

Multi-item Memory in the Primate Prefrontal Cortex

by

Melissa R. Warden

A.B., Molecular Biology
Princeton University, 1996

Submitted to the Department of Brain and Cognitive Sciences in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy in Systems Neuroscience

at the

Massachusetts Institute of Technology

June 2006

© 2006 Massachusetts Institute of Technology.
All rights reserved.

Signature of Author: _____
Department of Brain & Cognitive Sciences
April 28, 2006

Certified by: _____
Earl K. Miller
Picower Professor of Neuroscience
Associate Director, Picower Institute for Learning and Memory
Thesis Supervisor

Accepted by: _____
Matthew Wilson
Professor of Neurobiology
Picower Scholar
Chairman, Department Graduate Committee

Multi-item Memory in the Primate Prefrontal Cortex

by

Melissa R. Warden

Submitted to the Department of Brain and Cognitive Sciences
on April 7, 2006 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy in
Systems Neuroscience

ABSTRACT

The ability to retain multiple items in short-term memory is fundamental for cognition, yet almost nothing is known about its neural basis. To explore the mechanisms underlying this ability, we trained two monkeys to remember a sequence of two images across a short delay. We then recorded the activity of neurons from the lateral prefrontal cortex during task performance. We found that the majority of neurons showed delay activity that depended on the identity of both images (a minority reflected just one image), and that activity related to a given combination of images was only partially predictable from each neuron's activity to individual images. A model to predict the resultant neural activity was tested.

We also examined the effect of task demands on the neural representation of multiple images. Our first experiment showed that each of the two images in memory was represented with a certain strength, and that this strength was dependent on how long the image had been in memory; image strength decayed as time progressed. We found that changing the way that the memory of the images was reported, from a bar release to a sequence of eye movements, changed the relative strength of the image representations. In the eye-movement version of the task the strength of the representation of the image did not decay with time; in fact the strength of older images could even surpass the strength of newer images, depending on how frequently the tasks were switched. Further experiments showed that when the monkey switched between the two tasks individual neurons could turn their image coding on and off. We also found a substantial population of cells that directly represented the task that the animal was performing.

Thesis Supervisor: Earl K. Miller

Title: Picower Professor of Visual Neuroscience

Associate Director, Picower Institute for Learning and Memory

TABLE OF CONTENTS

Abstract	3
Table of Contents	5
Acknowledgments	6
Chapter 1. Introduction	7
The Anatomy of Working Memory	9
Memory for Multiple Items	10
Memory for Sequences	11
Task-Selective Responses in the Prefrontal Cortex	12
Interaction of Task-Selectivity and Image-Selectivity	14
Multi-Item Memory Experiment	14
Task-Dependent Multi-Item Memory Experiment	15
References	17
Chapter 2. Representation of Multiple Items in Prefrontal Delay Activity	21
Introduction	22
Experimental Design and Methods	25
Results	32
Discussion	43
Figure Legends	48
Figures	52
References	61
Chapter 3. Task-Specific Representation of Multiple Items in the Primate Prefrontal Cortex	65
Introduction	66
Experimental Design and Methods	71
Results	80
Discussion	88
Figure Legends	91
Figures	96
References	109
Chapter 4. Conclusion	112
The Representation of Multiple Items in Memory	113
Task-Dependent Representation of Multiple Items	114
Future Directions	115
Conclusion	117
References	118

ACKNOWLEDGMENTS

I would like to thank my advisor, Earl Miller, for giving me the opportunity to do my thesis research in his laboratory. I couldn't have asked for a richer or more stimulating environment or a better advisor, and have loved my time in the lab. During my stay here I have grown up as a scientist, and have watched myself transform from a naïve, accepting student into a critical thinker, largely due to his influence. Earl runs a very happy group, and it was a lot of fun to work in that environment. I will miss our poker nights.

I would also like to thank Bill Newsome, who gave me a tech job in his lab when I was transitioning from molecular biology and deeply conflicted about what I wanted to study. The experiments I observed in his lab convinced me that electrophysiology was the coolest thing in the world. Bill was always very generous with his time, and I learned a great deal about visual neuroscience and the way good science is done during my stay there. I would especially like to thank Greg Horwitz, who taught me how to record in the superior colliculus with good humor and cheer.

