict ensemble-level effects (Zohary et al., 1994; Averbeck et al., 2006; Averbeck and Lee, 2006; Shamir and Sompolinsky, 2006; Cohen and Kohn, 2011; Moreno-Bote et al., 2014). Thus, examining how information coding varies across different subsets or sub-ensembles of simultaneously-recorded neurons—what we refer to as the ensemble “configuration space”—could reveal insights overlooked by the constraint of analyzing only a single, fixed ensemble of all tuned neurons recorded during an experiment.

In order to determine how WM coding scales across ensemble configurations, we devised “ensemble construction” procedures. The procedures consisted of

iteratively constructing neuronal ensembles by drawing units from the pool of all simultaneously-recorded neurons, and quantifying the WM information using the decoder (Figure 4.3). We implemented two procedure variants. We refer to the first variant as the “best individual unit” method. This method examines the assumption that a neuronal ensemble is simply a collection of the best individually tuned neurons; accordingly, the method is agnostic to between-neuron information such as the ensemble r_{sc} and r_{signal} structures. It was implemented as follows (Figure 4.3a): We began by using the decoder to assess the WM information content of each individual unit in a single recording session. We then rank ordered the units based on their information content. An ensemble of two neurons was constructed using the two most informative neurons, and the decoding analysis performed on the ensemble of two neurons. This process was repeated iteratively, performing the decoding analysis using the n most informative neurons in the session, until the ensemble consisted of all the neurons recorded in the session. The results from applying the “best individual unit” method to an example session are depicted in Figure 4.3c (teal).

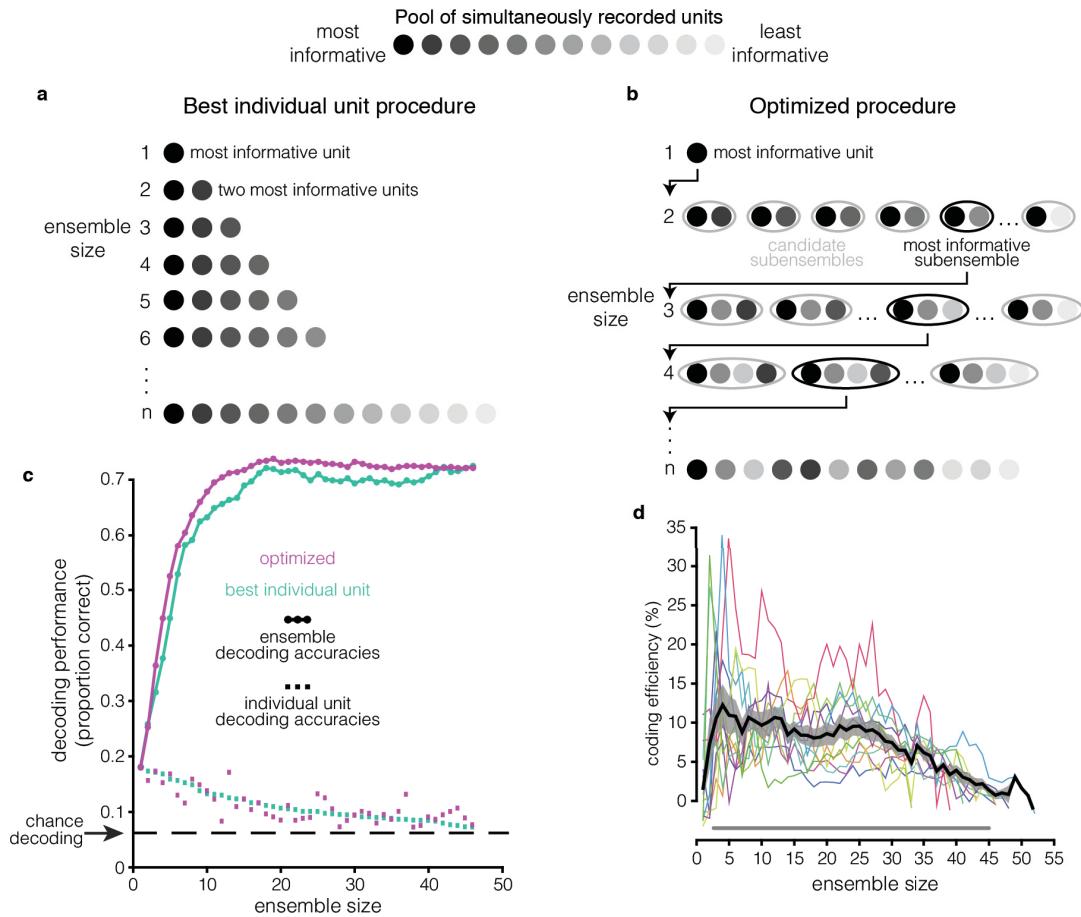


Figure 4.3: Accounting for between-neuron phenomena increases ensemble efficiency. Visualization of the (a) “best individual unit” ensemble construction procedure and (b) “optimized” ensemble construction procedure. Each circle represents a unit, and the shading represents that unit’s information content, as assessed using the decoder. (c) Decoding results for the best individual unit (teal) and optimized procedures (violet), applied to a single example session. The continuous line plot with circular markers shows the ensemble decoding accuracy (y-axis) as a function of size (x-axis). The square markers at the bottom of the plot denote the decoding accuracy (y-axis) of the individual unit added to the ensemble at a given size (x-axis). Both methods yield identical results for ensembles of the maximum size because these ensembles are identical; they consist of every simultaneously recorded unit in the session (i.e. the full ensemble). (d) Coding efficiency of the optimized method relative to the best individual unit method (y-axis) as a function of ensemble size (x-axis). Coding efficiency is quantified as

$((accuracy_{optimized}/accuracy_{best\ individual\ unit})-1) \times 100$. Colored lines are values for individual sessions. The thick black line is the across-session mean, and the gray shaded area is the S.E.M. The gray line running along the bottom indicates ensemble sizes for which the optimized method is significantly more efficient than the best individual unit method ($P < 0.05$, paired t-test, Hochberg-corrected).

The second variant of our ensemble construction procedure, which we refer to as the “optimized” method (Figure 4.3b; also referred to as “greedy forward selection” in the machine learning literature (Guyon and Elisseeff, 2003)), was designed to optimize WM information for a given ensemble size, accounting for the r_{sc} and r_{signal} structures that were ignored in the best individual unit method. The optimized method also began by rank ordering the information content of individual neurons within a given recording session using the decoder. However instead of starting with the two most informative individual neurons, as in the “best individual unit” method, we instead constructed all possible neuron pairs that contained the most informative unit. We then identified the most informative of these pairs, as assessed using the decoder. The most informative pair was then combined with each remaining neuron to generate a set of trios, from which the most informative trio was identified and used as the basis for of the most informative quartet, and so on until the ensemble consisted of all the neurons recorded in the session. Figure 4.3c shows the results of applying the optimized

method to an example session. Unlike the best individual unit method, the optimized method does not consider the information content of an individual unit in isolation, but instead how the neuron contributes to the information content of the ensemble to which it belongs.

The results of the two ensemble building methods are directly compared in Figure 4.3c & d. Notice that the optimized method yields more informative ensembles of a given size than the best individual unit method. We refer to this property—differing WM information content in ensembles of identical size—as *coding efficiency*; the optimized ensembles are more efficient than the best individual unit ensembles. Note that coding efficiency can also refer to the converse idea—identical WM information in ensembles of different size. We quantified coding efficiency as the percent change in decoding accuracy of the optimized method relative to the best individual unit method (similar to $\Delta_{shuffle}$ —Materials and Methods—Figure 4.3d). The optimized method becomes significantly more efficient than the best individual unit method starting at ensemble size $n = 3$ ($P < 0.05$, paired t-test, Hochberg-corrected). For certain sessions and ensemble sizes the relative efficiency can exceed 30%. Furthermore, the decoding performance approaches saturation more quickly in the optimized ensembles (Supplementary

Figure 4.2). Achieving 95% of maximum decoding accuracy using the optimized method requires only ~25% of the units recorded in a given session (~11 units), whereas the best individual unit method requires ~33% of the units (~14 units). In “random” ensembles—ensembles generated by randomly subsampling n units from a given recording session—approximately 85% of the units are necessary to reach 95% of maximum decoding accuracy. These results demonstrate that neuronal ensembles in the LPFC encode more information than single neurons, the most informative ensembles are not necessarily composed of the most informative individual units, and that a relatively small subset of neurons can represent nearly as much WM information as the full recorded population.

4.3.3 Effects of r_{sc} and r_{signal} structures on WM coding efficiency

In order to dissociate the effects of the r_{sc} and r_{signal} structures on WM coding efficiency in the optimized ensembles, we constructed new ensembles using the optimized procedure on firing rate data from which the r_{sc} structure had been removed via shuffling; the classifier was trained and tested on shuffled data for each ensemble size. We then compared the information content of these “ r_{signal} -only” ensembles to the information content of ensembles generated using the original, r_{sc} structure-intact data, which we now refer to as the “ $r_{signal} + r_{sc}$ ” ensembles. The

results for all three methods (best individual unit, $r_{signal} + r_{sc}$ optimized, and r_{signal} -only optimized) applied to an example session are compared in Figure 4.4a. The r_{signal} -only ensembles contain significantly more WM information than the best individual unit ensembles across sizes ranging from 2-47 neurons ($P < 0.05$, paired t-test, Hochberg-corrected; Figure 4.4b). However, the effect of the r_{sc} structure is variable: the $r_{signal} + r_{sc}$ ensembles are more efficient than the r_{signal} -only ensembles at smaller ensemble sizes, while this effect inverts at larger ensemble sizes ($P < 0.05$ for ensemble sizes of 11-15 and 43-45 neurons, paired t-test, Hochberg-corrected; Figure 4.4b-c). These changes in WM coding efficiency effected by the r_{sc} structure can reach $\pm 15\%$ across different recording sessions and ensemble sizes (Figure 4.4c), enough to double (or nullify) efficiency increases afforded by the r_{signal} structure alone. These results indicate that the r_{sc} structure significantly impacts WM coding, and can do so in a manner that varies non-monotonically with ensemble size. These results cannot be ascribed to idiosyncrasies of the SVM decoder, as repeating the same analyses using logistic regression yields similar results (Supplementary Figure 4.3). We also found a similar—though less consistent—effect during stimulus presentation, with considerably greater session-to-session variability (Supplementary Figure 4.4).

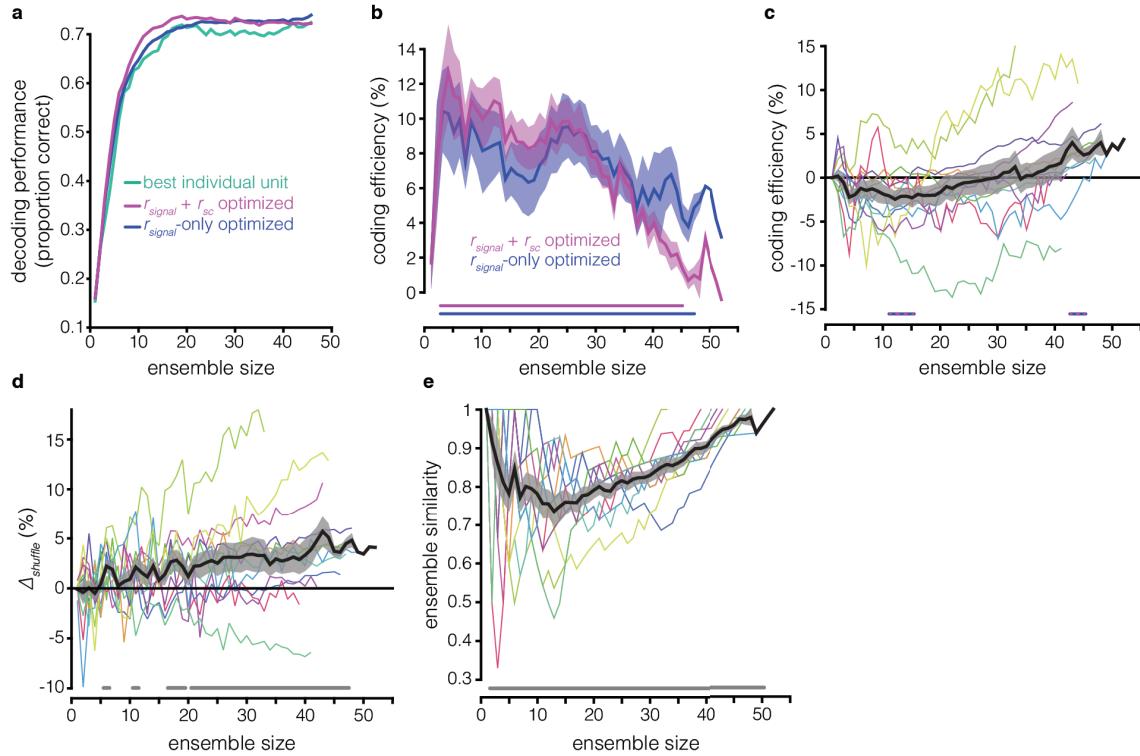


Figure 4.4: Effects of r_{sc} structure on ensemble coding efficiency and composition. (a) Decoding accuracy (y-axis) as a function of ensemble size (x-axis) for the best individual unit (teal), $r_{signal} + r_{sc}$ (violet), and r_{signal} -only (blue) methods for the same example session as in Figure 4.3c. Note that for the r_{signal} -only ensembles, the classifier was trained and tested on r_{sc} -shuffled data, whereas for the $r_{signal} + r_{sc}$ and best individual unit ensembles, the classifier was trained and tested on r_{sc} -intact data. (b) Coding efficiency of $r_{signal} + r_{sc}$ ensembles and r_{signal} -only ensembles, relative to the best individual unit ensembles (y-axis), as a function of ensemble size (x-axis). The violet line running along the bottom indicates ensemble sizes for which the $r_{signal} + r_{sc}$ ensembles are significantly more efficient than the best individual unit ensembles ($P < 0.05$, paired t-test, Hochberg-corrected); the blue line is similar, but for r_{signal} -only ensembles vs. best individual unit ensembles. Note that the coding efficiency of $r_{signal} + r_{sc}$ ensembles relative to best individual unit ensembles was previously shown in Figure 4.3d. (c) Coding efficiency of r_{signal} -only ensembles relative to $r_{signal} + r_{sc}$ ensembles; similar to Figure 4.3d. A positive value indicates that shuffling out the r_{sc} structure improves decoding. The striped blue and violet lines running along the bottom indicate ensemble sizes for which the efficiency of $r_{signal} + r_{sc}$ ensembles and r_{signal} -only ensembles are significantly different ($P < 0.05$,

paired t-test, Hochberg-corrected). (d) Decoding performance of r_{sc} -shuffled vs. r_{sc} -intact ensembles ($\Delta_{shuffle}$, y-axis) as a function of ensembles size (x-axis) for random ensembles. Ensembles were generated by randomly subsampling n units from the full recorded population in a given session. The gray lines running along the bottom indicate ensemble sizes for which the r_{sc} -shuffled vs. r_{sc} -intact ensembles are significantly different ($P < 0.05$, paired t-test, Hochberg-corrected). (e) Similarity between $r_{signal} + r_{sc}$ ensembles and r_{signal} -only ensembles (y-axis) as a function of ensemble size (x-axis). Ensemble similarity is quantified as the proportion of units common to the two ensembles for a given size. Note that ensemble similarity is 1 for ensembles of size $n = 1$, and for the largest ensemble size in a given session, because both ensemble-building procedures begin with the same unit, and the largest ensemble in each session consists of every simultaneously-recorded unit in that session. The gray line running along the bottom indicates ensemble sizes for which the similarity of the $r_{signal} + r_{sc}$ ensembles and r_{signal} -only ensembles is significantly less than 1 ($P < 0.05$, z-test of proportion, Hochberg-corrected).

It is possible that the observed effects of the r_{sc} structure on WM coding are simply a property of an ensemble's size, regardless of whether the ensemble is optimized for WM representation. In order to resolve this ambiguity, we compared the decoding performance of the random ensembles in which r_{sc} structure was intact vs. shuffled (Figure 4.4d). We found that shuffling out the r_{sc} structure significantly improved decoding in most ensembles of 6 or more units ($P < 0.05$, paired t-test, Hochberg-corrected), and that the magnitude of the decoding improvement was robustly and significantly correlated with the size of the ensemble in 8 out of 12 recording sessions (Spearman's $\rho \geq .53$; $P < 0.001$; Supplementary Figure 4.5a). While the r_{sc} structure appears to consistently impair decoding at the largest

ensemble sizes (Supplementary Figure 4.5b), these results demonstrate that WM coding in a neuronal ensemble consisting of randomly selected neurons will be impaired by the r_{sc} structure in a manner proportional to the size of the ensemble, but that the r_{sc} structure can actually improve WM coding in $r_{sc} + r_{signal}$ optimized ensembles.

4.3.4 Different ensemble configurations optimize WM coding when the r_{sc} structure is intact vs. removed

The previous results demonstrate that accounting for an ensemble's r_{sc} structure can significantly alter estimates of its WM information content. A complementary question is whether accounting for the r_{sc} structure also alters estimates of individual neurons' contributions to an ensemble's WM coding. Are ensembles that maximize coding efficiency when the r_{sc} structure is intact composed of the same neurons that maximize coding efficiency when the r_{sc} structure is shuffled out? In order to answer this question, we examined the proportion of units common to both the $r_{signal} + r_{sc}$ and r_{signal} -only ensembles for each ensemble size (Figure 4.4e). The proportion is significantly less than 1 for ensemble sizes of 2 to 50 neurons ($P < 0.05$, z-test of proportions, Hochberg-corrected), indicating that the ensembles generated by the two methods are not identical; the similarity within an individual

session can be as low as 33%. The $r_{signal} + r_{sc}$ and r_{signal} -only procedures also recruited units into ensembles in different sequences (Spearman's $\rho < 1$ in all sessions, mean $\rho = 0.713$; $P < 0.05$, Bonferroni-corrected; Supplementary Figure 4.6). These results demonstrate that different subpopulations of neurons optimize WM coding when the intrinsic r_{sc} structure is present vs. when it is absent, though some neurons strongly contribute to WM coding regardless of an ensemble's r_{sc} structure.

4.3.5 Ensembles optimized for WM representation are r_{signal} diverse and anatomically dispersed

One of our earlier analyses demonstrated that near-maximum decoding performance can be achieved with a relatively small proportion of recorded units, and that accounting for an ensemble's r_{signal} and/or r_{sc} structure can further enhance WM coding. If the WM coding is optimized in these ensembles by maximizing their representation of the stimulus space, their r_{signal} distributions should be broader than those of the full recorded ensembles. We tested this by examining the $r_{signal} + r_{sc}$ and r_{signal} -only ensembles that achieved $\geq 95\%$ of maximum decoding performance in each session (which we refer to as "near-max" ensembles). Indeed, we found that the r_{signal} distributions of the near-max $r_{signal} + r_{sc}$ ensembles, r_{signal} -only ensembles, and full ensembles were all significantly different from each other (Figure 4.5a; $P <$

0.001 for all comparisons, χ^2 -test, Bonferroni-corrected). The width of the r_{signal} distribution, measured as the mean absolute deviation (Materials and Methods), was larger for the near-max $r_{signal} + r_{sc}$ and r_{signal} -only ensembles than for all units (Figure 4.5b; $P << 0.001$ for both, F -test, Bonferroni-corrected), and larger in the r_{signal} -only than the $r_{signal} + r_{sc}$ ensembles ($P = 0.01$, F -test).

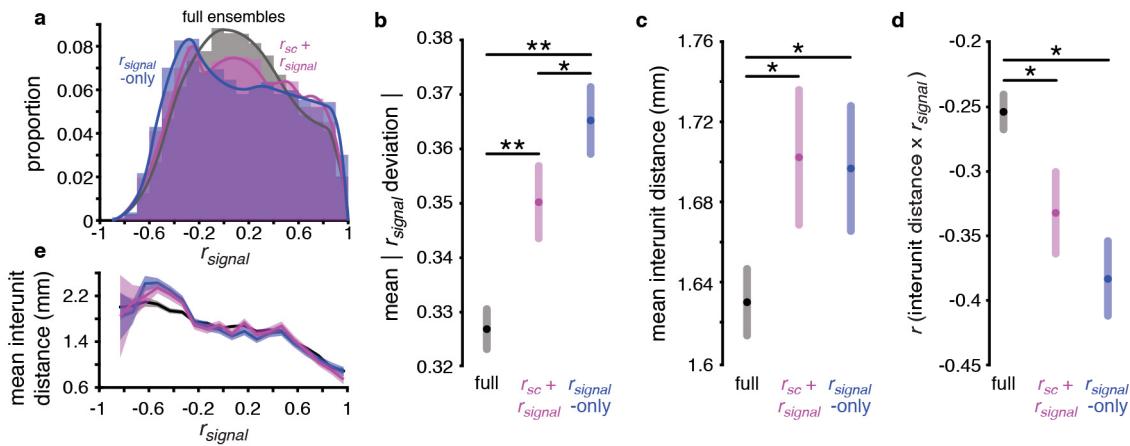


Figure 4.5: Ensembles optimized for WM representation are r_{signal} diverse and anatomically dispersed. (a) r_{signal} distributions for the full ensembles (grey; $n = 12,222$ units), near-max $r_{signal} + r_{sc}$ ensembles (violet; $n = 2,414$), and near-max r_{signal} -only ensembles (blue; $n = 2,724$), pooled across all sessions. All three distributions are significantly different from each other ($P << 0.001$, χ^2 -test, Bonferroni-corrected; computed using non-overlapping bins of size = 0.1). (b) Mean $|r_{signal}$ deviation | in the full (grey), near-max $r_{signal} + r_{sc}$ (violet), and near-max r_{signal} -only ensembles (blue). r_{signal} deviation is defined as the difference between a unit pair's r_{signal} and the mean r_{signal} of the ensemble to which the unit pair belongs. ** $P << 0.001$, Bonferroni-corrected, * $P = 0.01$, F -test—Online Methods. Shaded regions represent Bonferroni-corrected 95% comparison intervals between group means—Online Methods. (c) Mean interunit distance in each of the three ensemble groups. * $P < 0.005$, F -test, Bonferroni-corrected. Shaded regions represent Bonferroni-corrected 95% comparison intervals between group means. (d) Correlation between interunit

distance and r_{signal} in the three ensemble groups. * $P < 0.005$, bootstrap test. Shaded regions represent bootstrapped 95% confidence intervals. (e) Mean interunit distance (y-axis) as a function of r_{signal} in each of the ensemble groups, computed using non-overlapping r_{signal} bins of size 0.1. Shaded region denotes S.E.M.

Prior studies have reported weak topography for visual (Suzuki and Azuma, 1983; Leavitt et al., 2013) and mnemonic (Constantinidis et al., 2001a) space in LPFC; units' tuning similarity and the anatomical distance between them—the “interunit distance”—are negatively correlated. If the optimized ensembles reflect this topography, their broader representation of the stimulus space means they should encompass larger regions of cortex relative to the full recorded ensembles. Indeed, we found that the mean distance between units—or interunit distance—was larger in the near-max $r_{signal} + r_{sc}$ and r_{signal} -only ensembles than the full ensembles (Figure 4.5c $P < 0.005$ for both, *F*-test, Bonferroni-corrected—Materials and Methods). We also found that topography in the optimized ensembles was enhanced compared to the full ensembles (Figure 4.5d); the correlation between interunit distance and r_{signal} was significantly stronger in the near-max $r_{signal} + r_{sc}$ ensembles ($r = -0.33$) and r_{signal} -only ensembles ($r = -0.38$) compared to the full ensembles ($r = -0.26$; $P < 0.005$ for both, bootstrap test). A potential explanation for this difference is that the distance between units with negative r_{signal} is larger in the optimized

ensembles (Figure 4.5e). Remarkably, the mean interunit distance can reach 2.5mm in the near-max ensembles. Considering that cortical columns in LPFC could span ~0.7 millimeters (Bugbee and Goldman-Rakic, 1983), this result suggests that optimal ensembles extend across several cortical columns. These results link the spatial mnemonic topography of LPFC to principles of WM coding. They also demonstrate the utility of accounting for neuronal information content when examining cortical organization, compared to approaches that focus on neuronal tuning characteristics while leaving their effects on information implicit. These findings are also robust to the choice of “near-max” value, as repeating the analyses with different thresholds yielded similar results (Supplementary Figure 4.7).

4.3.6 Non-selective units can improve WM coding by modifying the r_{sc} structure

Given our observation that the r_{sc} structure can significantly affect the information content of a neuronal ensemble during WM, it is possible that neurons that do not contain task-related information in isolation could still influence the information content of an ensemble by modifying the r_{sc} structure (Figure 4.6a). The r_{signal} distribution of the $r_{signal} + r_{sc}$ ensembles in Figure 4.6a contains a peak near $r_{signal} = 0$, unlike the r_{signal} -only ensembles, suggesting that units with orthogonal and/or weak selectivity may contribute more to WM coding when the r_{sc} structure is intact.

Indeed, non-selective units were sometimes added to ensembles before selective units, and before decoding performance saturated (Figure 4.6b). In order to test whether these units were increasing ensemble WM information by modifying the r_{sc} structure, we identified all the non delay-selective units ($P \geq 0.05$, one-way Kruskal-Wallis ANOVA with stimulus location as the factor) that were added prior to decoding performance saturation in the $r_{signal} + r_{sc}$ ensembles (Figure 4.6b; 16 units in total). We then compared the amount of information these units contributed to an ensemble before and after shuffling out the r_{sc} structure (Figure 4.6c—see Materials and Methods). Removing the r_{sc} structure significantly decreased the amount of WM information contributed by these units ($P < 0.01$, signed rank test, paired), and the amount of WM information contributed after shuffling was not significantly different from zero ($P = 0.43$, Wilcoxon signed rank test, unpaired; additional descriptive statistics and control analyses for these units are provided in Supplementary Figure 4.8). We also found 15 non-selective, noise-shaping neurons during the stimulus epoch. Only one of the non-selective, noise-shaping neurons was common to both epochs. However, the decoding improvement contributed by these neurons, both before and after removing the r_{sc} structure, was more consistent during the delay than the stimulus epoch (Supplementary Figure 4.9). These results

demonstrate the existence of non-selective “noise-shaping” neurons: neurons that do not contain task-related information in isolation, but increase the information content of an ensemble entirely through modifying the r_{sc} structure.

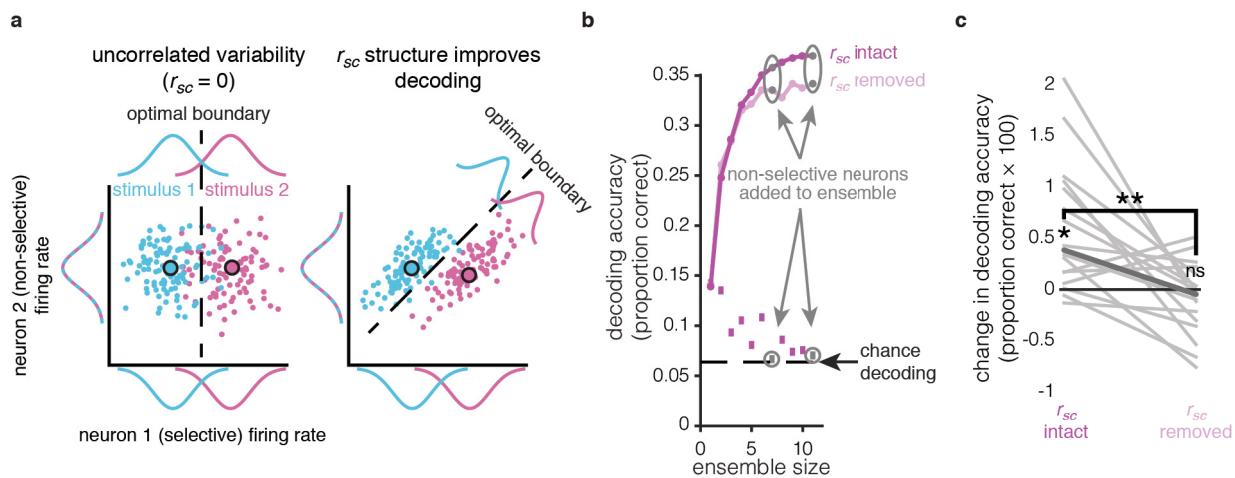


Figure 4.6: Non-selective neurons can increase ensemble information by modifying the r_{sc} structure. (a) Two-neuron conceptual diagram of how a non-selective neuron could increase ensemble information content. In the first scenario (left), one neuron differentiates between two stimuli (i.e. is selective; stimuli are denoted by blue and pink) and the other neuron does not (i.e. is not selective). The response variability of the two neurons is not correlated (i.e. $r_{sc} = 0$). In the second scenario (right), the individual neurons’ properties are identical, yet correlated response variability (i.e. the r_{sc} structure) improves discrimination between the two stimuli relative to the uncorrelated scenario. (b) The continuous line plots with circular markers show the ensemble decoding accuracy (y-axis) as a function of size (x-axis) for the $r_{sc} + r_{signal}$ optimized method for a single example ensemble, prior to decoding saturation, for r_{sc} -intact data (magenta) and r_{sc} -shuffled data (pale magenta). The square markers at the bottom of the plot denote the decoding accuracy (y-axis) of the individual unit added to the ensemble at a given size (x-axis). Notice units that are added to the population that are not selective (gray). (c) Change in decoding accuracy from adding non-selective units to pre-saturation ensembles (y-axis) when the r_{sc} structure is intact (left) and removed (right). Each line is the change for an individual unit. The bolded line is the median. Removing the r_{sc} structure

eliminates the information gain contributed by these units. $*P = 0.001$, signed-rank test; $**P < 0.003$, paired signed-rank test; ns (not significant), $P = 0.43$, signed-rank test; $n = 16$.

4.4 Discussion

By using microelectrode arrays to record from ensembles of LPFC neurons, we were able to elucidate the effects of the r_{sc} structure on WM coding, and more generally, how WM is represented in neuronal ensembles. We found that the relationship between r_{sc} and r_{signal} during WM maintenance was as predicted by connection topography in which similarly-tuned neurons are recurrently excitatory and dissimilarly-tuned neurons are mutually inhibitory. Using a linear decoder, we found that removing the r_{sc} structure could increase or decrease the information content of neuronal ensemble, depending on the size and composition of the ensemble. Consistent with previous findings, WM fidelity in ensembles of randomly-selected neurons was impaired by the r_{sc} structure, and the magnitude of the impairment was proportional to the size of the ensemble. However, the intrinsic r_{sc} structure improved WM coding in smaller neuronal ensembles of neurons optimized for WM representation ($r_{signal} + r_{sc}$ ensembles). The $r_{signal} + r_{sc}$ ensembles consisted of different neurons than ensembles optimized for WM representation in the absence of the r_{sc} structure (r_{signal} -only ensembles). The $r_{signal} + r_{sc}$ ensembles had a broader

r_{signal} distribution, were more anatomically dispersed, and exhibited stronger topography than the full population of recorded LPFC units. Finally, we found that individual units that did not encode WM in isolation (“non-selective” neurons), could still contribute to WM coding when part of an ensemble by altering the ensemble’s r_{sc} structure.

4.4.1 Recurrent network dynamics during WM coding

WM representations in LPFC are hypothesized to be maintained by a network structure of recurrent excitation and lateral inhibition (Amit and Brunel, 1997; Camperi and Wang, 1998; Compte et al., 2000; Durstewitz et al., 2000; Wang, 2001; Constantinidis and Wang, 2004; Wimmer et al., 2014). The resulting dynamics should manifest as changes in r_{sc} during WM maintenance (delay epoch) relative to other epochs. We observed this phenomenon in our data—mean r_{sc} is lower during the stimulus epoch compared to the delay epoch. A previous experiment (Constantinidis and Goldman-Rakic, 2002) reported this trend, but did not find a significant effect, perhaps due to a smaller sample size (295 pairs, compared to our 10,535 pairs). A second prediction is that WM maintenance should modify r_{sc} as a function of r_{signal} ; r_{sc} should be lower between neurons with dissimilar tuning than neurons with similar tuning (Constantinidis et al., 2001a; Constantinidis and

Goldman-Rakic, 2002; Qi and Constantinidis, 2012; Katsuki et al., 2014; Wimmer et al., 2014; Markowitz et al., 2015). Indeed, we found that the relationship between r_{sc} and r_{signal} changed as predicted during the delay compared to the fixation and stimulus epochs (Figure 4.2b). Our findings indicate that WM maintenance modulates the r_{sc} structure of LPFC neuronal ensembles in a manner consistent with recurrent excitation and lateral inhibition.

4.4.2 Decoding WM representations from LPFC neuronal ensembles

A prior study showed that using a pseudopopulation (asynchronously-recorded neurons) of the 8 most informative LPFC neurons to decode spatial WM information during a match/non-match WM task yielded nearly identical results as using the entire 600-neuron pseudopopulation (Meyers et al., 2012). We also showed that a small subensemble of the most informative neurons contain nearly as much WM information as the full recorded population. Importantly, we demonstrated that accounting for the r_{sc} structure increases ensemble efficiency, thus pseudopopulation analyses likely overestimate the number of neurons required to achieve a given decoding accuracy.

A second study (Markowitz et al., 2011) using simultaneous recordings from 32 electrodes was also able to decode which of 8 locations was being remembered

during a spatial WM task. However, their study was primarily concerned with how cortical depth affected the ability to decode a remembered location from local field potentials (LFPs), and contained minimal analysis of spiking activity or the impact of neuronal ensemble composition on WM coding.

4.4.3 Effects of r_{sc} and r_{signal} structures on WM coding

The observed patterns of r_{sc} and r_{signal} are thought to be indicative of a network structure that stabilizes WM representations over time (Laing et al., 2002; Polk et al., 2012; Wimmer et al., 2014). Our results demonstrate that these correlations can also affect the readout of WM representations from neuronal ensembles: If WM representations are read out from optimized ensembles, then the network correlation structure will favor WM coding; however, if WM representations are read out from ensembles that are “suboptimal”, then the correlation structure could impair WM coding. Our experiment shows that these changes in ensemble information content can reach 20%. Thus a mechanism that is thought to temporally stabilize WM representations can also affect the ability to read out these representations.

4.4.4 Effects of the r_{sc} structure on information in non-WM tasks

Reports of the effects of ensemble r_{sc} structure on information content vary significantly in sign and magnitude (Zohary et al., 1994; Abbott and Dayan, 1999; Averbeck et al., 2006; Shamir and Sompolinsky, 2006; Graf et al., 2011; Tremblay et al., 2014; Graf and Andersen, 2015); our results can help reconcile these disparate accounts. For example, previous studies that applied decoding techniques to simultaneously-recorded ensembles found that removing the r_{sc} structure decreased decoding accuracy for grating orientation (Graf et al., 2011) and remembered location (Graf and Andersen, 2015), while another study reported a positive effect of pairwise r_{sc} on information coding³². Moreover, spatial attention increases signal-to-noise primarily by reducing r_{sc} (Cohen and Maunsell, 2009; Mitchell et al., 2009; Tremblay et al., 2014). We found that the effect of r_{sc} structure on ensemble information varied dramatically depending on an ensemble's size and composition; removing the r_{sc} structure from the full recorded ensembles increased decoding accuracy, but removing the r_{sc} structure from the most informative sub-ensembles decreased decoding accuracy. The discrepancies across previous studies may arise from the location in “configuration space” of the neuronal ensembles under investigation (Hu et al., 2014). Importantly, they should caution us against making

broad conclusions concerning how variables such as r_{sc} shape information transmission across brain areas. In order to fully clarify this issue one must identify which neurons are contributing to coding, which poses a significant technical challenge.

Our ensemble construction procedures were designed in part to sidestep the challenge of identifying which neurons contribute to coding, and allow us to characterize the system at specific states of interest. One may argue that we did not examine the full ensemble configuration space. Such an undertaking would be computationally infeasible; there are approximately 10^{15} unique ensembles that could be created from 50 neurons. Thus, our results may actually underestimate the true range of effect sizes. Nevertheless, even a limited search of the full configuration space demonstrates the importance of the r_{sc} structure to WM coding in LPFC neuronal ensembles.

4.4.5 Generalization to larger ensembles

We demonstrated that decoding performance of an ensemble can be improved by adding neurons that are not the most informative in isolation. However, this effect seems to be maximized at ensemble sizes of approximately 3-30 neurons (e.g. Figure 4.3d). The origin of this effect seems to be that it benefits decoding to maximize the

width of an ensemble's r_{signal} distribution (e.g. Figure 4.5); if there is a stimulus that few neurons in an ensemble are selective for, it can be more beneficial to add a neuron that is weakly selective for the underrepresented stimulus than to add a neuron that is strongly selective for a stimulus that is already well-represented by the ensemble. The question remains whether this is simply an artifact of the sample size available for analysis, and whether the effect would extend to larger ensemble sizes. Given that this effect arises from the selectivity statistics of the population of recorded neurons, we expect it to scale to larger ensembles so long as the statistics of the population remain similar, with the following caveats: First, this effect will depend on the size of the stimulus space (i.e. the number of unique stimuli that need to be decoded). As the size of the stimulus space increases, so does the need to represent it. In the extreme example of two stimuli to decode, this effect should not exist. Second, as ensemble decoding approaches saturation (i.e. 100%), the decoding improvement from adding neurons to the ensemble will become negligible, rendering arbitrary the choice of which neuron is “optimal” to add to the ensemble. A generalized solution for determining the ensemble size at which decoding saturates is beyond the scope of this study, however it depends on many variables,

including the size of the stimulus space, the selectivity statistics of the neurons in the ensemble, and the number of trials in the data set.

4.4.6 Effects of spike sorting

As in prior similar studies of the effects of r_{sc} on neuronal information coding, we included both single- and multiunits for analysis (Cohen and Maunsell, 2009; Cohen and Kohn, 2011; Ruff and Cohen, 2014; Tremblay et al., 2014). This provides the dual advantages of greater statistical power, and a larger range of ensemble sizes across which to examine the effects of the r_{sc} structure. One caveat of this approach is that measurements of r_{sc} are known to be smaller for single units than for multi-unit clusters, which constitute the majority of our data set. However, our observed mean r_{sc} values are similar to reports in prior studies of WM in LPFC that exclusively examined single units, in which mean r_{sc} ranged from approximately 0.02-0.06, depending on the task epoch (Constantinidis and Goldman-Rakic, 2002; Qi and Constantinidis, 2012). Our observed r_{sc} values are also similar to two prior studies conducted in our laboratory that used identical subjects, microelectrode arrays, and microelectrode array implantation sites as the present study (Leavitt et al., 2013; Tremblay et al., 2014). One of these studies (Tremblay et al., 2014) found that the effects of r_{sc} on decoding spatial attentional information were robust to

changes in spike sorting; the r_{sc} structure impaired SVM decoding of attended location whether or not the activity on an electrode was sorted into separate single- and multiunit clusters, or remained a unitary channel of threshold crossings. Because we replicated previous findings, and observed ranges of values r_{sc} in accordance with similar previous studies, we doubt that our results can be ascribed entirely to the inclusion of multiunit clusters in our data set.

4.4.7 *Noise shaping units*

Neurophysiological studies typically assume that if an ensemble codes for some behavior, the individual neurons constituting that ensemble will also code for that behavior when examined in isolation. This assumption is implicit in the method that forms the bedrock of neurophysiological research: serial recording of individual neurons. However, this approach cannot account for simultaneity between neurons. The use of large-scale simultaneous ensemble recordings allowed us to find non-selective “noise-shaping” neurons: neurons that are not selective for a remembered location, but can improve the fidelity of WM representation in an ensemble by modifying the ensemble’s r_{sc} structure (Figure 4.6a). A similar phenomenon was shown in a prior fMRI study; voxels that do not contain stimulus information in isolation can improve decoding when part of an “ensemble” of voxels (Yamashita et

al., 2008). Their study and ours appear to report two different instances of the same general property of information coding in multidimensional systems: features (e.g. voxels or neurons) that do not contain information in isolation can still modify the amount of information in a system to which they belong by changing the structure of correlated variability. It remains to be observed whether non-selective noise-shaping neurons contribute to information coding in other tasks and brain regions.

4.5 Conclusion

We leveraged the simultaneous multi-neuron recording capabilities of microelectrode arrays to elucidate how WM is coded in LPFC neuronal ensembles. We found that the structure of the correlated variability (r_{sc}) supports current computational models of how sustained activity emerges in WM networks. A great deal of the power of modeling studies lies in their ability to explore parameter spaces, and we devised our ensemble construction procedures in an attempt to create an empirical analog of this capability. Applying these procedures revealed that the size, r_{signal} structure, and r_{sc} structure of an ensemble can profoundly impact WM coding. We also found that LPFC neuronal ensembles that optimize the coding of remembered locations are heterogeneously tuned and anatomically dispersed. Finally, we demonstrated that a ubiquitous assumption in neurophysiological

studies—only “selective” neurons contribute to information coding—is not justified in LPFC networks; non-selective neurons can contribute to information coding by shaping the r_{sc} structure. More generally, our results emphasize the relevance of ensemble-level phenomena in building a comprehensive understanding of brain networks.