When I first started graduate school I would have been lost without my conversations with Wael Asaad. Wael seemed (and still seems) infinitely wise to me, and I have a great respect not only for his knowledge of the field of neuroscience, but also for his ability to come up with a million new experiments in just a few seconds. It is always inspiring to talk to him. Kristin MacCully has been helpful with just about everything under the sun, and makes managing a lab look easy, even though I know it's not. The other members and former members of the Miller Lab (Scott Brincat, Tim Buschman, Cinira Diogo, Dave Freedman, Mark Histed, Christopher Irving, Michelle Machon, Anitha Pasupathy, Vicky Puig, Jefferson Roy, and Jon Wallis) have always been there to discuss problems, help fix bugs, or go out for a beer.

None of us could accomplish anything without the sage guidance of our vet, Bob Marini. I can't imagine working at a different university without him around to save the day or sing a funny song. Also, I really have to acknowledge my monkeys, Samantha and Arthur, even though it seems silly. If they didn't do their job I wouldn't have a thesis.

My thesis has benefited from the advice of my committee (Peter Schiller, Matt Wilson, Earl Miller, and Mark Goldman), and I thank them for their time and effort.

I couldn't have kept my sanity without the advice of my three best friends from college, Aylin Ulku, Dasein Sophia Howell, and Eiko Mishima, wise women all of them. And my life at MIT has been made happy and fulfilling by my friends here: Andrea Zanzotto, Benjamin Monreal, Daniel Collins, Jessie Shelton and Jun Song.

I would especially like to thank my husband, Ian Ellwood, for all the love and support he has given me over the years. I would like to thank my mom, a one-time monkey researcher herself, for instilling the love of knowledge in me and encouraging me to go down the academic path, and I thank my dad and my sister Sarah for their love and encouragement.

CHAPTER 1

Introduction

INTRODUCTION

As an organism interacts with its environment, it is faced with a constant barrage of stimuli that it must sense, remember, and use. In the majority of situations outside of the laboratory environment, an organism must remember not only the most recent stimulus, but must also integrate this stimulus with other items already present in memory. Remembering and integrating several relevant, recently-seen items is fundamental to the process of reasoning effectively about the environment, because creating appropriate behavioral strategies requires many pieces of information. The brain must be able to maintain and operate on multiple items in working memory.

Furthermore, it is important to construct an accurate representation of the temporal structure of events in order to determine cause and effect relationships and to plan useful action sequences. For example, the ability to hold a phone number in mind during the brief interval between reading and dialing requires both the ability to store the digits that make up the number, as well as the ability to store them in their correct temporal order. A more naturalistic example of this phenomenon would be a task that animals do all the time – they remember how to get from point A to point B, from their nest to the watering hole, a task that most animals accomplish by memorizing a temporal string of landmarks. Both the problem of temporal order and the problem of the simultaneous maintenance of multiple items are unsolved, although there have been some tantalizing hints at possible mechanisms.

The anatomy of working memory

What brain structure is responsible for the ability to maintain multiple items in short term memory? Decades of research have firmly established the dorsolateral prefrontal cortex as critical for the maintenance of single items in short term memory. This would, therefore, be the first place to look for a multi-item memory store. The first evidence that the prefrontal cortex might be important for memory came from anatomical and behavioral studies on non-human primates. Many studies found that if regions of the frontal lobe sustained damage, either through lesions or through reversible inactivation, short term memory was affected adversely (Blum, 1952; Harlow et al., 1952; Mishkin, 1957; Fuster and Alexander, 1970; Goldman and Rosvold, 1970; Butters et al., 1971; Passingham, 1975; Mishkin and Manning, 1978). These results, coupled with the development of the microelectrode, led to the first neurophysiological studies of prefrontal cortex, which established the existence of stimulus-selective delay period activity (Fuster and Alexander, 1971; Kubota and Niki, 1971; Fuster, 1973; Funahashi et al., 1989; Miller et al., 1996). This activity has since been interpreted as the neural signature of short term memory, an idea originally suggested by Donald Hebb (Hebb, 1949). Prefrontal cortex, therefore, has been established as a locus in the brain that not only was required for short term memory, but that also contained neurons with an attractive physiological profile. Subsequent physiological studies have found selective delay period activity in other areas of the brain, including inferotemporal cortex (Miyashita and Chang, 1988; Miller and Desimone, 1994) and parietal cortex (Gnadt and Andersen, 1988; Gnadt et al., 1991) among others; however, in no area of the brain is delay activity as robust and predominant as in the prefrontal cortex.