4.6 Materials and methods

4.6.1 Ethics Statement

The animal care and ethics are identical to those in (Leavitt et al., 2013), and were in agreement with Canadian rules and regulations and were pre-approved by the McGill University Animal Care Committee. Animals were pair-housed in enclosures according to Canadian Council for Animal Care guidelines. Interactive environmental stimuli were provided for enrichment. During experimental days, water was restricted to a minimum of 35ml/kg/day, which they could earn through successful performance of the task. Water intake was supplemented to reach this quantity if it was not achieved during the task and water restriction was lifted during non-experimental days. The animals were also provided fresh fruits and vegetables daily. Body weight, water intake, as well as mental and physical hygiene were monitored daily. Blood cell count, hematocrit, hemoglobin, and kidney function

were tested quarterly. If animals exhibited discomfort or illness, the experiment was stopped and resumed only after successful treatment and recovery. All surgical procedures were performed under general anesthesia. None of the animals were sacrificed for the purpose of this experiment.

4.6.2 Task

Trials were separated into four epochs: fixation, stimulus presentation (stimulus), delay, and response (Figure 4.1a). The animal initiated a trial by maintaining gaze on a central fixation spot (0.08 degrees²) and pressing a lever; the subject needed to maintain fixation within 1.4° of the spot until cued to respond. The fixation period lasted either 483, 636, or 789ms, determined randomly at the beginning of each trial. After fixation, a sine-wave grating (2.5 Hz/deg, 1° diameter, vertical orientation) appeared at one of 16 randomly selected locations for 505ms. The potential stimulus locations were arranged in a 4x4 grid, spaced 4.7° apart, centered around the fixation point. The stimulus period was followed by a randomly variable delay period of 496-1500ms. The delay period ended and the response period commenced when the fixation point was extinguished, cuing the animal to make a saccade to the location of the previously presented stimulus and then to release the lever. The animal had 650ms to respond. Successful completion of the

trial yielded a juice reward. The minimum duration between trials was 300ms. Fixation breaks during the trial or failure to saccade to the target in the allotted time resulted in immediate trial abortion without reward and a delay of 3.5 seconds before the next trial could be initiated.

4.6.3 Experimental Setup

The experimental setup is identical to (Leavitt et al., 2013; Tremblay et al., 2014). The stimuli were back-projected onto a screen located 1 meter from the subjects' eyes using a DLP video projector (NEC WT610, 1024x768 pixel resolution, 85 Hz refresh rate). Subjects performed the experiment in an isolated room with no illumination other than the projector. Eye positions were monitored using an infrared optical eye-tracker (EyeLink 1000, SR Research, Ontario, Canada). A custom computer program controlled stimulus presentation and reward dispensation, and recorded eye position signals and behavioral responses. Subjects performed the experiment while seated in a standard primate chair, and were delivered reward via a tube attached to the chair and an electronic reward dispenser (Crist Instruments, TX, USA) that interfaced with the computer. Prior to the experiments, subjects were implanted with head posts. The head post(s)

interfaced with a head holder to fix the monkeys' heads to the chair during experiment sessions.

4.6.4 Microelectrode Array (MEA) Implant

As in (Leavitt et al., 2013; Tremblay et al., 2014) we chronically implanted a 10x10, 1.5mm microelectrode array (Blackrock Microsystems LLC, Utah, USA; (Maynard et al., 1997; Normann et al., 1999)) in each monkey's left LPFC—anterior to the knee of the arcuate sulcus and caudal to the posterior end of the principal sulcus (area 8a) (Figure 4.1a). Detailed surgical procedures can be found in (Leavitt et al., 2013).

4.6.5 Recordings and Spike Detection

Data were recorded using a 'Cerebus Neuronal Signal Processor' (Blackrock Microsystems LLC, Utah, USA) via a Cereport adapter. Spike waveforms were detected online by thresholding. The extracted spikes (48 samples at 30 kHz) were re-sorted manually in 'OfflineSorter' (Plexon Inc, TX). The electrodes on each MEA were separated by at least 0.4 mm and were organized into three blocks of 32 electrodes. We collected data from one block during each recording session. Detailed recording procedures can be found in (Leavitt et al., 2013). We collected spike data across 12 recording sessions (7 in JL, 5 in F), yielding a total of 545 units: 164 single

neurons (99 in JL, 65 in F), and 381 multiunits (221 in JL, 160 in F). Multiunits were defined as threshold-crossing events with action potential-like morphology that were not similar enough to be included with any of the well-defined single units. Units with mean firing rates of less than 0.5 Hz during the stimulus or delay epoch and units that fired in fewer than 5% of trials were excluded from analysis.

4.6.6 Analysis Epochs

We analyzed the final 483ms of the fixation epoch, the initial 496ms of the stimulus epoch, and the initial 496ms of the delay epoch. These durations were selected to make the analysis epochs as similar as possible, given the constraints that the stimulus presentation software operated at a resolution of 85 Hz, and to include as many trials as possible. We only analyzed successfully completed trials. Data analysis was performed using MATLAB.

4.6.7 Spatial Selectivity

In order to determine whether a unit was selective for the stimulus location during a given epoch, we computed a one-way Kruskal-Wallis ANOVA on epoch-averaged firing rates with stimulus location as the independent variable. A unit was defined as selective if the test resulted in $P < 0.05$. The number of trials per location varied across recording sessions (mean = 21.5, minimum = 8, maximum = 36). Thus the

total number of trials across all locations (n trials x 16 locations) ranged from 128 to 576. 417 (76%) of the 545 recorded units exhibited delay epoch selectivity.

4.6.8 Spike Count Correlation (r_{sc}) and Signal Correlation (r_{signal}) Analysis

In order to compute r_{sc} , we first z-scored each unit's spike counts for each condition (i.e. stimulus location) in each epoch. Z-scoring separately for each condition removes the spike rate variability across conditions due simply to variability in firing rate responses to different stimuli (i.e. stimulus selectivity); z-scoring also removes differences in baseline firing rates for different neurons. We then grouped units into simultaneously recorded pairs ($n=12,006$) and computed Pearson's correlation coefficients (r_{sc}) between the z-scored spike counts (Cohen and Kohn, 2011; Leavitt et al., 2013). We minimized the risk of falsely inflating the correlation values by excluding correlations between units on the same electrode from analysis.

r_{sc} can covary with firing rate (la Rocha et al., 2007; Cohen and Kohn, 2011), so in order to ensure that differences in r_{sc} across epochs were not confounded by differences in firing rates, we implemented a distribution matching procedure as in (Churchland et al., 2010; Ruff and Cohen, 2014). To create matched distributions, we first computed the distribution of geometric mean firing rates for every pair of neurons included in the r_{sc} analysis, for each epoch. Next, we computed the greatest

common distribution present across all epochs. The distributions in each epoch were matched to the common distribution by randomly discarding data points from each bin of an epoch's distribution, until each epoch's bins contained the same number of data points as those of the common distribution. This reduced the number of correlation pairs from 12,006 to 10,535. This distribution-matching procedure was repeated 2000 times. The mean of these 2000 distributions is plotted in Figure 4.2a. We used a bootstrap test to determine whether the mean pairwise r_{sc} during a given epoch was different from zero and/or different from other epochs. We first computed the mean within each of the 2000 firing rate-matched r_{sc} distributions, yielding 2000 estimates of the mean. If the 0.1th percentile of this distribution of 2000 mean r_{sc} values was greater than zero, we deemed it significantly greater than zero. If the central 99.9% of the distributions of mean r_{sc} values for two epochs did not overlap, we deemed them significantly different.

Signal correlation (r_{signal}) was computed as the correlation of two neurons' mean responses to each of the 16 stimuli. To determine whether tuning similarity affects r_{sc} , we performed a bootstrap test similar to the one described above. For each of the 2000 firing rate-matched r_{sc} distributions, we subdivided the r_{sc} values based on their corresponding r_{signal} value. Neuron pairs with $r_{signal} > 0.25$ were

categorized as “similarly-tuned” and pairs with $r_{signal} < -0.25$ as “dissimililarly-tuned”. We then computed the median r_{sc} within each of the 2000 groups of similarly- and dissimililarly-tuned neuron pairs, yielding six distributions of median r_{sc} values (three epochs \times two tuning similarity groups). If the central 99.9% of two distributions of median r_{sc} values did not overlap, we deemed them significantly different.

4.6.9 Population Decoding

We used a support vector machine (SVM; Libsvm 3.14), a linear classifier, to extract task-related activity from the population-level representations of simultaneously-recorded neuronal ensembles (Cortes and Vapnik, 1995; Chang and Lin, 2011). The SVM used epoch-averaged firing rate data from an ensemble to predict at which of the 16 locations the stimulus appeared in a given trial during each the stimulus and delay epochs. Each neuron constitutes a dimension in the multidimensional population space, and the SVM seeks to find the boundaries that best distinguish between population responses to each stimulus location. Units that fired in fewer than 5% of trials, or fired at a mean rate of less than 0.5Hz were excluded from analysis. We scaled each unit’s firing rates to [-1, 1] by subtracting the midrange rate value $(\text{max}+\text{min})/2$ and dividing by one-half the range $(\text{max}-\text{min})/2$, in order to

prevent units with larger absolute changes in firing rate from dominating the classification boundaries. These two parameters were determined from the training set and applied to both the training and testing sets. We assessed the classifier's performance using five-fold cross-validation, such that 80% of the trials were used to train the decoder (the "training set"), and the decoder attempts to classify the remaining 20% of the trials (the "testing set").

In order to test whether the r_{sc} structure affects the fidelity of ensemble representations during WM, we removed the r_{sc} structure from the neural data using a shuffling procedure identical to that described in (Tremblay et al., 2014). The shuffling procedure consisted of randomizing the trial order within each location condition for each neuron, such that the condition (i.e. the remembered location) for a given trial remained the same for all neurons, but the firing rates for each neuron were drawn from different trials. This procedure destroys the simultaneity, and thus the intrinsic r_{sc} structure, in the recordings. The decoding analysis was then run on the shuffled firing rates. The shuffled decoding procedure was repeated 200 times, and the mean of these 200 iterations was taken as the " r_{sc} -shuffled" decoding accuracy. We refer to the percent change in decoding accuracy due to shuffling as $\Delta_{shuffle}$, defined as $((accuracy_{rsc\text{-}shuffled}/accuracy_{rsc\text{-}intact})-1)\times100\%$.

We define “coding efficiency” as the amount of WM information in an ensemble of a given size relative to another ensemble of the same size, computed as $((accuracy_{ensemble1}/accuracy_{ensemble2})-1) \times 100\%.$

In order to ensure that our results were robust to the choice of classifier, we repeated our decoding analyses using logistic regression instead of SVM. The data analysis procedure for the logistic regression was identical to that of SVM (i.e. excluding low firing rate units, scaling firing rates, and cross-validation procedure), except we used the LIBLINEAR library (Fan et al., 2008) to perform logistic regression instead of SVM. As in the SVM, each unit is a predictor, there are no interaction terms, and the neuronal firing rates are used to predict the remembered location. In the context of our analysis, the relevant differences between SVM and logistic regression are that they have different loss (sometimes called error) functions, and that logistic regression is probabilistic while SVM is deterministic (see (Fan et al., 2008) for more detail).

4.6.10 Functional anatomy

To compare the widths of the r_{signal} distributions of the near-max $r_{sc}+r_{signal}$ ensembles, near-max r_{signal} -only ensembles, and full recorded ensembles, we first computed the absolute value of the r_{signal} deviation ($|r_{signal}$ deviation|) of each unit pair in an

ensemble. The r_{signal} deviation is defined as the difference between a unit pair's r_{signal} and the mean r_{signal} of the ensemble to which the unit pair belongs. We then assessed the difference between the $|r_{signal}$ deviations| in each type of ensemble ($r_{sc+r_{signal}}$, r_{signal} -only, and full recorded ensemble) by fitting a linear mixed-effects model with ensemble type as a fixed effect and recording session as a random effect to predict $|r_{signal}$ deviation|. Pairwise significance between ensemble types was determined by a Bonferroni-corrected *F*-test on the difference between the two groups' coefficients, the degrees of freedom approximated using the Satterthwaite equation. Distance between units (interunit distance) in each ensemble type was compared using a similar linear mixed-effects model, but to predict interunit distance. The measures of variability displayed in Figure 4.6b&c are Bonferroni-corrected simultaneous comparison intervals, generated using equation 3.32 in (Hochberg and Tamhane, 1987).

The strength of topography was assessed by computing the Pearson correlation between every simultaneously-recorded unit pair's r_{signal} and interunit distance within each group of ensembles ($r_{sc+r_{signal}}$, r_{signal} -only, and full recorded ensemble). Significance was assessed using a bootstrap test: the distributions of r_{signal} and interunit distance were randomly sampled, with replacement, to generate

new vectors of length equal to the original distributions, and the Pearson correlation computed between the new resampled vectors. This procedure was repeated 10,000 times to generate a distribution of correlation coefficients. The strength of topography for two ensemble groups was significantly different at $P < \alpha$ if the central proportion of size $1 - \alpha$ of the two groups' bootstrapped correlation coefficients did not overlap.

4.6.11 Non-selective noise-shaping neuron analysis

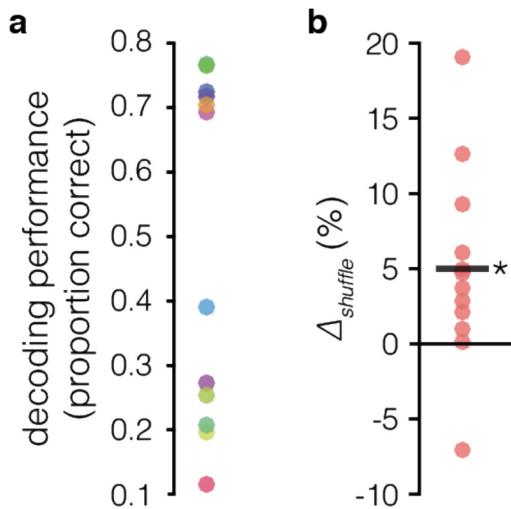
In order to find non-selective noise-shaping neurons, we examined the results of the $r_{signal} + r_{sc}$ optimized procedure and identified all instances in which adding a non-delay-selective neuron ($P \geq 0.05$, one-way Kruskal-Wallis ANOVA, firing rate \times location) increased the information content of the ensemble. We then used the near-max $r_{signal} + r_{sc}$ ensembles to decode firing rate data from which the r_{sc} structure had been shuffled out, and compared the amount of information these units contributed to an ensemble before and after shuffling. Note that using the near-max $r_{signal} + r_{sc}$ ensembles to decode firing data from which the r_{sc} structure has been shuffled out is not the same as the r_{signal} -only optimized procedure. While both involve decoding r_{sc} -shuffled firing rates, in the former case the ensembles are generated using r_{sc} -intact data, whereas in the latter case the ensembles are generated using r_{sc} -shuffled data.

We performed a control analysis to determine whether WM-selective units that contribute similar amounts of information to an ensemble as non-selective noise-shaping units also contribute information by modifying the r_{sc} structure (Supplementary Figure 4.8). We accomplished this analysis by using a distribution matching procedure similar to the one used to match distributions of firing rates in the pairwise r_{sc} analysis. First, we computed the distributions of improvements in decoding accuracy from non-selective noise-shaping units and selective units. Next, we randomly discarded data points from each bin of the selective units' distribution until it matched the non-selective noise-shaping units' distribution. The same data points were discarded from the distribution of selective units' decoding accuracy on the r_{sc} -shuffled firing rates. This procedure was repeated 2000 times. We then computed the median within each of these 2000 matched distributions for r_{sc} -intact data and the r_{sc} -shuffled data. If the central 95% of the two distributions of median improvement in decoding accuracy did not overlap, we deemed them significantly different. We also performed an additional control analysis using a standard ANOVA to assess selectivity instead of a non-parametric Kruskal-Wallis ANOVA.

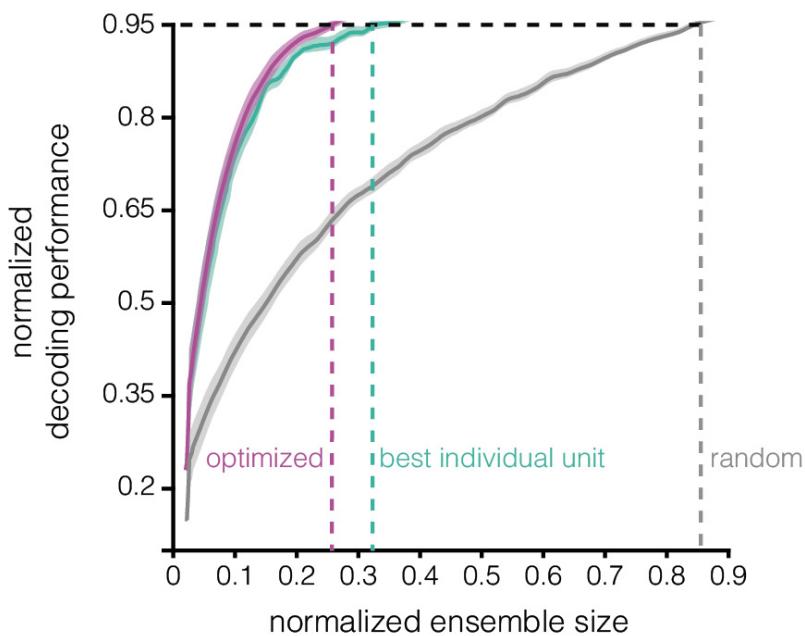
4.7 Acknowledgements

We thank Walter Kucharski for his fabrication expertise, Stephen Nuara for his zoological talents, Megan Schneiderman for assistance with the recordings, Rishi Rajalingham for scrutiny and code, Dr. Ruben Moreno-Bote, Ramon Nogueira, Roberto Gulli, and Lyndon Duong for helpful discussion, and the MT Lab for their support and feedback.

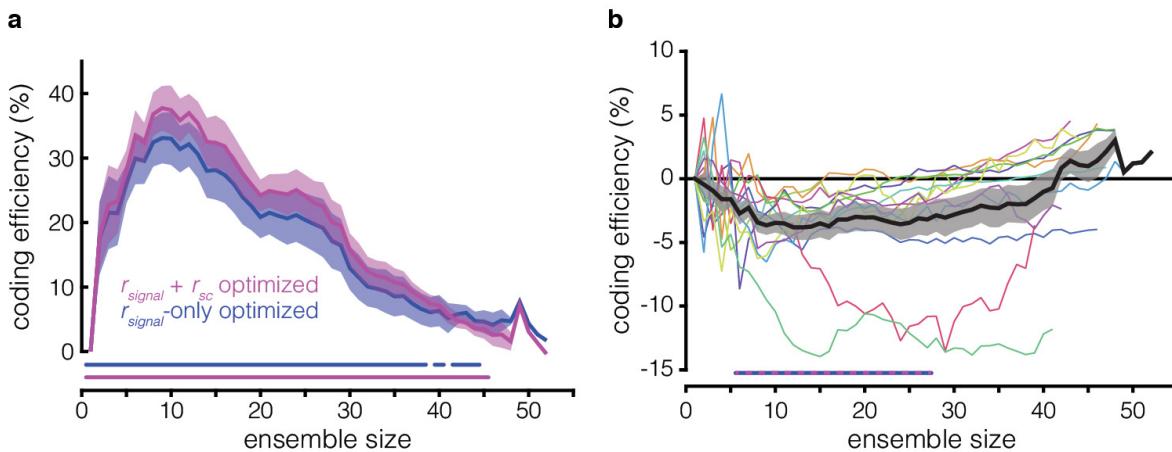
4.8 Supplemental Data



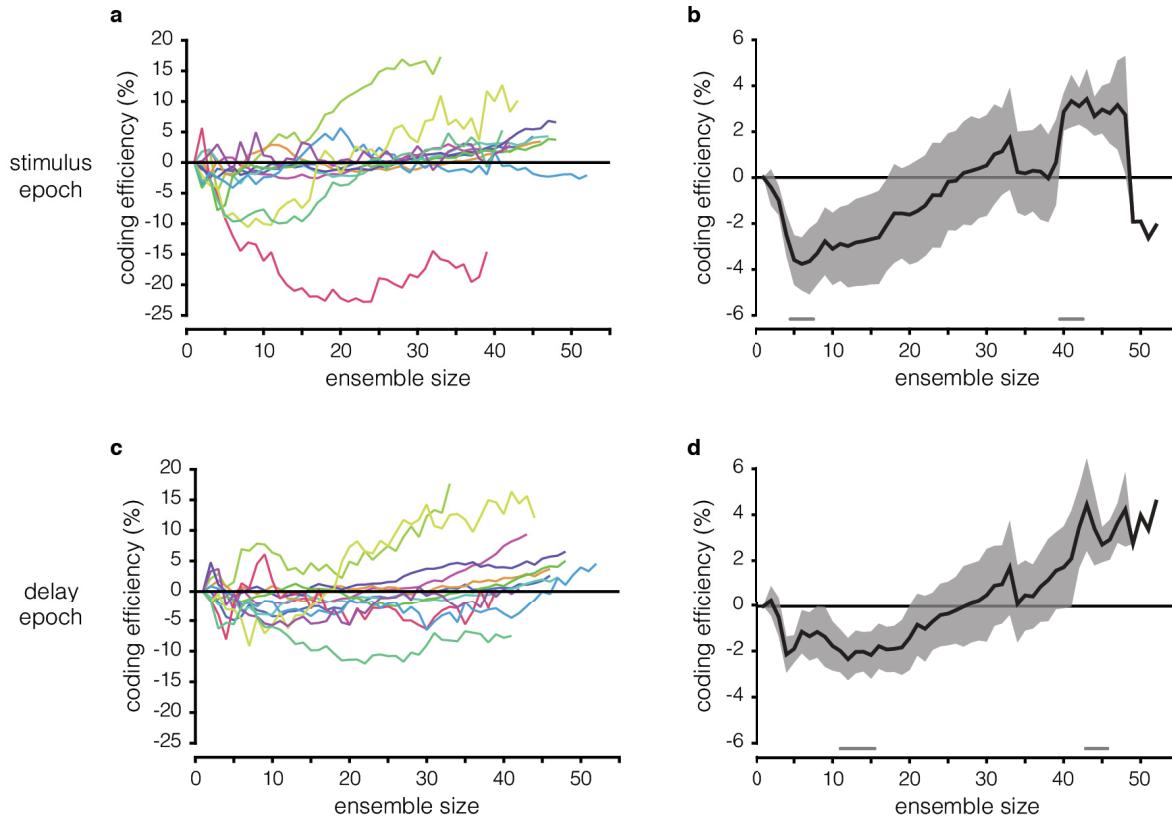
Supplementary Figure 4.1: Decoding in full ensembles. (a) Decoding performance for full ensembles in each session. (b) Effects of removing the r_{sc} structure ($\Delta_{shuffle}$) during the delay epoch for each session. Removing the r_{sc} shuffling has a net effect of improving decoding accuracy in the full ensembles (i.e. ensembles including all simultaneously-recorded neurons). The black line denotes the across-session mean.
 $*P = 0.0032$, paired t-test. $\Delta_{shuffle} = ((accuracy_{rsc-shuffled} / accuracy_{rsc-intact}) - 1) \times 100$.



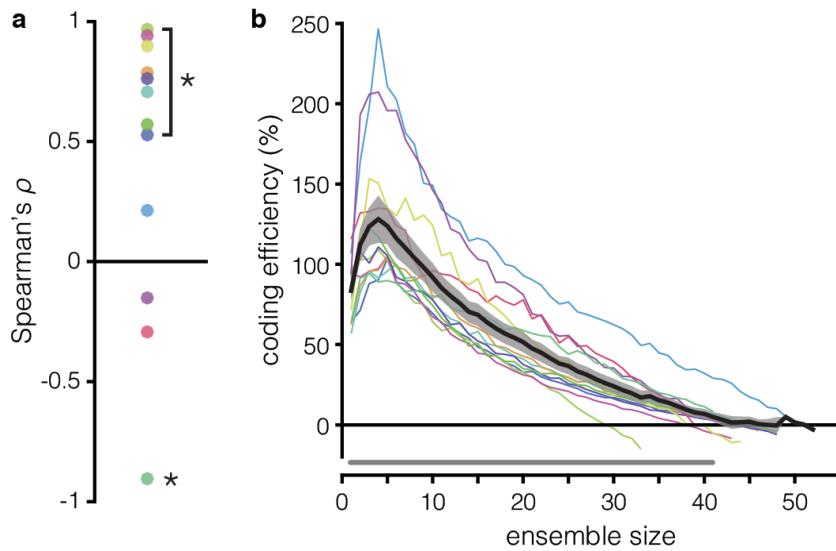
Supplementary Figure 4.2: Decoding saturation curves for the best individual unit vs. best subensemble methods. Normalized decoding accuracy (y-axis) as a function of normalized ensemble size (x-axis) is plotted for the best individual unit (teal), optimized (violet), and random (grey) ensembles. The normalized ensemble size at which normalized decoding accuracy of .95 is achieved is shown for the three procedures.



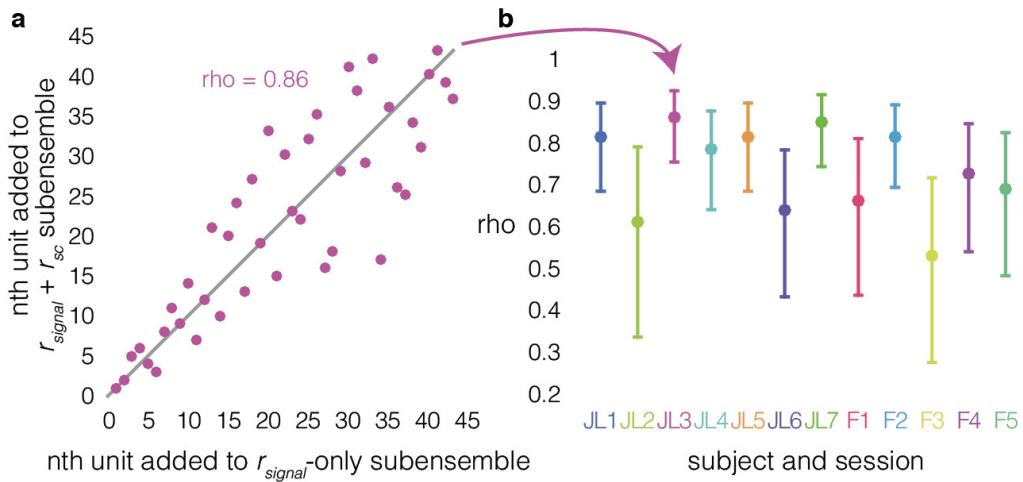
Supplementary Figure 4.3: Effects of r_{sc} structure on ensemble coding efficiency when using logistic regression. (a) Similar to Figure 4.4b, but using logistic regression instead of SVM. Coding efficiency of $r_{signal} + r_{sc}$ ensembles and r_{signal} -only ensembles, relative to the best individual unit ensembles (y-axis), as a function of ensemble size (x-axis). The violet line running along the bottom indicates ensemble sizes for which the $r_{signal} + r_{sc}$ ensembles are significantly more efficient than the best individual unit ensembles ($P < 0.05$, paired t-test, Hochberg-corrected); the blue line is similar, but for r_{signal} -only ensembles vs. best individual unit ensembles. (b) Coding efficiency of r_{signal} -only ensembles relative to $r_{signal} + r_{sc}$ ensembles; similar to Figure 4.4c, but using logistic regression instead of SVM. A positive value indicates that shuffling out the r_{sc} structure improves decoding. The striped blue and violet line running along the bottom indicates ensemble sizes for which the efficiency of $r_{signal} + r_{sc}$ ensembles and r_{signal} -only ensembles are significantly different ($P < 0.05$, paired t-test, Hochberg-corrected).



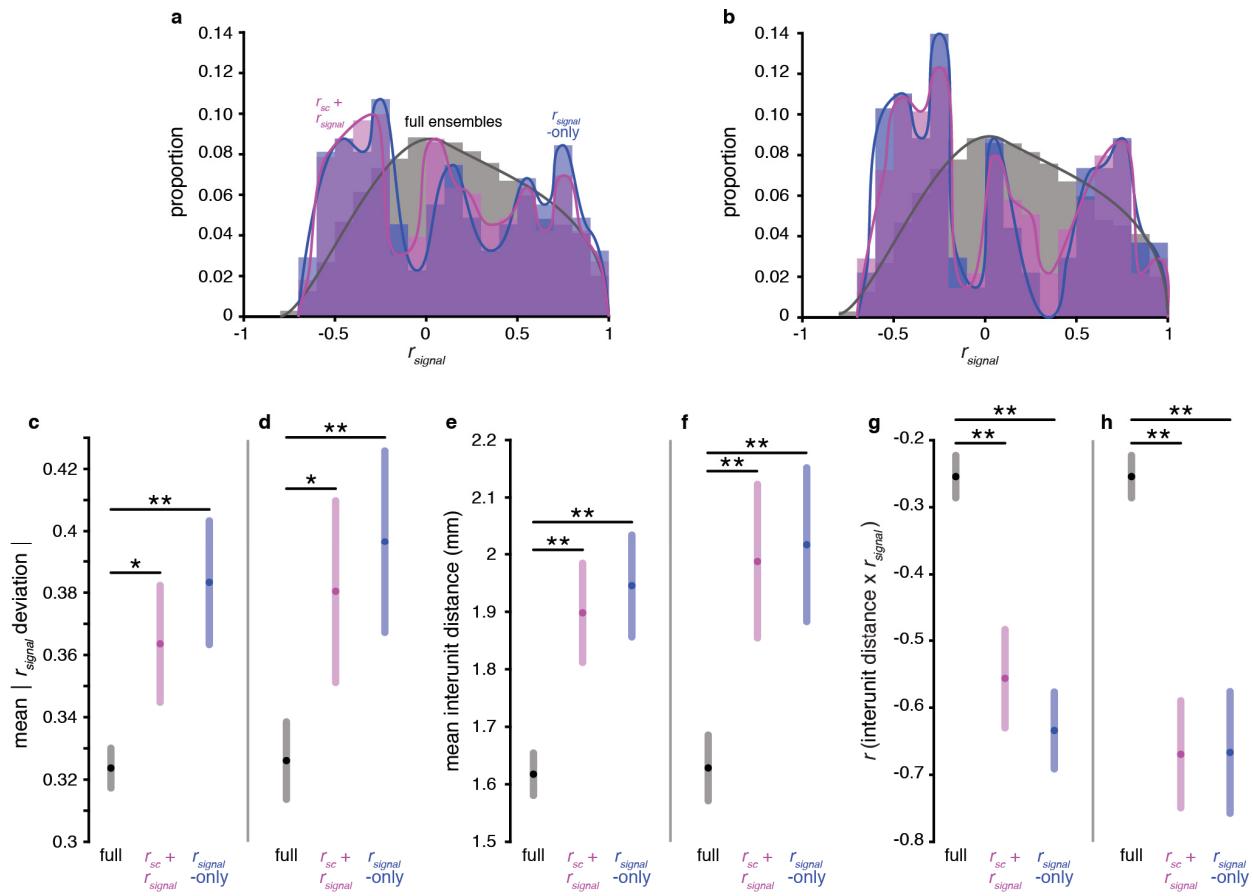
Supplementary Figure 4.4: Effects of the r_{sc} structure during the stimulus vs. delay epochs. (a) Similar to Figure 4.4c, but during the stimulus epoch. Coding efficiency of r_{signal} -only ensembles relative to $r_{signal} + r_{sc}$ ensembles for individual sessions. A positive value indicates that shuffling out the r_{sc} structure improves decoding. Each colored line is an individual session. (b) Across-session mean coding efficiency of r_{signal} -only ensembles relative to $r_{signal} + r_{sc}$ ensembles during the stimulus epoch. The gray lines running along the bottom indicate ensemble sizes for which the efficiency of $r_{signal} + r_{sc}$ ensembles and r_{signal} -only ensembles are significantly different ($P < 0.05$, paired t-test, Hochberg-corrected). (c) Identical data as Figure 4.4c, presented again here for comparison with stimulus epoch data. Coding efficiency of r_{signal} -only ensembles relative to $r_{signal} + r_{sc}$ ensembles for individual sessions. (d) Identical data as Figure 4.4c, presented again here for comparison with stimulus epoch data. Across-session mean coding efficiency of r_{signal} -only ensembles relative to $r_{signal} + r_{sc}$ ensembles.



Supplementary Figure 4.5: $\Delta_{shuffle}$ and coding efficiency in random vs. optimized ensembles. (a) $\Delta_{shuffle}$ and ensemble size are correlated in random ensembles. The Spearman rank correlation coefficient (ρ) between ensemble size and $\Delta_{shuffle}$ in random ensembles, for each of the 12 recording sessions. A positive correlation indicates that shuffling out the r_{sc} structure increases decoding accuracy more in larger ensembles. * $P < 0.01$, Spearman correlation. (b) $r_{signal} + r_{sc}$ ensembles are more efficient than random ensembles at nearly all ensemble sizes. Coding efficiency of the $r_{signal} + r_{sc}$ ensembles relative to random ensembles (y-axis) as a function of ensemble size (x-axis). A positive value indicates that the $r_{signal} + r_{sc}$ ensembles are more efficient for a given ensemble size. The gray line running along the bottom indicates ensemble sizes for which the $r_{signal} + r_{sc}$ ensembles are significantly more efficient than the random ensembles ($P < 0.05$, paired t-test, Hochberg-corrected).

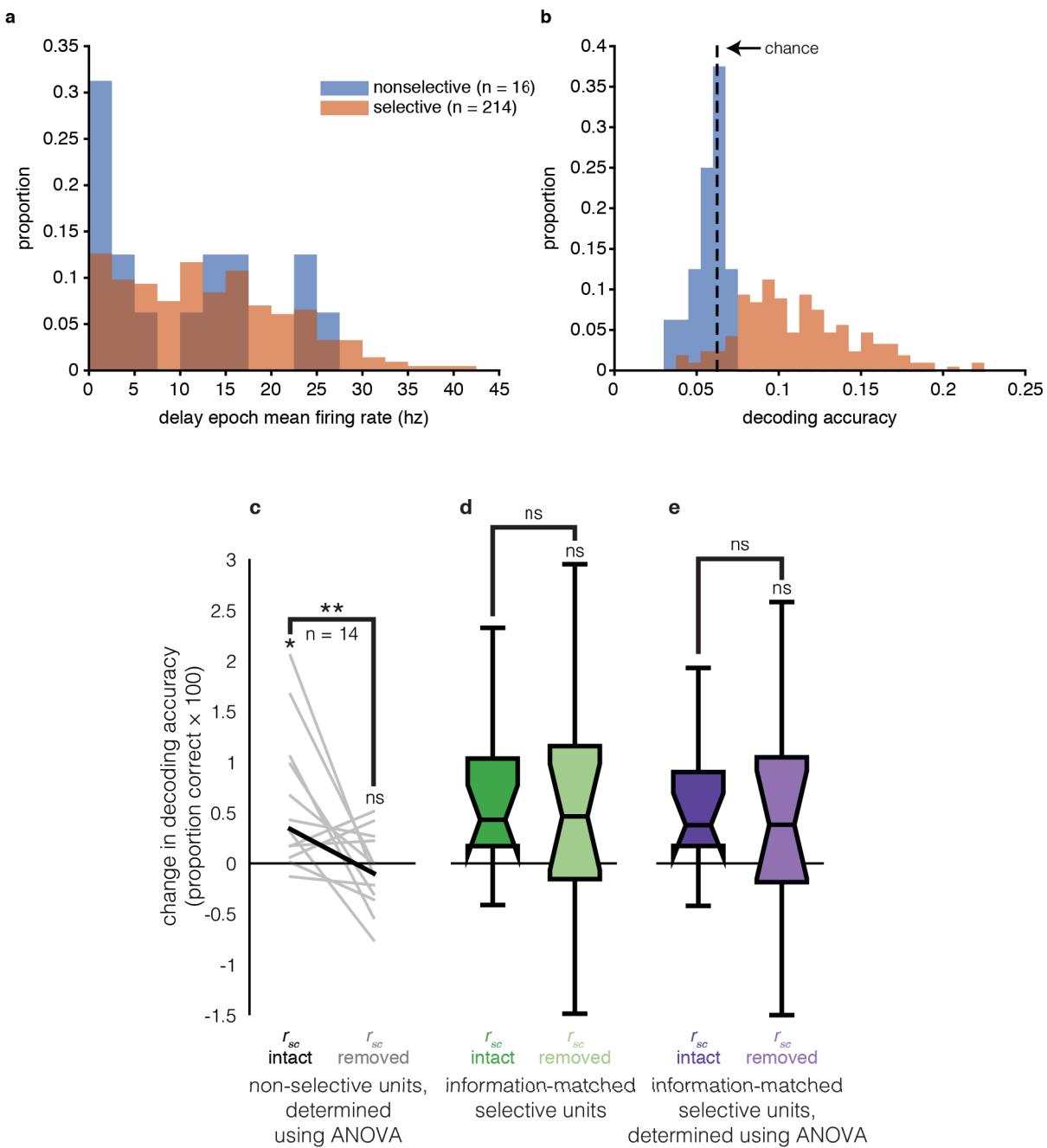


Supplementary Figure 4.6: Ensemble construction sequence correlation shows that different ensembles maximize WM information when the r_{sc} structure is intact vs. removed. (a) Sequence order correlation between $r_{\text{signal}} + r_{\text{sc}}$ ensembles (y-axis) and r_{signal} -only ensembles (x-axis) for an example session. The n^{th} unit added to the ensemble in the r_{signal} -only method (x-axis) is plotted against the n at which it was added to the ensemble in the $r_{\text{signal}} + r_{\text{sc}}$ method. For example, the 5th unit added in the r_{signal} -only method is added 4th in the $r_{\text{signal}} + r_{\text{sc}}$ method. If the sequence in which the two methods added units was identical, all the points would fall along the unity line (gray), and the Spearman rank correlation coefficient (ρ) between the two sequences would equal 1, indicating that identical ensembles maximize WM information regardless of whether the r_{sc} structure is intact. Likewise, $\rho = 0$ means that the relationship between the sequences is random; the ensembles that maximize WM information when the r_{sc} structure is intact are entirely distinct from the ensembles that maximize WM information when the r_{sc} structure has been removed. (b) Spearman rank correlation coefficient (ρ) between the $r_{\text{signal}} + r_{\text{sc}}$ ensembles and r_{signal} -only ensembles (y-axis) for each session (x-axis). Error bars are Bonferroni-corrected 95% confidence intervals. The correlation between ensemble sequences is significantly less than 1 in every session ($P < 0.05$, Bonferroni-corrected), indicating that the ensembles that maximize WM information when the r_{sc} structure is intact are built in a different sequence than the ensembles that maximize WM information when the r_{sc} structure has been removed. The range of ρ values (0.53-0.86) shows that there is some degree of similarity to the sequences in which the two methods recruit neurons to the ensembles.



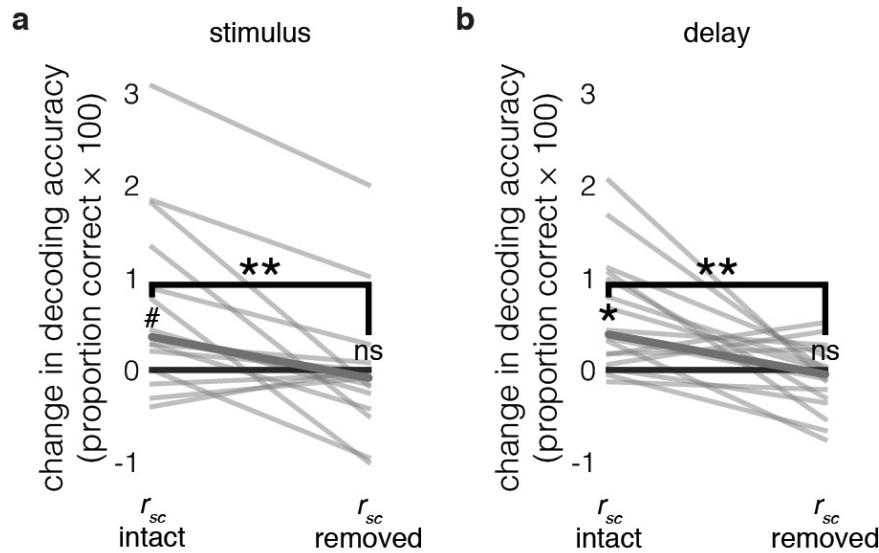
Supplementary Figure 4.7: Functional anatomy analyses are consistent even when using different decoding saturation thresholds. (a) r_{signal} distributions for the full ensembles (grey), near-max $r_{signal} + r_{sc}$ ensembles (violet), and near-max r_{signal} -only ensembles (blue), pooled across all sessions. Identical to Figure 4.5a, but using a threshold of 90% of maximum decoding. The $r_{signal} + r_{sc}$ ensemble and r_{signal} -only ensemble distributions are both significantly different from the full ensemble distributions ($P < 0.001$, χ^2 -test, Bonferroni-corrected; computed using non-overlapping bins of size = 0.1). (b) Identical to panel (a), but using a threshold of 80% of maximum decoding. All three distributions are significantly different from each other ($P < 0.001$, χ^2 -test, Bonferroni-corrected; computed using non-overlapping bins of size = 0.1). (c) Mean $|r_{signal}$ deviation | in the three categories of ensembles. Identical to Figure 4.5b, but using a threshold of 90% of maximum decoding. r_{signal} deviation is defined as the difference between a unit pair's r_{signal} and the mean r_{signal} of the ensemble to which the unit pair belongs. $**P <<0.001$; $*P = 0.01$, F -test—Materials and Methods. Shaded regions represent Bonferroni-

corrected 95% comparison intervals between group means—Materials and Methods. (d) Identical to panel (c), but using a threshold of 80% of maximum decoding. (e) Mean interunit distance in each of the three ensemble groups. Identical to Figure 4.5c, but using a threshold of 90% of maximum decoding. ** $P << 0.001$, F -test. Shaded regions represent Bonferroni-corrected 95% comparison intervals between group means. (f) Identical to panel (e), but using a threshold of 80% of maximum decoding. (g) Correlation between interunit distance and r_{signal} in the three ensemble groups. Identical to Figure 4.5d, but using a threshold of 90% of maximum decoding. ** $P << 0.001$, bootstrap test. Shaded regions represent bootstrapped 95% confidence intervals. (h) Identical to panel (g), but using a threshold of 80% of maximum decoding.