Memory for multiple items

Several major issues confront us when we think about how the fundamental problem of storing multiple short-term memories might be solved. There is now an abundance of evidence that single items are represented in PFC activity by a rate code (Fuster and Alexander, 1971; Kubota and Niki, 1971; Fuster, 1973; Funahashi et al., 1989; Miller et al., 1996). What about multiple items? Is information about all of them reflected in PFC activity? If so, how might this be accomplished? One possible scenario is that there is a separate population of cells responsible for maintaining the memory of each item. This mechanism would be analogous to the concept of the address in computer memory, where each item to be stored is placed in its own 'box'. Another possible way to store multiple memories would be to suppose that the PFC contains single cells that are capable of representing more than one item simultaneously. In this scenario, the representation of each item would be distributed among the entire population of prefrontal cells, or at least a large fraction of the population. If this is so, it is interesting to think about the form that this type of storage might take, particularly with regard to how multiple, sequentially presented items might be stored. Does the delay activity of a neuron representing a single memory relate in a straightforward way that neuron's to multi-item activity?

Most of the research on the memory for multiple items thus far has been at a psychological level, and there have been some interesting results. One topic of research has been the capacity of short term memory. Early studies showed that short term memory has a capacity of 7 ± 2 items (Miller, 1956), and many studies have shown that the capacity of working memory is of approximately this size (Nickerson, 1965).

However, more recent results suggest that the number may be closer to 3 or 4 if the items to be remembered are pictographic and not verbal (Luck and Vogel, 1997). Another well studied and related topic is the memory of lists of items. Many behavioral results have been reported, but one of the most important and robust is the serial position effect. When subjects are asked to maintain a list of several items in short term memory, their memory for the first and last items tends to be better than their memory for items that fall in the middle of the sequence (Murdock, 1962). Short term memory capacity and the serial position effect have both been investigated using fMRI, and a mixed pattern of results has emerged (Callicott et al., 1999; Osaka et al., 2003; Vogel and Machizawa, 2004; Talmi et al., 2005).

Many theorists have attempted to explain how multiple items could be stored in the brain (Lashley, 1951; Jensen and Lisman, 1996; Tanaka, 2002; Amit et al., 2003). These models usually also have a component that deals with the representation of temporal order, and range from the physiologically inspired to the more abstract. However, none have yet been directly experimentally verified. These models will be discussed in more detail in Chapter 2.

Memory for sequences

Although memory for single items, both objects and spatial targets, has been studied very thoroughly in the prefrontal cortex, there has been little physiological research on the problem of simultaneously representing multiple items. However, related work has been done on the problem of representing entire sequences of events. Researchers have found single neurons in the prefrontal cortex that respond selectively to

particular sequences of objects or spatial locations (Shima and Tanji, 2000; Averbeck et al., 2003; Ninokura et al., 2003; Averbeck et al., 2006), as well as single neurons that encode the temporal position within a given sequence (Hahnloser et al., 2002; Fujii and Graybiel, 2003; Ninokura et al., 2004).

At this point in time, however, no one has yet attempted to examine the signal related to each item stored in memory independently. There are a number of interesting directions that this line of research could take. For example, does the delay period activity corresponding to one item change as more items are added to memory? If it does change, does it do so in a systematic way that can be predicted from the activity related to a single item? When several memories are stored at the same time, are they all stored at the same strength, or does this depend on serial order? These issues are very different from the question of whether or not sequence-selective cells exist.

Task-selective responses in the prefrontal cortex

Until this point we have solely discussed the mnemonic functions of the prefrontal cortex. However, this function is only a small part of its overall role. Over a century of research has established that the prefrontal cortex is essential in all tasks that require anything but the most rote, automatic behavior. It is a crucial component in the transformation of stimulus to response, and allows us to behave flexibly when context demands it, rather than mindlessly mapping the same stimulus to the same response every time. It also allows us to control our impulse to respond to a stimulus by inhibiting inappropriate behaviors; one of the hallmarks of prefrontal damage is the lack of impulse control.

One of the first hints that the prefrontal cortex was important for impulse control and the control of context-dependent behavior was the case of frontal damage in the patient Phineas Gage (Harlow, 1848, 1868). Gage recovered from the passage of an iron rod through his frontal cortex and exhibited what are now known as classic symptoms of prefrontal syndrome. Although previously calm and reliable, after the accident he was capricious, irreverent, and impulsive, exhibited inappropriate behavior, and was unable to hold a job. Similar symptoms have since been described in many patients suffering from prefrontal damage.

Subsequent studies have confirmed the importance of the prefrontal cortex in the appropriate mapping of stimulus to response. The Wisconsin Card Sorting Test was developed to probe this aspect of prefrontal function. In this task, subjects are required to follow a continually evolving rule instructing them how to sort cards into piles. Patients with prefrontal damage are inevitably unable to perform well on this task. They are typically able to learn the first card-pile association rule, but are unable to suppress this mapping when the rule changes.

Recent years have seen several intriguing neurophysiological studies that shed light on this aspect of prefrontal function. In particular, several studies have reported the existence of cells that encode rules, or are context- or task-selective (White and Wise, 1999; Asaad et al., 2000; Wallis et al., 2001; Wallis and Miller, 2003). These cells have the potential to form the core of a processing system that allows for the flexible control of behavior. At this point there is an overwhelming amount of evidence that implicates the prefrontal cortex as the central locus of executive control.

Interaction of task-selectivity and image-selectivity

Given that the prefrontal cortex is responsible for both the short term memory buffer and executive control, the natural question is whether context has an effect on the representation of items in short term memory, or if these two processes are completely independent. In particular, we are interested in determining whether the context within which a multi-item memory task is performed has an impact on how each item is represented. Also, it will be interesting to determine whether individual cells carry information about both the stimuli and the behavioral context, or whether separate networks of cells are responsible for carrying these pieces of information. These issues will be addressed.

Multi-item memory experiment

To address the questions we have outlined above, we designed an experiment to study the mechanism that the brain employs to represent multiple items. Monkeys were trained to remember two items presented sequentially at the fovea, and to release a lever when a matching sequence was seen. In this first set of experiments we found that the monkeys remembered the sequences of items at a high level of performance, as judged by their patterns of responses. We recorded neuronal activity from the prefrontal cortex while the monkeys were performing this task, and discovered that the majority of single cells encoded the identity of both items simultaneously, rather than the alternative possibility of a separate population of cells for each item. Also, the activity related to a given sequence of items was only partially predictable from each neuron's response to

individual items. Finally, the strength of coding of an individual item decayed as time progressed, leading to a stronger representation of more recently seen items.

Task-dependent multi-item memory experiment

In the second set of experiments we studied the representation of multiple items in memory under different behavioral contexts. In one version of the task, the monkeys were required to report their memory of the sequence with a bar release. In another version, their memory was reported with a sequence of two eye movements. In both versions of the task the sample stimulus to keep in memory was exactly the same; the only thing that differed was how the animal would act on this information. This experiment was designed to investigate the influence of behavioral context on the representation of multiple items in memory. Following training on the saccade task, prefrontal recordings confirmed that the behavioral report did indeed have a profound impact on the way sequences were coded. Instead of a decaying memory trace associated with each item, the older item now had a stronger representation than the newer item during the memory delay period.

The purpose of the third set of experiments was to determine how this behavioral report-dependent change of coding was manifested at the single cell level. For these experiments we trained monkeys to switch back and forth between the two types of behavioral report, bar-release and eye-movement. When we recorded neuronal activity from the prefrontal cortex, we found that single cells were capable of exhibiting both types of activity, and could flip their coding back and forth depending on which task was

being performed. We also found neurons that directly represented the task that the monkey was currently performing.