Supplementary Figure 4.8: Descriptive statistics and control analyses for non-selective “noise-shaping” units. (a) Histogram of delay epoch firing rates of non-selective ($P \geq 0.05$, Kruskal-Wallis ANOVA; blue) and selective ($P < 0.05$, Kruskal-Wallis ANOVA; orange) units in best ensembles. Firing rate bins are plotted on the x-axis, and bin counts are plotted on the y-axis. (b) Histogram of delay epoch

decoding accuracies of non-selective and selective units in best ensembles. Decoding accuracy bins are plotted on the x-axis, and bin counts are plotted on the y-axis. (c) Parametric selectivity control. We reran the analysis in Figure 4.6, defining selectivity as $P < 0.05$, ANOVA, instead of Kruskal-Wallis ANOVA. The change in decoding accuracy from adding non-selective units to best ensembles (y-axis) is compared when the r_{sc} structure is intact (left) and removed (right). Each line is the change for an individual unit. The bolded line is the median. Similar to when using a non-parametric measure of WM selectivity, removing the r_{sc} structure eliminates the information gain contributed by these units. * $P = 0.001$, signed-rank test; ** $P = 0.01$, paired signed-rank test; ns $P = 0.58$, signed-rank test; $n = 13$. (d) Selective unit control. We used a distribution-matching procedure (see methods) to obtain a distribution of WM-selective units ($P < 0.05$, Kruskal-Wallis ANOVA) that contribute equivalent amounts of information to an ensemble as the non-selective units and performed a similar analysis as in Figure 4.6. The center of each box is the median, and the notches extend from the median $\pm 1.57(q_3-q_1)/\sqrt{n}$, where q_3 is the 75th percentile, q_1 is the 25th percentile, and $n = 13$, the sample size of a single matched distribution. The bottom box edge = q_1 , top edge = q_3 , and the whiskers extend approximately 99.3% distribution coverage. Removing the r_{sc} structure does not change the magnitude of information added to the ensemble by these units ($P > 0.05$, bootstrap test—see methods), thus it is not simply the case that units that weakly improve ensemble information do so by modifying the r_{sc} structure. ns (not significant), $P > 0.05$, bootstrap test. (e) Selective unit + parametric selectivity control. Same as (b), but defining selectivity as $P < 0.05$, ANOVA, instead of Kruskal-Wallis ANOVA. The results are similar for the two methods. ns $P > 0.05$, bootstrap test.



Supplementary Figure 4.9: Comparison of untuned noise-shaping neurons during stimulus and delay epochs. (a) Similar to Figure 4.6c, but during the stimulus epoch. Change in decoding accuracy from adding non-selective units to pre-saturation ensembles (y-axis) when the r_{sc} structure is intact (left) and removed (right). Each line is the change for an individual unit. The bolded line is the median. $\#P = 0.013$, signed-rank test; $**P < 0.003$, paired signed-rank test; ns (not significant), $P > 0.05$, signed-rank test; $n = 15$. (b) Contribution of untuned, noise-shaping units during the delay epoch. Identical to Figure 4.6c, presented again here for comparison with stimulus epoch data. $*P = 0.001$, signed-rank test, $**P < 0.003$, paired signed-rank test; ns $P > 0.05$, signed-rank test; $n = 16$.

CHAPTER 5

GENERAL DISCUSSION

5.1 Summary

In Experiment 1 we determined the tuning properties of area 8a neuron during a simple visually-guided saccade task. We then examined the relationship between neurons' tuning and r_{sc} during fixation, prior to the presentation of visual stimuli. We found that r_{sc} decreased as a function of distance between neurons, positive r_{sc} were stronger between similarly-tuned neurons, and negative r_{sc} were stronger between dissimilarly-tuned neurons. Most importantly, we found that r_{sc} between anatomically distant ($>3\text{mm}$) neurons with dissimilar tuning were predominantly negative, suggestive of tonic resting-state inhibition. This pattern of intrinsic functional connectivity supports the model of a recurrent-excitatory lateral-inhibitory network structure in dlPFC. This experiment addressed Aim 2, determining the magnitude and spatial scale of functional interactions in dlPFC in the absence of stimulus input.

Experiment 2 was designed to address three aims: Aim 1, to determine how ensembles of simultaneously-recorded neurons in dlPFC area 8a represent SWM, and how it differs from single neuron representations; Aim 3, quantifying whether functional (i.e. task-related) properties of neurons in dlPFC exhibit topography; and Aim 4, to examine whether neuronal tuning for visual space during WM is biased by

the meridians of the visual field. In this experiment the animals performed an oculomotor delayed-response (ODR) task. We found that neuronal SWM representations were asymmetrically biased not to extend beyond the meridians of the visual field induce field. This bias is present at multiple levels of examination: in single neuron firing rates, the r_{sc} structure, and neuronal ensemble representations. The subjects' saccades also exhibited similar biases to those reported in the psychophysical literature. These results could be considered neuronal correlates of known biases in WM-based judgments of visual space. We also found that dlPFC neurons are anatomically clustered in a non-retinotopic manner that partially reflects the organization of visual space.

The final experiment addressed Aim 1, determining how ensembles of simultaneously-recorded neurons in dlPFC area 8a represent SWM and how it differs from single neuron representations, and Aim 3, quantifying whether functional (i.e. task-related) properties of neurons in dlPFC exhibit topography. This experiment utilized the same ODR task as Experiment 2. We found that WM maintenance modulated the r_{sc} structure in manner indicative of increased coupling between similarly-tuned neurons and increased inhibition between dissimilarly-tuned neurons. We then examined the effects of the r_{sc} structure on WM coding, and

found that it could facilitate or impair the readout of WM representations, depending on the size of the ensemble and tuning properties of its constituent neurons. Interestingly, ensembles of neurons that contained the most robust SWM representations were more anatomically dispersed than predicted by the statistics of the full recorded population of neurons. Finally, we found neurons that contained little WM information when examined in isolation, but contributed to coding when part of an ensemble by shaping the ensemble's r_{sc} structure. This final result is a powerful example of a phenomenon that is inaccessible in individually-recorded neurons, and emphasizes the importance of multi-neuron recordings for elucidating the neuronal mechanisms of cognition.

5.2 Reconciling the relationship between tuning and r_{sc} across experiments

All three experiments found a similar relationship between pairs of neurons' tuning and r_{sc} during fixation; r_{sc} positively correlates with tuning similarity. Experiments 2 and 3 each extend this analysis to the stimulus and delay epochs, but the results in each experiment appear to contradict one another (Figure 3.5, Figure 4.2b). Experiment 3 found that the positive correlation between r_{sc} and tuning observed during fixation increased in strength during the stimulus and

delay epochs (Figure 4.2b). Experiment 2, however, found that this effect appears to invert during the stimulus epoch; r_{sc} between similarly-tuned neurons are lower than between dissimilarly-tuned neurons (Figure 3.5). Experiment 2 also found that during the delay epoch, the effect is non-linearly meridian-dependent (Figure 3.5). The apparent discrepancies between Experiments 2 and 3 can be reconciled by the difference in how the two experiments quantify tuning similarity. Experiment 3 quantified tuning similarity as the signal correlation (r_{signal}) between two neurons. In contrast, Experiment 2 first fit a 2d polynomial to each neuron's responses to the 16 stimulus locations, then identified the peak (i.e. maximum) of the fitted function (Figure 3.4), and assigned each pair of neurons to a group depending on the spatial relationship of the quadrants containing the two neurons' fitted response peaks (Figure 3.5, top). Furthermore, neurons in dlPFC—and across multiple regions of PFC—are known to have non-traditional tuning functions unlike neurons in sensory cortex, which are typically Gaussian or parametric (Rainer et al., 1998; Constantinidis et al., 2002; Rigotti et al., 2013; Fusi et al., 2016). The non-traditional tuning and multidimensional selectivity of PFC neurons suggests that standard measures of tuning similarity may be overly simplistic for characterizing the functional properties of PFC (Fusi et al., 2016).

5.3 Stability of prefrontal WM representations

Goldman-Rakic's domain-specific model of PFC WM function, at its most extreme, proposes that individual subregions of lateral PFC intrinsically encode WM for individual features or sensory modalities. As discussed in the introduction of this thesis, such a claim is no longer tenable. There arguably exists an equally-extreme inverse to the domain-specific model. It is referred to as the "adaptive coding" model of PFC function, and it posits that PFC neurons have no intrinsic selectivity for specific features, and instead their selectivity adapts to represent inputs according to their task-relevance (Duncan, 2001). While there is no doubt that selectivity in the PFC is remarkably plastic and often represents supersensory and/or abstract features, especially those defined by behavioral relevance (Miller and Cohen, 2001), the wealth of domain-specific deficits observed in focal PFC lesions undermine the extreme interpretation of this model (Petrides, 2000a; 2005a; 2005b). Furthermore, a series of experiments examining lateral PFC activity in monkeys before and after training on WM tasks demonstrated the existence of sustained, spatially-selective responses after stimulus offset prior to training, and changes in the proportion of tuned neurons, the tuning specificity of single neurons, and the patterns of cross-correlations and r_{sc} after training, and found that many of these effects exhibited

differential anatomical specificity (see (Constantinidis and Klingberg, 2016) for a review).

Some of the results presented in this thesis potentially provide a counterpoint to recent findings interpreted to support the adaptive coding model. In an experiment conducted by Stokes and colleagues, monkeys first learned a set of arbitrary associations between cue and target stimuli (e.g. if ‘house’ is the cue, ‘banana’ is the target; if ‘car’ is the cue, ‘tree’ is the target) (Stokes et al., 2013). The researchers then recorded single cell activity in areas 9/46 and 8 while monkeys performed a task in which a cue object was initially presented, then a series of non-matching targets were presented during which the animal had to withhold a response, and finally respond via saccade when the matching target was presented. They then performed a cross-temporal decoding analysis on the neuronal pseudopopulation, which consisted of training a classifier on the pseudopopulation firing rates integrated over a single 50ms window at a given point in the task, and testing it on every other 50ms window across the remainder of the task. They found that the decoding performance dropped precipitously as the training window moved further in time away from the testing window. This effect was particularly pronounced when training within the cue and testing within the delay epoch. The

authors interpreted these results to indicate that the sustained dlPFC activity observed during WM is not a temporally-consistent representation of remembered stimuli or upcoming response features, but instead a context-dependent dynamic code that reflects the current task demands, in this case whether each sequentially-presented target stimulus should elicit a response.

In Experiments 2 and 3 of this thesis, I presented data showing it is possible to decode which of 16 stimulus locations was being remembered using ensemble firing rates integrated over timescales ranging from 345-1349ms. Such robust decoding of high-resolution SWM information using firing rates integrated over large and variable timescales necessitates temporally-stable dimension(s) in the ensemble code. This appears to conflict with the findings of Stokes and colleagues. One could attempt to account for this discrepancy by appealing to the lack of simultaneity in their neuronal recordings, but the magnitudes of their reported changes in decoding performance exceed the range of effects of the r_{sc} structure reported in Experiment 3. This would also require one to contrive a model with temporal and task-variable modulation of the r_{sc} structure. Alternatively, the ensemble dimensionality is high enough that strong temporal dynamics and stable stimulus coding could co-exist in separate subspaces. Indeed, Murray et al. found exactly this when analyzing

pseudopopulation data from an ODR and a somatosensory delayed-match-to-sample task (Murray et al., 2017). Through clever use of a traditional dimensionality-reduction technique (PCA), they showed that heterogeneous temporal dynamics in single-neuron and ensemble activity could exist alongside temporally stable population coding of the remembered stimulus. These findings provide yet another example of the utility of ensemble-level approaches for elucidating the mechanisms of cognition and neural coding.

It remains possible that there is additional task-relevant information encoded in the temporal dynamics of dlPFC WM activity that is not present in the stable coding subspace. One way to test this would be to compare the performance of a rate classifier to that of a temporal pattern classifier. In a rate classification scheme, which is what was used in Experiments 2 and 3 of this thesis, the inputs to the classifier are simply the firing rates of each neuron integrated over some time window. In temporal pattern classification, the spiking activity over the same time window as used in the rate classifier is subdivided into smaller time windows, the spike rate integrated over these smaller time windows, and then all the smaller time windows are provided to the classifier simultaneously. As an example, a rate classifier would use the firing rates of 20 neurons integrated across a 500ms time

window, while the input to the temporal pattern classifier would consist of spike rates integrated over every non-overlapping 50ms time window contained in the same 500ms time window used for the rate classifier. Thus the rate classifier would have n_{neuron} inputs (20 in the present example), while the temporal pattern classifier would have $n_{neuron} \times n_{subwindows}$ inputs (200 in the present example) which also contain the temporal dynamics. Nandy et al. applied this methodology to V4 neurons recorded during passive receptive field mapping using stimuli with strong spatiotemporal dynamics (Nandy et al., 2016). They found that the temporal pattern classifier significantly outperformed the rate classifier, demonstrating that there is considerable stimulus information encoded in the temporal dynamics of V4 neurons. Applying this approach to dlPFC recordings such as those collected for thesis could clarify the role of temporal dynamics in dlPFC representations and hone the specificities of the dynamic coding model of PFC.

5.4 Layer-specific functional properties of dlPFC

dlPFC, like cortex in general, is known to have different input sources and output destinations, and contain different types and proportions of neurons (Kritzer and Goldman-Rakic, 1995; Petrides and Pandya, 1999). Thus depth information is extremely important for advancing our knowledge of the mechanisms of working

memory and cortical circuits in general. One shortcoming of the microelectrode arrays we used for our recordings is that they do not provide any control over or explicit information about penetration depth.

The electrodes on the arrays we used were 1.5mm long, however the penetration depth was likely shallower for multiple reasons. The first reason originates from our surgical procedures for the array implantation. The array is implanted into the cortical surface by means of a pneumatic impactor (not unlike a captive bolt gun), which can be adjusted to obtain different electrode penetration depths. We set the impactor to a 1mm penetration depth despite our electrodes being 1.5mm, as we found during prior implantation surgeries that this approach mitigated local cerebral trauma around the implantation site. The second factor affecting the electrode penetration depth is that the electrodes are not resting against the first cortical layer before implantation; pia matter and blood vessels are present between the electrode tips and the cortex. Furthermore, the brain is not mechanically static. Hemodynamic pulsations can cause variability in the array implantation depth depending on when in the pulse cycle the array is implanted. Finally, the array encompasses a large enough area of 8a (4mmx4mm) that the underlying cortex potentially curves away from the array by a non-trivial amount.

Accounting for the factors affecting the electrode penetration depth our arrays, we estimate that our electrodes were no more than 1mm deep. Thus we were likely recording from the lower layer III (i.e. layer IIIc) or layer IV (Kritzer and Goldman-Rakic, 1995; Petrides and Pandya, 1999). Layer IIIc in this area has been observed to have distant and extensive lateral projections (Kritzer and Goldman-Rakic, 1995). Interestingly, the targets of these projections are not homogeneous, and instead exhibit distinct spatial periodicity. It is possible that the strength of functional interactions between neurons in this layer (e.g. r_{sc}) exhibit corresponding spatial periodicity, a hypothesis that could potentially be addressed with the data collected for this thesis.

Research on the layer-specific and network-level functional properties of neurons in dlPFC is unfortunately limited, due both to the difficulty of precise depth measurements and novelty of large-scale recording technologies. However, a handful of studies auger advancement in this domain. One study recorded from chronically-implanted depth-adjustable microelectrode arrays implanted across areas 8a and 9/46 while macaques performed an ODR task (Markowitz et al., 2015). The researchers segregated their recorded neurons based on whether they exhibited “storage”- or “response”-related activity, and found that “storage” neurons were

located primarily within the top 1mm of 8a, while “response” neurons were predominantly found at depths of 1-2mm in area 9/46. These results provide electrophysiological corroboration of anatomical studies linking superficial lateral projections to maintenance or storage functions of WM, and descending projections from deeper PFC layers to subcortical oculomotor structures which are thought to be involved in response generation (Kritzer and Goldman-Rakic, 1995). Another series of studies (see (Opris and Casanova, 2014) for a review) used arrays of multi-contact laminar probes to record from layers 2/3 and 5 of areas 9/46d, 8ad, and 6 during delayed spatial and object match-to-sample tasks. They found segregation of task-specific responses across cortical layers, as well as layer- and microcolumn-dependent differences in cross-correlation strength and latency between neurons. They also found that applying microstimulation using a pattern similar to that which was recorded during correct behavioral performance improves the animals’ performance. Extending these techniques to simultaneous recordings of dlPFC and its projection targets such as parietal and extrastriate cortex could elucidate the micro- and macro-network mechanisms that give rise to WM.

5.5 FEF, 8a, and prefrontal cortex

The frontal eye fields (FEF), located in the arcuate sulcus, are commonly described as part of PFC, and area 8a as an extension of FEF (Ferrera et al., 1999; Sommer and Wurtz, 2001; Gregoriou et al., 2009; Schafer and Moore, 2011; Squire et al., 2013; Bichot et al., 2015). Such descriptions overlook the anatomical and physiological differences between 8a and FEF; FEF bears more similarity to premotor cortex, while 8a is much more like other regions in PFC. While only 10–20% of neurons in FEF exhibit ipsilateral selectivity (Bruce and Goldberg, 1985; Bruce et al., 1985; Sommer and Wurtz, 2000), a larger proportion of neurons in area 8a—up to 40%—have ipsilateral stimulus selectivity (Lennert and Martinez-Trujillo, 2011). Area 8a also has a well-defined granular layer while FEF is distinctly dysgranular (Petrides, 2005a). Furthermore, low-current microstimulation (<50mA) does not evoke saccades in area 8a, while it does in FEF (Bruce et al., 1985; Schall et al., 1995; Petrides, 2005a). Finally, 8a exhibits selectivity for non-spatial object features (Miller and Cohen, 2001; Hussar and Pasternak, 2010; Mendoza-Halliday et al., 2014), while such selectivity has not been found in FEF (Ferrera et al., 1999). Thus, the properties of area 8a neurons during visual and WM tasks cannot be equated to the properties of FEF neurons.

FEF primarily has receptive fields in the contralateral visual field, even though it is considered to be downstream from areas MST and IT, both of which encode substantial regions of the ipsilateral visual field (Desimone et al., 1985; Raiguel et al., 1997). It is possible that FEF lies closer to the “motor” end of the visuomotor transformation process, after receptive fields have reverted to a more retinotopic organization and recovered their contralateral representational bias. This could be a fundamental feature of information transmission to motor structures, as movement coding is usually restricted to effectors in the contralateral side of the body. Given that the superior colliculus, a structure known to lie near the distal end of the visuomotor transformation, also exhibits a hemifield bias similar to that found in FEF (Goldberg and Wurtz, 1972a; 1972b; Wurtz and Goldberg, 1972), this explanation appears parsimonious.