REFERENCES

- Amit DJ, Bernacchia A, Yakovlev V (2003) Multiple-object working memory--a model for behavioral performance. *Cereb Cortex* 13:435-443.
- Asaad WF, Rainer G, Miller EK (2000) Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 84:451-459.
- Averbeck BB, Sohn JW, Lee D (2006) Activity in prefrontal cortex during dynamic selection of action sequences. *Nat Neurosci* 9:276-282.
- Averbeck BB, Chafee MV, Crowe DA, Georgopoulos AP (2003) Neural activity in prefrontal cortex during copying geometrical shapes. I. Single cells encode shape, sequence, and metric parameters. *Exp Brain Res* 150:127-141.
- Blum RA (1952) Effects of subtotal lesions of frontal granular cortex on delayed reaction in monkeys. *AMA Arch Neurol Psychiatry* 67:375-386.
- Butters N, Pandya D, Sanders K, Dye P (1971) Behavioral deficits in monkeys after selective lesions within the middle third of sulcus principalis. *J Comp Physiol Psychol* 76:8-14.
- Callicott JH, Mattay VS, Bertolino A, Finn K, Coppola R, Frank JA, Goldberg TE, Weinberger DR (1999) Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cereb Cortex* 9:20-26.
- Fujii N, Graybiel AM (2003) Representation of action sequence boundaries by macaque prefrontal cortical neurons. *Science* 301:1246-1249.
- Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol* 61:331-349.
- Fuster JM (1973) Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *J Neurophysiol* 36:61-78.
- Fuster JM, Alexander GE (1970) Delayed response deficit by cryogenic depression of frontal cortex. *Brain Res* 20:85-90.

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

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

Gnadt JW, Bracewell RM, Andersen RA (1991) Sensorimotor transformation during eye movements to remembered visual targets. *Vision Res* 31:693-715.

Goldman PS, Rosvold HE (1970) Localization of function within the dorsolateral prefrontal cortex of the rhesus monkey. *Exp Neurol* 27:291-304.

Hahnloser RH, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419:65-70.

Harlow HF, Davis RT, Settlage PH, Meyer DR (1952) Analysis of frontal and posterior association syndromes in brain-damaged monkeys. *J Comp Physiol Psychol* 45:419-429.

Harlow JM (1848) Passage of an iron rod through the head. *Boston Medical and Surgical Journal* 39:389-393.

Harlow JM (1868) Recovery after severe injury to the head. *Publications of the Massachusetts Medical Society* 2:327-347.

Hebb DO (1949) *The Organization of Behaviour*. New York: John Wiley & Sons Inc.

Jensen O, Lisman JE (1996) Novel lists of 7 +/- 2 known items can be reliably stored in an oscillatory short-term memory network: interaction with long-term memory. *Learn Mem* 3:257-263.

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

Lashley KS (1951) The problem of serial order in behavior. In: *Cerebral mechanisms in behavior* (Jeffress LA, ed). New York: Wiley.

- Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390:279-281.
- Miller EK, Desimone R (1994) Parallel neuronal mechanisms for short-term memory. *Science* 263:520-522.
- Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16:5154-5167.
- Miller GA (1956) The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 63:81-97.
- Mishkin M (1957) Effects of small frontal lesions on delayed alternation in monkeys. *J Neurophysiol* 20:615-622.
- Mishkin M, Manning FJ (1978) Non-Spatial Memory after Selective Prefrontal Lesions in Monkeys. *Brain Research* 143:313-323.
- Miyashita Y, Chang HS (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* 331:68-70.
- Murdock BBJ (1962) The serial position effect in free recall. *Journal of Experimental Psychology* 64:482-488.
- Nickerson RS (1965) Short-Term Memory for Complex Meaningful Visual Configurations: A Demonstration of Capacity. *Can J Psychol* 19:155-160.
- Ninokura Y, Mushiake H, Tanji J (2003) Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *J Neurophysiol* 89:2868-2873.
- Ninokura Y, Mushiake H, Tanji J (2004) Integration of temporal order and object information in the monkey lateral prefrontal cortex. *J Neurophysiol* 91:555-560.
- Osaka M, Osaka N, Kondo H, Morishita M, Fukuyama H, Aso T, Shibasaki H (2003) The neural basis of individual differences in working memory capacity: an fMRI study. *Neuroimage* 18:789-797.