However, it is unclear whether area 8a and FEF are serially connected in the stream of visuomotor processing. An alternative possibility is that association areas such as 8a utilize connections with other motor areas such as FEF to monitor and/or select between representations of sensory information and motor plans during complex tasks (Petrides, 2005a; 2005b). Supporting this hypothesis, patients with prefrontal lesions do not show sensory or motor deficits in simple visuomotor tasks

such as visually guided saccades, while they are strongly impaired when performing tasks that require contextual flexibility of behavior (Petrides, 1982; 1987; 2005a; 2005b). Thus, within this framework the dlPFC may play a role in tasks in which the transformation of sensory information into motor commands is not direct, but requires flexible behavior (Petrides, 2005a; Fuster, 2008). Anatomical and evolutionary evidence supports this hypothesis, as the dlPFC is one of the regions that has undergone the most significant relative size increases in primates, compared to other animals with less sophisticated behavioral repertoires (Fuster, 2008).

5.6 Sustained activity underlying WM: A ubiquitous cortical phenomenon?

The data gathered for this thesis were recorded exclusively from macaque dlPFC area 8a, falling in with the preponderance of literature focusing on the role of dlPFC in working memory. However, reports of WM-related sustained activity in parietal cortex (Andersen et al., 1985) and the temporal lobe (Fuster and Jervey, 1981) date back to the 1980s. These findings led to the current view that WM-related sustained activity is not exclusive to prefrontal neurons, but exists within a network of association cortices (Goldman-Rakic, 1995). The role of each area in the origin of

this phenomenon and in working memory in general remains unclear. Studies over the last decades have also reported sustained activity across a range of cortical areas (predominantly in extrastriate and higher-level sensory regions—see (Pasternak and Greenlee, 2005) for a review), and subcortical structures (Hikosaka et al., 1989; Watanabe and Funahashi, 2004; Takaura et al., 2011) the thalamus, superior colliculus, and the basal ganglia. These findings are often cited to support the theory that sustained, WM-related activity is a ubiquitous feature of cortical networks (Christophel et al., 2017).

Of particular note are findings of sustained activity in primary visual cortex (V1), reported in at least two studies. The first study used a figure-ground segregation paradigm in which V1 receptive fields were tonically stimulated, providing a potential explanation for the reported elevation in background firing during the memory period (Supèr et al., 2001). The second study using a similar paradigm but controlling for this potential confound reporting that sustained activity is present even in the absence of a stimulus in the receptive field during the first 500 milliseconds of the delay period (van Kerkoerle et al., 2017). However, at least three other studies have reported an absence of sustained activity in V1 when using match-to-sample tasks for orientation (Luck et al., 1997; McAdams and

Maunsell, 1999), and color/shape (Fuster, 1990). Interestingly, all these negative results were reported in publications containing positive results for other cortical areas, raising the possibility that additional negative results may be unreported due to the scientifically-deleterious publication bias against negative results (Dickersin et al., 1987; Easterbrook et al., 1991; Button et al., 2013). The discrepant findings reported in these studies can be accounted for by multiple factors: the studies used different tasks; the studies reporting positive results recorded multi-unit activity, while the studies reporting negative results recorded from isolated single neurons; finally, the “sustained activity” reported in the positive studies decreased substantially after the visual cue offset and decayed to baseline within 500ms. Furthermore, negative reports of WM-related activity in visual area MT (Ferrera et al., 1994; Mendoza-Halliday et al., 2014), and the lack of feature dimension-specificity among mixed reports of WM-related activity in V4 (Luck et al., 1997; McAdams and Maunsell, 1999; Chelazzi et al., 2001; Hayden and Gallant, 2013) call into question the cortical ubiquity of local network structures capable of supporting WM.

5.7 Reconciling fMRI and electrophysiological findings

Findings from fMRI research have been used to support claims that WM-related activity is cortically ubiquitous. Recent studies using multivoxel pattern-classification analysis (MVPA) have been able decode the contents of WM from early sensory cortices including V1, even if the mean signal intensity in these areas remains near baseline levels during memory delays (Harrison and Tong, 2009; Serences et al., 2009; Emrich et al., 2013). Interestingly, these methods often fail to extract feature-specific WM information when applied to PFC, which has been used to argue against a feature-specific role for PFC in WM (Riggall and Postle, 2012). These failures to decode WM feature information from PFC is peculiar considering the decades of electrophysiological studies reporting feature-selective WM responses in PFC neurons. Experiments 2 and 3 of this thesis can be counted among these studies.

Some aspects of Experiments 2 and 3 bear more similarity to fMRI MVPA approaches than to traditional single neuron studies. MVPA is much more similar to the population decoding methods used in Experiments 2 and 3 than to traditional single neuron measures of information content, and we analyzed signals over areas of cortex atypical of single neuron studies (multiple mm²). Yet we were able to

robustly decode which of 16 locations was being held in WM. This discrepancy likely arises from the interaction of two factors: First, an fMRI voxel integrates BOLD responses over 1-27mm³ of cortical volume. The second factor has to do with cortical organization. Sensory cortex has a topographical organization characterized by cortical columns for relevant features (e.g., space, orientation or motion direction) (Hubel et al., 1977), whereas cortical organization in association areas may not follow the same principles, and single neurons or groups of neurons with different tuning may cluster within a space smaller than the voxel resolution. Accordingly, voxels in sensory cortex likely contain neurons with homogeneous or similar tuning, whereas voxels in PFC contain neurons with heterogeneous response properties that are unresolvable with the BOLD signal. Findings from Experiment 2 speak directly to this issue. While we found that spatial tuning for WM clustered anatomically in area 8a, the strength of this clustering decreased monotonically over larger spatial scales and was strongest at scales ≤1.5mm (Figure 3.2), which is at or below the spatial resolution of fMRI. Additionally, while we found some degree of retinotopy for contralateral visual-mnemonic space across the region of cortex covered by the microelectrode arrays, representation of the ipsilateral hemifield was

not organized in this manner, which likely also contributes to the failure of fMRI to decode feature-specific WM information in this region.

5.8 Features of neural circuits allowing sustained activity and WM

Findings reported in this thesis support the theory that the sustained activity underlying working memory maintenance requires cortical circuits with recurrent network dynamics (Wang, 2001; Wimmer et al., 2014; Murray et al., 2017).

However, there is a rich literature on the cellular, anatomical, and neuromodulatory features of WM circuits that are inaccessible using electrophysiological methods. Wang et al. have proposed that a microcircuit composed of pyramidal neurons and three different types of interneurons containing different calcium binding proteins (parvalbumin (PV), calbindin (CB) and calretinin (CR)) can generate sustained activity in the prefrontal cortex. Within such a circuit, activity is sustained after the number of recurrent connections exceed a certain threshold (Wang, 2006). Considering this modeling work, one would expect to find microcircuit differences between regions known to exhibit sustained spiking activity during WM tasks and regions in which such sustained activity is not found, such as early sensory cortices. Such differences have indeed been reported. For example, layer III pyramidal neurons in the prefrontal cortex have 16 times more spines than

in area V1. This is not simply due to larger cell bodies in pyramidal neurons, but to an increase in spine density, indicating that pyramidal neurons in the PFC are more heavily interconnected than area V1 (Elston, 2007).. Moreover, anatomical studies in the PFC of macaques have revealed stripe-like patterns and long lateral connections originating in layer IV that have the potential to connect cortical columns over millimeters (Goldman-Rakic and Schwartz, 1982).

Another important difference between PFC and early sensory areas in which delay activity is not observed, such as V1, is the proportions of different types of interneurons. These cells are thought to play an important role in recurrent dynamics in neural circuits, and therefore in the origin of sustained activity (Wang et al., 2004). In area V1, PV are the most prevalent interneuron type. By contrast in the prefrontal cortex, CR cells are the most prevalent type (Wang, 2013). Interestingly, a study has reported that in human and monkey brain slices of prefrontal and temporal cortex, CR neurons generate sustained spiking activity after stimulation by a single action potential, and that spiking is terminated by bursts of action potentials (Wang et al., 2015). CR interneurons have been shown to selectively drive the activity of CB interneurons that project to dendritic trees of pyramidal cells in the rat hippocampus (Gulyás et al., 1996). Unfortunately many

of these relationships between cell type, network structure, and function are difficult to test electrophysiologically, and the results of such research would ultimately still be correlational. Transgenic techniques may prove extremely powerful in this domain. Optogenetics and DREADDs (designer receptors exclusively activated by designer drugs (Urban and Roth, 2015)) can target specific cell types, affording cell type-specific manipulation for precise deconstruction of neuronal circuits. Optogenetic manipulation of frontal cortex parvalbumin neurons in mice has demonstrated their modulatory effects on signal-to-noise in local circuits (Sohal et al., 2009). Given the challenges of implementing optogenetic techniques in macaques, widespread use of these techniques to investigate the issues posed here are contingent on methodological advancements in macaque optogenetics (Gerits and Vanduffel, 2013). Pursuing these questions in marmosets may be a promising alternative (Kinoshita and Isa, 2015; MacDougall et al., 2016).

The PFC is also a target of afferents from dopaminergic, noradrenergic, serotonergic, and cholinergic systems (Fuster, 2008). Studies in areas 8a and 9/46 have shown that sustained activity and performance on WM tasks are sensitive to dopaminergic modulation (Sawaguchi and Goldman-Rakic, 1991; Williams and Goldman-Rakic, 1995). Computational modeling work has suggested that dopamine

D1 receptor stimulation strengthens recurrent excitation and lateral inhibition during WM, which increases the signal-to-noise of the maintained representation and makes it more robust to interfering signals (Durstewitz and Seamans, 2002). This effect would likely be visible as a more extreme version of the r_{sc} structure modulation observed in Experiments 2 and 3. Unfortunately this has yet to be verified experimentally, highlighting the importance of future studies that combine neurotransmitter and pharmacological manipulations with large-scale neuronal recordings.

5.9 Future directions for WM and cognition research in PFC

The experiments presented in this thesis investigated neuronal activity in area 8a during a WM task. But, as mentioned previously in this thesis, it is likely the case that area 8a, and dlPFC more generally, is not involved in WM *per se*. There are a staggering variety of behaviors and features found to be reliant on or represented in the dlPFC, these include: auditory, visual, and somatosensory object features; top-down attention; behavioral context; reward; stimulus-response associations; and often retrospective and prospective versions of these phenomena as well. Accommodating such an expansive range of capabilities in a single model has been challenging. Miller and Cohen's model of PFC function—that it is “*responsible for*

active maintenance of patterns of activity...that represent goals and the means to achieve them”—is justifiably hailed for clearly and articulately synthesizing so many seemingly disparate observations. Its generality, however, cuts both ways, as it suffers from a lack of testable mechanistic predictions.

Petrides proposes a more exacting model that, while anatomically limited to lateral PFC, provides greater specificity and testable hypotheses (Petrides, 2005a; 2005b). Based primarily on behavioral deficits observed after focal lesions of different macaque LPFC subregions, he proposes a “rostral-caudal axis of cognitive control”. In this model, caudal LPFC (areas 8 and 6DR) are responsible for learning and contextually selecting between different sensorimotor mappings, evidenced by lesion-induced deficits in conditional associative learning and top-down attention tasks (e.g. ‘green square means pick the left object and red square means pick the right object’, or ‘respond to change in the cued object but not the distractors’). What Petrides refers to as the mid-dlPFC (areas 9, 46, and 9/46v) is responsible for *“monitoring of selections from a set of stimuli or the occurrence of stimuli from an expected set”* (Petrides, 2005a). The prototypical example of this behavior involves a task in which monkeys initially select one object from an array of familiar objects, then following a delay are presented with the initially-selected object and another

object, and must select the object they did not initially select. Monkeys with mid-dlPFC lesions exhibit no impairment if the initially-selected object is presented with a novel object during the test phase, but are severely impaired if the initially-selected object is presented with other familiar objects; their recognition memory is intact, but they are unable to “monitor” their prior choices. Finally, the mid-vLPFC (areas 47/12 and 45) is responsible for “*active effort to retrieve specific information guided by the subject’s intentions and plans*” (Petrides, 1996). This concept is operationalized in experiments comparing free recall vs. recognition. Lesions of mid-vLPFC do not impair recognition, nor simple recall such as in a single-location ODR task, but severely impair free recall of larger sets of memoranda or complex patterns. It is noteworthy that the behaviors ascribed to each of these LPFC subregions are fully dissociated: lesions of one subregion do not cause deficits in the tasks reliant on other subregions. While lesions of 8a may not cause deficits on simple WM tasks, the ventral portion of area 8a is reciprocally connected with subcortical structures and regions of the parietal, temporal, and occipital cortices involved in oculomotor and spatial processing, which could explain why neurons in area 8a exhibit selectivity across such a broad range of visuomotor tasks (Petrides, 2005a).

Given the strength of the evidence supporting Petrides' model, why has there been such debate over the role of PFC in WM? One straightforward explanation is differences in experimental design across studies. One study found that PFC lesion-induced WM performance deficits could be rescued by darkening the room in which the experiment was conducted (Malmo, 1942). Another study found that saccades to incorrect target locations in an ODR task caused by area 9/46 GABA_A inactivation were often followed by a second saccade to the correct stimulus location, but that these corrective saccades were not included for analysis in prior publications (Tsujimoto and Postle, 2012).

Another concern is that methodological limitations have left much of the neuroscientific research examining WM as correlational. Neither human functional imaging nor primate electrophysiology can parse the causal relationships between behavioral phenomena and neural structures, this generates data that are unable to confirm that the signals being analyzed are relevant to the behavior of interest. This limited use of causal methods undermines systems and behavioral neuroscience in general, beyond simply studies of WM. The widespread adoption of optogenetics in rodent models provides a promising way forward, but this technology is not yet mature enough for widespread adoption in primate research.

Causal methods are especially critical for understanding the role of the PFC and other neuronal structures underlying cognition.

Traditional electrophysiological research methods may also be mismatched for work on the PFC. These analytical approaches were developed to characterize specific components of the nervous system, involved in sensory coding, which process information that directly maps to well-defined, measurable physical quantities. These methods are still regularly applied when investigating cognitive phenomena, despite the extreme difficulty in quantifying cognitive processes both abstract and nebulous. And, it is possible that the operations performed on cognitive signals are unlike those performed by the sensory cortex on sensory signals. These methods may confound our observations of signals related to myriad task parameters in PFC; this difficulty is not solely caused by the common behavioral demands across varied task designs, but may be amplified by our reliance on inappropriate analytical techniques.

The PFC is essential for behaviors that require complex yet plastic associations across diverse and disparate signals. This unique facility may explain the non-traditional tuning and multidimensional representations that are a hallmark of PFC neurons. When introducing an explanation for “mixed selectivity”

in PFC neurons, Fusi elegantly explains the dilemma posed by traditional approaches:

The traditional view of brain function is that individual neurons and even whole brain areas are akin to gears in a clock. Each is thought to be highly specialized for specific functions. This, however, does not fit with many observations, especially in higher-order cortex. For example, training monkeys on a cognitive-demanding task engages huge proportions of neurons in the prefrontal cortex (~40% of randomly sampled cells). This means that training either hijacks a huge slice of cortical tissue (and monkeys can only learn 2–3 tasks before their brains reach capacity). Or instead that neurons can do more than one thing (Fusi et al., 2016).

Thus traditional electrophysiological approaches may be failing to capture the full functional properties of the PFC, which may be better understood through the application of ensemble-level analyses. This more nuanced approach permits the examination of multidimensional representations and single-trial variability in behavioral and neuronal activity.

The greatest challenge to studying neuronal substrates of cognitive phenomena may be experimenters' difficulty in conceptualizing the phenomena under investigation. Sophisticated experimental techniques are only as useful as our ability to interpret their results. Traditional neurophysiological analyses have likely remained popular in part because they are relatively straightforward,

allowing researchers to readily comprehend and communicate their results. High-dimensional statistics and dynamical systems are not typically a component of neuroscientific curricula, though this may be a detriment to the field. They could provide a conceptual framework and common language for understanding PFC function.

The legacy of cognitive psychology continues to be a mixed blessing for the neuroscientific study of cognition. Cognitive psychology was invaluable in providing the conceptual groundwork and motivation for cognitive neuroscience. However, models created without regard for the underlying biological substrates of behavior will eventually hamstring advancement of the field. How much effort has been expended designing and interpreting experiments to conform to the impressionistic introspections of 19th century aristocrats? Concepts such as “attention” and “working memory” are ultimately both semantic and normative in origin.

Distinguished scientific careers have been forged by identifying and reconciling these semantic and conventional inconsistencies across researchers and fields. Advancing the field of cognitive neuroscience is best achieved by taking a “brain-first” approach: this would primarily rely on conceptualizing behavioral phenomena through the neural structures they require. This will no doubt be

challenging, as the brain is not organized for the express purpose of being understood by its owner. However difficult, this challenge should be viewed as an invitation. Humankind's myriad and astounding capabilities—critiquing a Rembrandt, engineering a skyscraper, passively acquiring language, traveling extraterrestrially, driving an automobile, or even recollecting a fond moment of connection with a loved one—all, ultimately arise from the nearly invisible activity of the brain. Only by accurately capturing this microcosm may we develop a deeper understanding of the physical substrate that underlies our shared humanity.

REFERENCES

- Abbott, L.F., and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11, 91–101.
- Abrams, J., Nizam, A., and Carrasco, M. (2012). Isoeccentric locations are not equivalent: The extent of the vertical meridian asymmetry. *Vision Res.* 52, 70–78.
- Alvarez, G.A., Gill, J., and Cavanagh, P. (2012). Anatomical constraints on attention: Hemifield independence is a signature of multifocal spatial selection. *Journal of Vision* 12, 9–9.
- Amit, D.J., and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex* 7, 237–252.
- Andersen, R.A., Essick, G.K., and Siegel, R.M. (1985). Encoding of spatial location by posterior parietal neurons. *Science* 230, 456–458.
- Atick, J.J. (2011). Could information theory provide an ecological theory of sensory processing? *Network: Computation in Neural Systems* 22, 4–44.
- Averbeck, B.B., and Lee, D. (2006). Effects of noise correlations on information encoding and decoding. *Journal of Neurophysiology* 95, 3633–3644.
- Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat Rev Neurosci* 7, 358–366.
- Awh, E., and Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends Cogn. Sci. (Regul. Ed.)* 5, 119–126.
- Awh, E., Vogel, E.K., and Oh, S.H. (2006). Interactions between attention and working memory. *Neuroscience* 139, 201–208.
- Baddeley, A. (1992). Working memory. *Science*.
- Baddeley, A.D., and Hitch, G. (1974). Working memory. In *The Psychology of Learning and Motivation: Advances in Research and Theory*, G.H. Bower, ed. (New York: Academic Press).