Passingham RE (1975) Delayed Matching in Rhesus-Monkeys with Selective Prefrontal Lesions. *Brain Research* 85:190-191.

Shima K, Tanji J (2000) Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *J Neurophysiol* 84:2148-2160.

Talmi D, Grady CL, Goshen-Gottstein Y, Moscovitch M (2005) Neuroimaging the serial position curve. A test of single-store versus dual-store models. *Psychol Sci* 16:716-723.

Tanaka S (2002) Multi-directional representation of spatial working memory in a model prefrontal cortical circuit. *Neurocomputing* 44-46:1001-1008.

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

Wallis JD, Miller EK (2003) From rule to response: neuronal processes in the premotor and prefrontal cortex. *J Neurophysiol* 90:1790-1806.

Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953-956.

White IM, Wise SP (1999) Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res* 126:315-335.

CHAPTER 2

Representation of Multiple Items in Prefrontal Delay Activity

INTRODUCTION

Although much is known about how neurons represent single items in working memory, relatively little is known about how neurons encode several items simultaneously. This is an important question, because in order to effectively interact with the world an animal must be able to maintain more than one item at a time in memory. Reasoning about cause and effect relationships requires this ability, as does learning about associations between items. Any number of demonstrations of this ability can be drawn from the spectrum of mental operations we are able to perform. For example, in order to add two numbers together we must be able to simultaneously hold in mind the identities of the two numbers we are adding, and, furthermore, we must also remember the rule (addition) that we are going to apply to the numbers. This requires multi-item memory.

We know that the primate prefrontal cortex (PFC) plays a critical role in the maintenance of items in working memory, as has been demonstrated by several lines of evidence. Decades of investigation have shown that lesioning or reversibly inactivating the lateral PFC in non-human primates causes deficits in performance on delayed-response tests (Mishkin, 1957; Gross and Weiskrantz, 1962; Fuster and Alexander, 1970; Goldman and Rosvold, 1970; Goldman et al., 1971; Passingham, 1975; Mishkin and Manning, 1978), and that comparable lesions in human subjects cause short-term memory deficits (Muller et al., 2002). Functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) have provided evidence that prefrontal regions in humans and monkeys are selectively activated in tasks utilizing working memory

(Jonides et al., 1993; Petrides et al., 1993; McCarthy et al., 1994; Swartz et al., 1995; Fiez et al., 1996; Leung et al., 2002; Inoue et al., 2004). Neurophysiological data from non-human primates has suggested the idea that working memory has as its neural basis the activity of single cells in the PFC. Beginning with the pioneering studies of the early 1970's, many experiments have shown that PFC cells are selectively activated during delay periods in tasks that require the maintenance in short-term memory of single items or spatial locations (Fuster and Alexander, 1971; Kubota and Niki, 1971; Fuster, 1973; Funahashi et al., 1989; Miller et al., 1996). The overwhelming weight of evidence tells us that the prefrontal cortex is the locus of short term memory, and it is therefore the logical place to record when looking for the networks of cells responsible for multi-item memory.

There is evidence from human behavioral studies that working memory has the capacity of several items, with estimates ranging from 3-4 items to "7 plus or minus 2" items, depending on the type of item to be stored (Miller, 1956; Luck and Vogel, 1997; Vogel and Machizawa, 2004). The subject of storing multiple memories has been investigated through human neuroimaging, particularly with respect to memory load (Braver et al., 1997), through computational modeling (Sompolinsky and Kanter, 1986; Jensen and Lisman, 1996; Amit et al., 2003), and through behavioral work in nonhuman primates (Sands and Wright, 1980; Swartz et al., 1991; Swartz et al., 2000; Orlov et al., 2002), but not yet through neurophysiology.