- Bair, W., Zohary, E., and Newsome, W.T. (2001). Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J. Neurosci.* *21*, 1676–1697.
- Barton, E.J., and Sparks, D.L. (2001). Saccades to remembered targets exhibit enhanced orbital position effects in monkeys. *Vision Res.* *41*, 2393–2406.
- Batuev, A.S. (1986). Neuronal mechanisms of goal-directed behavior in monkeys. *Neurosci. Behav. Physiol.* *16*, 459–465.
- Batuev, A.S. (1994). Two neuronal systems involved in short-term spatial memory in monkeys. *Acta Neurobiol Exp (Wars)* *54*, 335–344.
- Batuev, A.S., Pirogov, A.A., and Orlov, A.A. (1979). Unit activity of the prefrontal cortex during delayed alternation performance in monkey. *Acta Physiol Acad Sci Hung* *53*, 345–353.
- Batuev, A.S., Pirogov, A.A., Orlov, A.A., and Sheaffer, V.I. (1980). Cortical Mechanisms of Goal-directed Motor Acts in the Rhesus Monkey. *Acta Neurobiol Exp.*
- Batuev, A.S., Shaefer, V.I., and Orlov, A.A. (1985). Comparative characteristics of unit activity in the prefrontal and parietal areas during delayed performance in monkeys. *Behav. Brain Res.* *16*, 57–70.
- Bianchi, L. (1895). The Functions of the Frontal Lobes. *Brain* *18*, 497–522.
- Bianchi, L. (1922). The mechanism of the brain: and the function of the frontal lobes (Edinburgh: E. & S. Livingstone).
- Bichot, N.P., Heard, M.T., DeGennaro, E.M., and Desimone, R. (2015). A Source for Feature-Based Attention in the Prefrontal Cortex. *Neuron*.
- Blum, R.A. (1952). Effects of subtotal lesions of frontal granular cortex on delayed reaction in monkeys. *AMA Arch Neurol Psychiatry* *67*, 375–386.
- Boch, R.A., and Goldberg, M.E. (1989). Participation of prefrontal neurons in the preparation of visually guided eye movements in the rhesus monkey. *J Neurophysiol.*

References

- Botta, F., Santangelo, V., Raffone, A., Lupiáñez, J., and Belardinelli, M.O. (2010). Exogenous and endogenous spatial attention effects on visuospatial working memory. *Q J Exp Psychol (Hove)* *63*, 1590–1602.
- Bruce, C.J., and Goldberg, M.E. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. *Journal of Neurophysiology* *53*, 603–635.
- Bruce, C.J., Goldberg, M.E., Bushnell, M.C., and Stanton, G.B. (1985). Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. *Journal of Neurophysiology* *54*, 714–734.
- Bugbee, N.M., and Goldman-Rakic, P.S. (1983). Columnar organization of corticocortical projections in squirrel and rhesus monkeys: similarity of column width in species differing in cortical volume. *J. Comp. Neurol.* *220*, 355–364.
- Bullock, K.R., Pieper, F., Sachs, A.J., and Martinez-Trujillo, J.C. (2017). Visual and presaccadic activity in area 8Ar of the macaque monkey lateral prefrontal cortex. *Journal of Neurophysiology* *jn.00278.2016*.
- Buschman, T.J., Siegel, M., Roy, J.E., and Miller, E.K. (2011). Neural substrates of cognitive capacity limitations. *Pnas* *108*, 11252–11255.
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., and Munafò, M.R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* *14*, 365–376.
- Camperi, M., and Wang, X.J. (1998). A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *Journal of Computational Neuroscience* *5*, 383–405.
- Carlson, T.A., Alvarez, G.A., and Cavanagh, P. (2007). Quadrantic deficit reveals anatomical constraints on selection. *Pnas* *104*, 13496–13500.
- Carrasco, M., Talgar, C.P., and Cameron, E.L. (2001). Characterizing visual performance fields: effects of transient covert attention, spatial frequency, eccentricity, task and set size. *Spat Vis* *15*, 61–75.
- Chang, C.C., and Lin, C.J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* *2*, 27.

- Changizi, M.A., Hsieh, A., Nijhawan, R., Kanai, R., and Shimojo, S. (2008). Perceiving the present and a systematization of illusions. *Cogn Sci* 32, 459–503.
- Chelazzi, L., Miller, E.K., Duncan, J., and Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb. Cortex* 11, 761–772.
- Christophel, T.B., Klink, P.C., Spitzer, B., Roelfsema, P.R., and Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends Cogn. Sci. (Regul. Ed.)* 21, 111–124.
- Churchland, M.M., Yu, B.M., Cunningham, J.P., Sugrue, L.P., Cohen, M.R., Corrado, G.S., Newsome, W.T., Clark, A.M., Hosseini, P., Scott, B.B., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience* 13, 369–378.
- Cohen, J.Y., Crowder, E.A., Heitz, R.P., Subraveti, C.R., Thompson, K.G., Woodman, G.F., and Schall, J.D. (2010). Cooperation and competition among frontal eye field neurons during visual target selection. *J. Neurosci.* 30, 3227–3238.
- Cohen, M.R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience* 14, 811–819.
- Cohen, M.R., and Maunsell, J.H.R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience* 12, 1594–1600.
- Cohen, M.R., and Newsome, W.T. (2008). Context-Dependent Changes in Functional Circuitry in Visual Area MT. *Neuron* 60, 162–173.
- Compte, A. (2006). Computational and in vitro studies of persistent activity: edging towards cellular and synaptic mechanisms of working memory. *Neuroscience* 139, 135–151.
- Compte, A., Brunel, N., Goldman-Rakic, P.S., and Wang, X.-J. (2000). Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model. *Cerebral Cortex* 10, 910–923.
- Connolly, M., and Van Essen, D. (1984). The representation of the visual field in

References

- parvicellular and magnocellular layers of the lateral geniculate nucleus in the macaque monkey. *J. Comp. Neurol.* *226*, 544–564.
- Constantinidis, C., and Goldman-Rakic, P. (2002). Correlated Discharges Among Putative Pyramidal Neurons and Interneurons in the Primate Prefrontal Cortex. *Journal of Neurophysiology* *88*, 3487–3497.
- Constantinidis, C., and Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nat Rev Neurosci* *17*, 438–449.
- Constantinidis, C., and Procyk, E. (2004). The primate working memory networks. *Cognitive, Affective, & Behavioral Neuroscience* *4*, 444–465.
- Constantinidis, C., and Wang, X.-J. (2004). A Neural Circuit Basis for Spatial Working Memory. *The Neuroscientist* *10*, 553–565.
- Constantinidis, C., Franowicz, M.N., and Goldman-Rakic, P.S. (2001a). Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. *J. Neurosci.* *21*, 3646–3655.
- Constantinidis, C., Franowicz, M.N., and Goldman-Rakic, P.S. (2001b). The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nature Neuroscience* *4*, 311–316.
- Constantinidis, C., Williams, G.V., and Goldman-Rakic, P.S. (2002). A role for inhibition in shaping the temporal flow of information in prefrontal cortex. *Nature Neuroscience* *5*, 175–180.
- Conway, A.R.A., Kane, M.J., and Engle, R.W. (2003). Working memory capacity and its relation to general intelligence. *Trends Cogn. Sci. (Regul. Ed.)* *7*, 547–552.
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Machine Learning* *20*, 273–297.
- Courtney, S.M. (2004). Attention and cognitive control as emergent properties of information representation in working memory. *Cognitive, Affective, & Behavioral Neuroscience* *4*, 501–516.
- Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*.

- D'Esposito, M., and Postle, B.R. (2015). The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* *66*, 115–142.
- Dakin, S.C., Athorp, D., and Alais, D. (2010). Anisotropies in judging the direction of moving natural scenes. *Journal of Vision* *10*, 5.
- Dayan, P., and Abbott, L.F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (Cambridge: MIT Press).
- de Fockert, J.W. (2001). The Role of Working Memory in Visual Selective Attention. *Science* *291*, 1803–1806.
- Delvenne, J.-F. (2005). The capacity of visual short-term memory within and between hemifields. *Cognition* *96*, B79–B88.
- Delvenne, J.-F., Kaddour, L.A., and Castronovo, J. (2011). An electrophysiological measure of visual short-term memory capacity within and across hemifields. *Psychophysiology* *48*, 333–336.
- Desimone, R., Albright, T.D., Gross, C.G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *The Journal of Neuroscience* *4*, 2051–2062.
- Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*
- Desimone, R., Schein, S.J., Moran, J., and Ungerleider, L.G. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Res.*
- Dickersin, K., Chan, S., Chalmers, T.C., Sacks, H.S., and Smith, H. (1987). Publication bias and clinical trials. *Control Clin Trials* *8*, 343–353.
- Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nat Rev Neurosci* *2*, 820–829.
- Durstewitz, D., and Seamans, J.K. (2002). The computational role of dopamine D1 receptors in working memory. *Neural Networks* *15*, 561–572.
- Durstewitz, D., Seamans, J.K., and Sejnowski, T.J. (2000). Neurocomputational models of working memory. *Nature Neuroscience* *3*, 1184–1191.

References

- Easterbrook, P.J., Gopalan, R., Berlin, J.A., and Matthews, D.R. (1991). Publication bias in clinical research. *The Lancet*.
- Ecker, A.S., Berens, P., Keliris, G.A., and Bethge, M. (2010). Decorrelated Neuronal Firing in Cortical Microcircuits. *Science*.
- Elston, G.N. (2007). Specialization of the Neocortical Pyramidal Cell during Primate Evolution. In *Evolution of Nervous Systems*, J.H. Kaas, ed. (Elsevier), pp. 191–242.
- Emrich, S.M., Riggall, A.C., Larocque, J.J., and Postle, B.R. (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.
- Engle, R.W., Tuholski, S.W., Laughlin, J.E., and Conway, A.R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *J Exp Psychol Gen* *128*, 309–331.
- Everling, S., Tinsley, C.J., Gaffan, D., and Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neuroscience* *5*, 671–676.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research* *9*, 1871–1874.
- Ferrera, V.P., Cohen, J.K., and Lee, B.B. (1999). Activity of prefrontal neurons during location and color delayed matching tasks. *Neuroreport* *10*, 1315–1322.
- Ferrera, V.P., Rudolph, K.K., and Maunsell, J.H. (1994). Responses of neurons in the parietal and temporal visual pathways during a motion task. *The Journal of Neuroscience* *14*, 6171–6186.
- Freedman, D.J., and Miller, E.K. (2008). Neural mechanisms of visual categorization: insights from neurophysiology. *Neurosci Biobehav Rev* *32*, 311–329.
- Funahashi, S. (2006). Prefrontal cortex and working memory processes. *Neuroscience*.
- Funahashi, S. (2013). Space representation in the prefrontal cortex. *Prog. Neurobiol.* *103*, 131–155.

- Funahashi, S., and Inoue, M. (2000). Neuronal interactions related to working memory processes in the primate prefrontal cortex revealed by cross-correlation analysis. *Cerebral Cortex* *10*, 535–551.
- Funahashi, S., and Kubota, K. (1994). Working memory and prefrontal cortex. *Neuroscience Research* *21*, 1–11.
- Funahashi, S., and Takeda, K. (2002). Information processes in the primate prefrontal cortex in relation to working memory processes. *13*, 313–345.
- Funahashi, S., Bruce, C., and Goldman-Rakic, P.S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology* *61*, 331–349.
- Fusi, S., Miller, E.K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* *37*, 66–74.
- Fuster, J. (2008). *The Prefrontal Cortex* (London: Elsevier).
- Fuster, J.M. (1973). Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *Journal of Neurophysiology* *36*, 61–78.
- Fuster, J.M. (1990). Inferotemporal units in selective visual attention and short-term memory. *Journal of Neurophysiology* *64*, 681–697.
- Fuster, J.M., and Alexander, G.E. (1971). Neuron activity related to short-term memory. *Science* *173*, 652–654.
- Fuster, J.M., and Jervey, J.P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* *212*, 952–955.
- Gawne, T.J., Kjaer, T.W., Hertz, J.A., and Richmond, B.J. (1996). Adjacent visual cortical complex cells share about 20% of their stimulus-related information. *Cereb. Cortex* *6*, 482–489.
- Gazzaley, A., and Nobre, A.C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends Cogn. Sci. (Regul. Ed.)* *16*, 129–135.
- Gerits, A., and Vanduffel, W. (2013). Optogenetics in primates: a shining future?

References

- Trends in Genetics 29, 403–411.
- Gnadt, J.W., and Andersen, R.A. (1988). Memory related motor planning activity in posterior parietal cortex of macaque. *Experimental Brain Research* 70, 216–220.
- Gnadt, J.W., Bracewell, R.M., and Andersen, R.A. (1991). Sensorimotor transformation during eye movements to remembered visual targets. *Vision Res.* 31, 693–715.
- Goffart, L., Quinet, J., Chavane, F., and Masson, G.S. (2006). Influence of background illumination on fixation and visually guided saccades in the rhesus monkey. *Vision Res.*
- Goldberg, M.E., and Wurtz, R.H. (1972a). Activity of superior colliculus in behaving monkey. I. Visual receptive fields of single neurons. *Journal of Neurophysiology* 35, 542–559.
- Goldberg, M.E., and Wurtz, R.H. (1972b). Activity of superior colliculus in behaving monkey. II. Effect of attention on neuronal responses. *Journal of Neurophysiology* 35, 560–574.
- Goldman-Rakic, P.S. (1995). Cellular basis of working memory. *Neuron* 14, 477–485.
- Goldman-Rakic, P.S. (1996). Regional and cellular fractionation of working memory. *Pnas* 93, 13473–13480.
- Goldman-Rakic, P.S. (2011). Circuitry of Primate Prefrontal Cortex and Regulation of Behavior by Representational Memory (Hoboken, NJ, USA: John Wiley & Sons, Inc.).
- Goldman-Rakic, P.S., and Schwartz, M.L. (1982). Interdigitation of contralateral and ipsilateral columnar projections to frontal association cortex in primates. *Science* 216, 755–757.
- Goodwin, S.J., Blackman, R.K., Sakellaridi, S., and Chafee, M.V. (2012). Executive control over cognition: stronger and earlier rule-based modulation of spatial category signals in prefrontal cortex relative to parietal cortex. *J. Neurosci.* 32, 3499–3515.

- Graf, A.B.A., and Andersen, R.A. (2015). Predicting oculomotor behaviour from correlated populations of posterior parietal neurons. *Nat Comms* *6*, 6024.
- Graf, A.B.A., Kohn, A., Jazayeri, M., and Movshon, J.A. (2011). Decoding the activity of neuronal populations in macaque primary visual cortex. *Nature Neuroscience* *14*, 239–245.
- Gregoriou, G.G., Gotts, S.J., Zhou, H., and Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* *324*, 1207–1210.
- Gross, C.G., Bender, D.B., and Rocha-Miranda, C.E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* *166*, 1303–1306.
- Gu, Y., Liu, S., Fetsch, C.R., Yang, Y., Fok, S., Sunkara, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* *71*, 750–761.
- Gulyás, A.I., Hájos, N., and Freund, T.F. (1996). Interneurons containing calretinin are specialized to control other interneurons in the rat hippocampus. *The Journal of Neuroscience* *16*, 3397–3411.
- Gutnisky, D.A., and Dragoi, V. (2008). Adaptive coding of visual information in neural populations. *Nature* *452*, 220–224.
- Guyon, I., and Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research* *3*, 1157–1182.
- Harrison, S.A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature* *458*, 632–635.
- Haun, D.B.M., Allen, G.L., and Wedell, D.H. (2005). Bias in spatial memory: a categorical endorsement. *Acta Psychol (Amst)* *118*, 149–170.
- Hayden, B.Y., and Gallant, J.L. (2013). Working memory and decision processes in visual area v4. *Frontiers in Neuroscience* *7*, 18.
- Hebb, D.O. (1949). *The Organization of Behavior: A Neuropsychological Theory* (New York: John Wiley & Sons, Inc.).

References

- Hebb, D.O. (2005). *The Organization of Behavior: A Neuropsychological Theory* (New York: John Wiley & Sons, Inc.).
- Heeger, D. (2000). Poisson Model of Spike Generation.
- Hikosaka, O., Sakamoto, M., and Usui, S. (1989). Functional properties of monkey caudate neurons. I. Activities related to saccadic eye movements. *Journal of Neurophysiology* *61*, 780–798.
- Hirata, Y., and Sawaguchi, T. (2008). Functional columns in the primate prefrontal cortex revealed by optical imaging in vitro. *Neuroscience Research* *61*, 1–10.
- Hochberg, Y., and Tamhane, A.C. (1987). Single-Step Procedures for Pairwise and More General Comparisons among All Treatments. In *Multiple Comparison Procedures*, (John Wiley & Sons, Inc.), pp. 72–109.
- Hu, Y., Zylberberg, J., and Shea-Brown, E. (2014). The sign rule and beyond: boundary effects, flexibility, and noise correlations in neural population codes. *PLoS Comput. Biol.* *10*, e1003469.
- Hubel, D.H., Wiesel, T.N., and Stryker, M.P. (1977). Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique. *Nature* *269*, 328–330.
- Hussar, C., and Pasternak, T. (2010). Trial-to-trial variability of the prefrontal neurons reveals the nature of their engagement in a motion discrimination task. *Pnas* *107*, 21842–21847.
- Huttenlocher, J., Hedges, L.V., and Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological Review* *98*, 352–376.
- Huttenlocher, J., Hedges, L.V., Corrigan, B., and Crawford, L.E. (2004). Spatial categories and the estimation of location. *Cognition* *93*, 75–97.
- Ikkai, A., and Curtis, C.E. (2011). Common neural mechanisms supporting spatial working memory, attention and motor intention. *Neuropsychologia* *49*, 1428–1434.
- Jacobsen, C.F. (1935). FUNCTIONS OF FRONTAL ASSOCIATION AREA IN PRIMATES. *Arch NeurPsych* *33*, 558–569.

- Jacobsen, C.F., and Nissen, H.W. (1937). Studies of cerebral function in primates. IV. The effects of frontal lobe lesions on the delayed alternation habit in monkeys. *Journal of Comparative Psychology*.
- Jacobsen, C.F., Wolfe, J.B., and Jackson, T.A. (1935). AN EXPERIMENTAL ANALYSIS OF THE FUNCTIONS OF THE FRONTAL ASSOCIATION AREAS IN PRIMATES. *J. Nerv. Ment. Dis.* *82*, 1.
- Jones, B.W. (2009). Utah Array (Salt Lake City).
- Kandel, E.R., Schwartz, J.H., and Chao, J. (2000). *Principles of neural science* (New York: McGraw-Hill).
- Kanitscheider, I., Coen-Cagli, R., Kohn, A., and Pouget, A. (2015). Measuring Fisher information accurately in correlated neural populations. *PLoS Comput. Biol.* *11*, e1004218.
- Katsuki, F., and Constantinidis, C. (2013). Time course of functional connectivity in primate dorsolateral prefrontal and posterior parietal cortex during working memory. *PLoS ONE* *8*, e81601.
- Katsuki, F., Qi, X.-L., Meyer, T., Kostelic, P.M., Salinas, E., and Constantinidis, C. (2014). Differences in intrinsic functional organization between dorsolateral prefrontal and posterior parietal cortex. *Cerebral Cortex* *24*, 2334–2349.
- Kety, S.S. (1957). The general metabolism of the brain in vivo. In *Metabolism of the Nervous System*, (London: Pergamon), pp. 221–237.
- Khayat, P.S., Niebergall, R., and Martinez-Trujillo, J.C. (2010). Attention differentially modulates similar neuronal responses evoked by varying contrast and direction stimuli in area MT. *J. Neurosci.* *30*, 2188–2197.
- Kiani, R., Cueva, C.J., Reppas, J.B., Peixoto, D., Ryu, S.I., and Newsome, W.T. (2015). Natural Grouping of Neural Responses Reveals Spatially Segregated Clusters in Prearcuate Cortex. *Neuron*.
- Kinoshita, M., and Isa, T. (2015). Potential of Optogenetics for the Behavior Manipulation of Non-human Primates. In *Optogenetics*, (Tokyo: Springer Japan), pp. 279–290.

References

- Kohn, A., and Smith, M.A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.* *25*, 3661–3673.
- Kojima, S., and Goldman-Rakic, P.S. (1982). Delay-related activity of prefrontal neurons in rhesus monkeys performing delayed response. *Brain Research* *248*, 43–49.
- Kojima, S., and Goldman-Rakic, P.S. (1984). Functional analysis of spatially discriminative neurons in prefrontal cortex of rhesus monkey. *Brain Research* *291*, 229–240.
- Komatsu, H., and Wurtz, R.H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *Journal of Neurophysiology* *60*, 580–603.
- Kritzer, M.F., and Goldman-Rakic, P.S. (1995). Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* *359*, 131–143.
- Kubota, K., and Niki, H. (1971). Prefrontal cortical unit activity and delayed alternation performance in monkeys. *Journal of Neurophysiology* *34*, 337–347.
- la Rocha, de, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature* *448*, 802–806.
- LaBar, K.S., Gitelman, D.R., Parrish, T.B., and Mesulam, M. (1999). Neuroanatomic overlap of working memory and spatial attention networks: a functional MRI comparison within subjects. *Neuroimage* *10*, 695–704.
- Laing, C.R., Troy, W.C., Gutkin, B., and Ermentrout, G.B. (2002). Multiple Bumps in a Neuronal Model of Working Memory. *SIAM Journal on Applied Mathematics* *63*, 62–97.
- Lashley, K.S. (1920). Studies of cerebral function in learning. *Psychobiology* *2*, 55.
- Leavitt, M.L., Mendoza-Halliday, D., and Martinez-Trujillo, J.C. (2017a). Sustained Activity Encoding Working Memories: Not Fully Distributed. *Trends in Neurosciences* *40*, 328–346.