To address the questions we have outlined above, we designed an experiment to study the mechanism that the brain employs to represent multiple items. Monkeys were trained to remember two items presented sequentially at the fovea, and to release a lever

when a matching sequence was seen. In this set of experiments we found that the monkeys remembered both items at a high level of performance, as judged by their pattern of responses. We recorded neuronal activity from the prefrontal cortex while the monkeys were performing this task, and discovered that the majority of single cells encoded the identity of both items simultaneously, rather than the alternative possibility of a separate population of cells for each item. Also, the activity related to a given sequence of items was only partially predictable from each neuron's response to individual items.

EXPERIMENTAL DESIGN AND METHODS

Subjects

The subjects were two rhesus monkeys (*Macaca mulatta*), one male and one female, weighing 6.0 and 6.5 kg. Eye movements were monitored and stored using an infrared eye-tracking system (ISCAN, Burlington, MA). Using previously described methods, monkeys were implanted with recording chambers and with a head bolt to immobilize the head during neuronal recordings. The location of the recording chambers and the location of recording penetrations were determined by structural magnetic resonance imaging scans. Recording chambers were placed over the lateral prefrontal cortex, centered over the principal sulcus and anterior to the arcuate sulcus. All surgeries were performed under aseptic conditions while the animals were anesthetized with isoflurane. The animals received post-operative antibiotics and analgesics and were always handled in accord with NIH guidelines and the recommendations of the MIT Animal Care and Use Committee.

Bar-release sequence task

Monkeys performed a two-item image sequence memory task (delayed-match-to-sequence, Figure 2.1) that required them to judge if two successively presented sequences of two natural images were the same. The task was administered and behavior monitored by two computers running the “CORTEX” real-time control system (<http://www.cortex.salk.edu>). The trial began when the monkeys grasped a lever and fixated a small (0.15°) white spot at the center of a CRT screen. They were required to

maintain fixation within a $\pm 1.5^\circ$ square window around the fixation spot for the entire trial. After the initial 1,000 ms of fixation, an image was presented at the center of the screen for 500 ms. The image was then extinguished and was followed by a 1,000 ms memory delay (the one-item memory delay). A second image was then presented for 500 ms and was also followed by a 1,000 ms memory delay (the two-item memory delay). The presentation of these two images constituted the sample phase of the task, because the monkeys were required to remember both of these images throughout the duration of the trial. The sample phase was followed by the presentation of a temporally identical test sequence, again consisting of two images presented on the screen for 500 ms each, separated by a 1,000 ms delay. If the test sequence exactly matched the sample sequence, the monkeys were required to release the lever before 900 ms following the onset of the second test item in order to receive a juice reward. If the test sequence differed in any way from the original sample sequence (if either of the images was different, or if their order was reversed), the monkey was required to continue holding the lever until a second test sequence was presented. This second test sequence was always a match and thus required a lever release. As a result, a sequence judgment was only required for the first test sequence; the second test sequence was used so that a behavioral response would be required on every trial. This ensured that the monkeys were always paying attention. Note that with this design, the behavioral response (lever release) is not uniquely associated with a sequence (it was used to signal “match”, not a particular sequence) and, further, the monkeys could not predict whether the first test sequence would require a response. Thus any differential activity to the sample sequences could not be related to

the behavioral response. 50% of all trials were nonmatch trials, and 50% were match trials. A 1,000 ms inter-trial interval followed all trials.

For each recording session, four novel cue stimuli, never before seen by the animal, were chosen at random from a database of images (Corel, Ottawa, Canada). The stimuli were small complex images about $2^{\circ} \times 2^{\circ}$ in size. The images were presented on a computer screen positioned directly in front of the animal. We made no attempt to determine which features of particular images were responsible for the cells' responses; for this study, it was necessary only that different cues elicited selective activity from a number of PFC neurons. Complex images were used because they have been shown to elicit robust activity from lateral prefrontal neurons (Miller et al. 1996). Each of the four images had the same chance (25%) of appearing as the first cue and of appearing as the second cue (25%). All combinations of two images in sequence were used, including the four sequences composed of a single image shown twice, leading to a total of 16 sequences. The design was completely balanced, in that each possible first image was followed equally often by each possible second image. The converse was also true; each possible second image was preceded equally often by each possible first image. This allowed us to disambiguate the signals related to the first and second images, and to follow each signal independently throughout the course of the trial. This design ensures that if the second image simply erased the effects of the first image (as one might expect to find in a primary sensory area), the cell would show no selectivity for the first image during the latter phase of the trial. However, if activity related to the first image was still carried by the cell, this task design would allow us to extract that signal.