- Leavitt, M.L., Pieper, F., Sachs, A., Joober, R., and Martinez-Trujillo, J.C. (2013). Structure of spike count correlations reveals functional interactions between neurons in dorsolateral prefrontal cortex area 8a of behaving primates. *PLoS ONE* 8, e61503.
- Leavitt, M.L., Pieper, F., Sachs, A.J., and Martinez-Trujillo, J.C. (2017b). A Quadrantic Bias in Prefrontal Representation of Visual-Mnemonic Space. *Cerebral Cortex* 1–17.
- Leavitt, M.L., Pieper, F., Sachs, A.J., and Martinez-Trujillo, J.C. (2017c). Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles. *Pnas* 114, E2494–E2503.
- Lebedev, M.A., Messinger, A., Kralik, J.D., and Wise, S.P. (2004). Representation of attended versus remembered locations in prefrontal cortex. *PLoS Biol.* 2, e365.
- Lee, D. (2008). Game theory and neural basis of social decision making. *Nature Neuroscience* 11, 404–409.
- Lennert, T., and Martinez-Trujillo, J.C. (2011). Strength of response suppression to distracter stimuli determines attentional-filtering performance in primate prefrontal neurons. *Neuron* 70, 141–152.
- Lennert, T., and Martinez-Trujillo, J.C. (2013). Prefrontal neurons of opposite spatial preference display distinct target selection dynamics. *J. Neurosci.* 33, 9520–9529.
- Levitt, J.B., Lewis, D.A., Yoshioka, T., and Lund, J.S. (1993). Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). *J. Comp. Neurol.* 338, 360–376.
- Liu, T., Heeger, D.J., and Carrasco, M. (2006). Neural correlates of the visual vertical meridian asymmetry. *Journal of Vision* 6, 1294–1306.
- Liu, T., Jiang, Y., Sun, X., and He, S. (2009). Reduction of the crowding effect in spatially adjacent but cortically remote visual stimuli. *Current Biology* 19, 127–132.
- Loffler, G., and Orbach, H.S. (2001). Anisotropy in judging the absolute direction of motion. *Vision Res.* 41, 3677–3692.

References

- Luck, S.J., Chelazzi, L., Hillyard, S.A., and Desimone, R. (1997). Neural Mechanisms of Spatial Selective Attention in Areas V1, V2, and V4 of Macaque Visual Cortex. *Journal of Neurophysiology* *77*, 24–42.
- MacDougall, M., Nummela, S.U., Coop, S., Disney, A., Mitchell, J.F., and Miller, C.T. (2016). Optogenetic manipulation of neural circuits in awake marmosets. *Journal of Neurophysiology* *116*, 1286–1294.
- Malmo, R.B. (1942). Interference factors in delayed response in monkeys after removal of frontal lobes. *Journal of Neurophysiology*.
- Markowitz, D.A., Curtis, C.E., and Pesaran, B. (2015). Multiple component networks support working memory in prefrontal cortex. *Pnas* *112*, 11084–11089.
- Markowitz, D.A., Wong, Y.T., Gray, C.M., and Pesaran, B. (2011). Optimizing the decoding of movement goals from local field potentials in macaque cortex. *J. Neurosci.* *31*, 18412–18422.
- Matsushima, A., and Tanaka, M. (2014). Different neuronal computations of spatial working memory for multiple locations within versus across visual hemifields. *J. Neurosci.* *34*, 5621–5626.
- Maunsell, J.H., and Van Essen, D.C. (1987). Topographic organization of the middle temporal visual area in the macaque monkey: representational biases and the relationship to callosal connections and myeloarchitectonic boundaries. *J. Comp. Neurol.* *266*, 535–555.
- Maynard, E.M., Nordhausen, C.T., and Normann, R.A. (1997). The Utah intracortical Electrode Array: a recording structure for potential brain-computer interfaces. *Electroencephalogr Clin Neurophysiol* *102*, 228–239.
- McAdams, C.J., and Maunsell, J.H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *The Journal of Neuroscience* *19*, 431–441.
- Mendoza, D., Schneiderman, M., Kaul, C., and Martinez-Trujillo, J. (2011). Combined effects of feature-based working memory and feature-based attention on the perception of visual motion direction. *Journal of Vision* *11*, 11–11.

- Mendoza-Halliday, D., Torres, S., and Martinez-Trujillo, J.C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nature Neuroscience* *17*, 1255–1262.
- Merchant, H., Crowe, D.A., Robertson, M.S., Fortes, A.F., and Georgopoulos, A.P. (2011). Top-down spatial categorization signal from prefrontal to posterior parietal cortex in the primate. *Front Syst Neurosci* *5*, 69.
- Merchant, H., Fortes, A.F., and Georgopoulos, A.P. (2004). Short-term memory effects on the representation of two-dimensional space in the rhesus monkey. *Anim Cogn* *7*, 133–143.
- Messinger, A., Lebedev, M.A., Kralik, J.D., and Wise, S.P. (2009). Multitasking of attention and memory functions in the primate prefrontal cortex. *J. Neurosci.* *29*, 5640–5653.
- Meyers, E.M., Qi, X.-L., and Constantinidis, C. (2012). Incorporation of new information into prefrontal cortical activity after learning working memory tasks. *Pnas* *109*, 4651–4656.
- Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* *24*, 167–202.
- Miller, E.K., Erickson, C., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience* *16*, 5154–5167.
- Milner, B. (1982). Some cognitive effects of frontal-lobe lesions in man. *Philosophical Transactions of the Royal Society B: Biological Sciences* *298*, 211–226.
- Milner, B., and Petrides, M. (1984). Behavioural effects of frontal-lobe lesions in man. *Trends in Neurosciences* *7*, 403–407.
- Mishkin, M., and Pribram, K.H. (1955). Analysis of the effects of frontal lesions in monkey. I. Variations of delayed alternation. *Journal of Comparative and Physiological Psychology* *48*, 492–495.
- Mishkin, M., and Pribram, K.H. (1956). Analysis of the effects of frontal lesions in monkey. II. Variations of delayed response. *Journal of Comparative and*

References

- Physiological Psychology 49, 36–40.
- Mitchell, J.F., Sundberg, K.A., and Reynolds, J.H. (2009). Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. *Neuron* 63, 879–888.
- Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika* 37, 17–23.
- Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. *Nature Neuroscience*.
- Murray, J.D., Bernacchia, A., Roy, N.A., Constantinidis, C., Romo, R., and Wang, X.-J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Pnas* 114, 394–399.
- Nandy, A.S., Mitchell, J.F., Jadi, M.P., and Reynolds, J.H. (2016). Neurons in Macaque Area V4 Are Tuned for Complex Spatio-Temporal Patterns. *Neuron* 91, 920–930.
- Nienborg, H., and Cumming, B. (2010). Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron's causality? *Curr. Opin. Neurobiol.* 20, 376–381.
- Niki, H. (1974a). Differential activity of prefrontal units during right and left delayed response trials. *Brain Research* 70, 346–349.
- Niki, H. (1974b). Prefrontal unit activity during delayed alternation in the monkey. II. Relation to absolute versus relative direction of response. *Brain Research* 68, 197–204.
- Niki, H., and Watanabe, M. (1976). Prefrontal unit activity and delayed response: Relation to cue location versus direction of response. *Brain Research*.
- Niven, J.E., and Laughlin, S.B. (2008). Energy limitation as a selective pressure on the evolution of sensory systems. *Journal of Experimental Biology*.
- Normann, R.A., Maynard, E.M., Rousche, P.J., and Warren, D.J. (1999). A neural interface for a cortical vision prosthesis. *Vision Res.* 39, 2577–2587.

- Noudoost, B., Chang, M.H., Steinmetz, N.A., and Moore, T. (2010). Top-down control of visual attention. *Curr. Opin. Neurobiol.* *20*, 183–190.
- Opris, I., and Casanova, M.F. (2014). Prefrontal cortical minicolumn: from executive control to disrupted cognitive processing. *Brain* *137*, 1863–1875.
- Owen, A.M., Evans, A.C., and Petrides, M. (1996). Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: a positron emission tomography study. *Cerebral Cortex* *6*, 31–38.
- Palm, G., Aertsen, A.M., and Gerstein, G.L. (1988). On the significance of correlations among neuronal spike trains. *Biol Cybern* *59*, 1–11.
- Passingham, R.E. (1993). The frontal lobes and voluntary action. (Oxford University Press).
- Pasternak, T., and Greenlee, M.W. (2005). Working memory in primate sensory systems. *Nat Rev Neurosci* *6*, 97–107.
- Petrides, M. (1982). Motor conditional associative-learning after selective prefrontal lesions in the monkey. *Behav. Brain Res.* *5*, 407–413.
- Petrides, M. (1987). Conditional learning and the primate frontal cortex. In *The Frontal Lobes Revisited*, E. Perecman, ed. (New York: The IRBN Press), pp. 91–108.
- Petrides, M. (1994). Frontal lobes and behaviour. *Curr. Opin. Neurobiol.*
- Petrides, M. (1996). Specialized Systems for the Processing of Mnemonic Information within the Primate Frontal Cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences* *351*, 1455–1462.
- Petrides, M. (2000a). Dissociable Roles of Mid-Dorsolateral Prefrontal and Anterior Inferotemporal Cortex in Visual Working Memory. *The Journal of Neuroscience*.
- Petrides, M. (2000b). The role of the mid-dorsolateral prefrontal cortex in working memory. In *Executive Control and the Frontal Lobe: Current Issues*, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 44–54.
- Petrides, M. (2005a). Lateral prefrontal cortex: architectonic and functional organization. *Proc. R. Soc. B* *360*, 781–795.

References

- Petrides, M. (2005b). The rostral–caudal axis of cognitive control within the lateral frontal cortex. In *From Monkey Brain to Human Brain*, S. Dehaene, J.-R. Duhamel, M.D. Hauser, and G. Rizzolatti, eds. (Cambridge: MIT Press).
- Petrides, M., and Pandya, D.N. (1999). Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *Eur J Neurosci* *11*, 1011–1036.
- Petrides, M., and Pandya, D.N. (2007). Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *J. Neurosci.* *27*, 11573–11586.
- Polk, A., Litwin-Kumar, A., and Doiron, B. (2012). Correlated neural variability in persistent state networks. *Pnas* *109*, 6295–6300.
- Postle, B.R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience* *139*, 23–38.
- Pribram, K.H., and Mishkin, M. (1956). Analysis of the effects of frontal lesions in monkey. III. Object alternation. *Journal of Comparative and Physiological Psychology* *49*, 41–45.
- Qi, X.-L., and Constantinidis, C. (2012). Correlated discharges in the primate prefrontal cortex before and after working memory training. *Eur J Neurosci* *36*, 3538–3548.
- Quijan Quiroga, R., and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci* *10*, 173–185.
- Raiquel, S., Van Hulle, M.M., Xiao, D.K., Marcar, V.L., Lagae, L., and Orban, G.A. (1997). Size and shape of receptive fields in the medial superior temporal area (MST) of the macaque. *Neuroreport* *8*, 2803–2808.
- Rainer, G., Asaad, W.F., and Miller, E.K. (1998). Memory fields of neurons in the primate prefrontal cortex. *Pnas*.
- Rao, S.C., RAINER, G., and Miller, E.K. (1997). Integration of what and where in the primate prefrontal cortex. *Science* *276*, 821–824.
- Rao, S.G., Williams, G.V., and Goldman-Rakic, P.S. (1999). Isodirectional tuning of

- adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization in PFC. *Journal of Neurophysiology* 81, 1903–1916.
- Reich, D.S., Mechler, F., and Victor, J.D. (2001). Independent and redundant information in nearby cortical neurons. *Science* 294, 2566–2568.
- Reuter-Lorenz, P.A., and Fendrich, R. (1992). Oculomotor readiness and covert orienting: differences between central and peripheral precues. *Percept Psychophys* 52, 336–344.
- Reynolds, J.H., and Chelazzi, L. (2004). Attentional modulation of visual processing. *Annu. Rev. Neurosci.* 27, 611–647.
- Riggall, A.C., and Postle, B.R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *J. Neurosci.* 32, 12990–12998.
- Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590.
- Riley, M.R., and Constantinidis, C. (2015). Role of Prefrontal Persistent Activity in Working Memory. *Front Syst Neurosci* 9, 181.
- Rizzolatti, G., Riggio, L., Dascola, I., and Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia* 25, 31–40.
- Romo, R., Hernandez, A., Zainos, A., and Salinas, E. (2003). Correlated Neuronal Discharges that Increase Coding Efficiency during Perceptual Discrimination. *Neuron* 38, 649–657.
- Ruff, D.A., and Cohen, M.R. (2014). Attention can either increase or decrease spike count correlations in visual cortex. *Nature Neuroscience* 17, 1591–1597.
- Sakurai, Y., and Takahashi, S. (2006). Dynamic synchrony of firing in the monkey prefrontal cortex during working-memory tasks. *J. Neurosci.* 26, 10141–10153.
- Sawaguchi, T., and Goldman-Rakic, P.S. (1991). D1 dopamine receptors in prefrontal cortex: involvement in working memory. *Science* 251, 947–950.

References

- Schafer, R.J., and Moore, T. (2011). Selective attention from voluntary control of neurons in prefrontal cortex. *Science* *332*, 1568–1571.
- Schall, J.D., Morel, A., King, D.J., and Bullier, J. (1995). Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. *The Journal of Neuroscience* *15*, 4464–4487.
- Scott, A.J., and Seber, G.A.F. (1983). Difference of Proportions from the Same Survey. *The American Statistician* *37*, 319–320.
- Selemon, L.D., and Goldman-Rakic, P.S. (1988). Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. *The Journal of Neuroscience* *8*, 4049–4068.
- Serences, J.T., Ester, E.F., Vogel, E.K., and Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychol Sci* *20*, 207–214.
- Shadlen, M.N., and Newsome, W.T. (1994). Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* *4*, 569–579.
- Shadlen, M.N., and Newsome, W.T. (1998). The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding. *The Journal of Neuroscience*.
- Shamir, M., and Sompolinsky, H. (2006). Implications of neuronal diversity on population coding. *Neural Comput* *18*, 1951–1986.
- Sheliga, B.M., Riggio, L., and Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research* *98*.
- Smith, D.J. (2003). From Frontal Lobe Syndrome to Dysexecutive Syndrome. [Http://Www.Smithsrisc.co.Uk/](http://Www.Smithsrisc.co.Uk/).
- Smith, D.T., and Schenk, T. (2012). The Premotor theory of attention: Time to move on? *Neuropsychologia* *50*, 1104–1114.
- Smith, M.A., and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* *28*, 12591–12603.

- Sohal, V.S., Zhang, F., Yizhar, O., and Deisseroth, K. (2009). Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* *459*, 698–702.
- Sommer, M.A., and Wurtz, R.H. (2000). Composition and topographic organization of signals sent from the frontal eye field to the superior colliculus. *Journal of Neurophysiology* *83*, 1979–2001.
- Sommer, M.A., and Wurtz, R.H. (2001). Frontal eye field sends delay activity related to movement, memory, and vision to the superior colliculus. *Journal of Neurophysiology* *85*, 1673–1685.
- Squire, R.F., Noudoost, B., Schafer, R.J., and Moore, T. (2013). Prefrontal contributions to visual selective attention. *Annu. Rev. Neurosci.* *36*, 451–466.
- Stokes, M.G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron* *78*, 364–375.
- Stuss, D.T., and Benson, D.F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin* *95*, 3.
- Supèr, H., Spekreijse, H., and Lamme, V.A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science* *293*, 120–124.
- Suzuki, H., and Azuma, M. (1983). Topographic studies on visual neurons in the dorsolateral prefrontal cortex of the monkey. *Experimental Brain Research* *53*, 47–58.
- Szabo, M., Almeida, R., Deco, G., and Stetter, M. (2004). Cooperation and biased competition model can explain attentional filtering in the prefrontal cortex. *Eur J Neurosci* *19*, 1969–1977.
- Takaura, K., Yoshida, M., and Isa, T. (2011). Neural substrate of spatial memory in the superior colliculus after damage to the primary visual cortex. *J. Neurosci.* *31*, 4233–4241.
- Takeda, K., and Funahashi, S. (2002). Prefrontal Task-Related Activity Representing Visual Cue Location or Saccade Direction in Spatial Working Memory Tasks. *Journal of Neurophysiology* *87*, 567–588.
- Theeuwes, J., Belopolsky, A., and Olivers, C.N.L. (2009). Interactions between

References

- working memory, attention and eye movements. *Acta Psychol (Amst)* *132*, 106–114.
- Tremblay, S., Doucet, G., Pieper, F., Sachs, A., and Leavitt, M. (2015). Single-Trial Decoding of Visual Attention from Local Field Potentials in the Primate Lateral Prefrontal Cortex Is Frequency-Dependent. *J. Neurosci.* *35*, 9038–9049.
- Tremblay, S., Pieper, F., Sachs, A., and Leavitt, M. (2014). Attentional Filtering of Visual Information by Neuronal Ensembles in the Primate Lateral Prefrontal Cortex. *Neuron*.
- Treue, S., and Martinez-Trujillo, J.C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* *399*, 575–579.
- Tsujimoto, S., and Postle, B.R. (2012). The prefrontal cortex and oculomotor delayed response: a reconsideration of the "mnemonic scotoma". *J Cogn Neurosci* *24*, 627–635.
- Tsujimoto, S., Genovesio, A., and Wise, S.P. (2008). Transient neuronal correlations underlying goal selection and maintenance in prefrontal cortex. *Cerebral Cortex* *18*, 2748–2761.
- Urban, D.J., and Roth, B.L. (2015). DREADDs (designer receptors exclusively activated by designer drugs): chemogenetic tools with therapeutic utility. *Annu. Rev. Pharmacol. Toxicol.* *55*, 399–417.
- Van Essen, D.C., Newsome, W.T., and Maunsell, J.H. (1984). The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Res.* *24*, 429–448.
- van Kerkoerle, T., Self, M.W., and Roelfsema, P.R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat Comms* *8*, 13804.
- Virsu, V., and Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research* *37*, 475–494.
- Visscher, K., Viets, E., and Snyder, L.H. (2003). Effects of training on memory-guided saccade performance. *Vision Res.* *43*, 2061–2071.
- Vogel, E.K., and Machizawa, M.G. (2004). Neural activity predicts individual

- differences in visual working memory capacity. *Nature* *428*, 748–751.
- Wang, B., Yin, L., Zou, X., Ye, M., Liu, Y., He, T., Deng, S., Jiang, Y., Zheng, R., Wang, Y., et al. (2015). A Subtype of Inhibitory Interneuron with Intrinsic Persistent Activity in Human and Monkey Neocortex. *Cell Rep.*
- Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences* *24*, 455–463.
- Wang, X.J. (2006). A microcircuit model of prefrontal functions: Ying and Yang of reverberatory neurodynamics in cognition. In *The Frontal Lobes: Development, Function and Pathology*, J. Risberg, and J. Grafman, eds. (Cambridge University Press), pp. 92–127.
- Wang, X.J. (2013). The Prefrontal Cortex as a Quintessential “Cognitive-Type” Neural Circuit: Working Memory and Decision Making. In *Principles of Frontal Lobe Function*, D.T. Stuss, and R.T. Knight, eds. (Oxford University Press), pp. 226–248.
- Wang, X.J., Tegnér, J., Constantinidis, C., and Goldman-Rakic, P.S. (2004). Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory. *Pnas* *101*, 1368–1373.
- Watanabe, Y., and Funahashi, S. (2004). Neuronal activity throughout the primate mediodorsal nucleus of the thalamus during oculomotor delayed-responses. I. Cue-, delay-, and response-period activity. *Journal of Neurophysiology* *92*, 1738–1755.
- White, J.M., Sparks, D.L., and Stanford, T.R. (1994). Saccades to remembered target locations: an analysis of systematic and variable errors. *Vision Res.* *34*, 79–92.
- Williams, G.V., and Goldman-Rakic, P.S. (1995). Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature* *376*, 572–575.
- Wilson, F.A., Scalaidhe, S.P., and Goldman-Rakic, P.S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* *260*, 1955–1958.
- Wimmer, K., Nykamp, D.Q., Constantinidis, C., and Compte, A. (2014). Bump

References

- attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*.
- Wise, S.P., and Murray, E.A. (2000). Arbitrary associations between antecedents and actions. *Trends in Neurosciences* *23*, 271–276.
- Wurtz, R.H., and Goldberg, M.E. (1972). Activity of superior colliculus in behaving monkey. III. Cells discharging before eye movements. *Journal of Neurophysiology* *35*, 575–586.
- Xiao, D., Zikopoulos, B., and Barbas, H. (2009). Laminar and modular organization of prefrontal projections to multiple thalamic nuclei. *Neuroscience* *161*, 1067–1081.
- Yamashita, O., Sato, M.-A., Yoshioka, T., Tong, F., and Kamitani, Y. (2008). Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. *Neuroimage* *42*, 1414–1429.
- Zipser, D., Kehoe, B., Littlewort, G., and Fuster, J. (1993). A spiking network model of short-term active memory. *The Journal of Neuroscience* *13*, 3406–3420.
- Zohary, E., Shadlen, M.N., and Newsome, W.T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* *370*, 140–143.
- Zuur, A., Ieno, E.N., and Smith, G.M. (2007). *Analysing Ecological Data*.