Three types of nonmatching test sequences were used to ensure that the monkeys were remembering the sequence correctly (Figure 2.2A). One type of nonmatch was that in which the first image changed and the second image remained the same. This nonmatch was used to ensure that the monkey remembered the first cue – it would be impossible to correctly respond to this type of trial if the monkey only remembered the second cue. The second type of nonmatch was a sequence in which the first image stayed the same but the second image changed. This was used to test the memory of the second image. The third type of nonmatch was that in which the same images were used, but they were presented in the reverse order. This type of nonmatch was used to ensure that the monkeys were remembering the images in the correct order. The monkeys performed well on all types of trials (Figure 2.2B; first cue 91% correct; second cue 85% correct; order 95% correct; chance on all conditions was 50%), indicating that they were remembering both items and the order in which they were presented.

Recording technique

Electrode penetration sites (Figure 2.3A) were determined using magnetic resonance imaging (MRI) scans obtained prior to surgery. The recording chambers were positioned stereotaxically over the left lateral PFC of each animal such that the principal sulcus and lateral prefrontal cortex were readily accessible (Figure 2.3A).

Monkeys were seated in primate chairs within sound-attenuating enclosures (Crist Instruments, Damascus, MD). Their heads were restrained, and a juice spout was placed at their mouths for automated reward delivery. Recordings were made using arrays of eight independently moveable dura-puncturing tungsten microelectrodes (FHC

Instruments, Bowdoinham, ME). The electrodes were advanced using custom-made screw-driven mini-microdrives (Nichols et al. 1998) mounted on a plastic grid (Crist Instruments, Damascus, MD) with 1-mm spacing between adjacent locations. Neuronal activity was amplified, filtered, and stored for off-line sorting into individual neuron records (Plexon Systems, Dallas, TX). We did not prescreen neurons for task-related activity such as visual responsiveness or stimulus selectivity. Rather, we randomly selected neurons for study by advancing each electrode until the activity of one or more neurons was well isolated, and then began data collection. This procedure was used to ensure an unbiased estimate of prefrontal activity. In any given session, we were able to simultaneously record the activity of up to 12 individual neurons (an average of 5.8 per recording session).

Analysis of neural data

Data were analyzed using custom-written routines in MATLAB (Mathworks, Natick, MA). Trials were divided into five epochs for the analysis of neural activity. The ‘fixation’ epoch consisted of the 500 ms immediately preceding stimulus onset. The ‘first cue’ epoch began 100 ms after the onset of the first cue and had a duration of 400 ms. The first 100 ms were excluded to compensate for the minimum latency of visual responses in PF cortex. The ‘first delay’ epoch started 200 ms after the offset of the first cue and had a duration of 800 ms. Likewise, the ‘second cue’ epoch started 100 ms after the onset of the second cue and had a duration of 400 ms, and the ‘first delay’ period started 200 ms after the offset of the first cue and had a duration of 800 ms. These

epochs were chosen for simplicity. The results reported here were insensitive to the exact time windows used.

To assess the effect of each of the two cues on neural activity, a two-way ANOVA was performed for each cell on the activity from each epoch. A significant effect of the first or second cue stimulus means that activity varied significantly with the identity of the first or second cue during the analysis epoch. If the effect of one of the cues on neural activity depended on the identity of the other cue, this would produce a significant interaction between cues. All ANOVAs were evaluated at $p \leq 0.05$. All neural activity histograms were calculated with a resolution of 1 ms, and then smoothed with a boxcar window.

To generate the normalized data in Figures 2.4, 2.5, and 2.8, we divided the firing rate obtained with a particular image during an epoch by the average firing rate during that epoch, which transformed the mean firing rate during that epoch to 1. This was in order to be able to compare epochs with very different firing rates. We did this for each cell. To generate the population averages in Figure 2.5 we averaged all the cells together.

The purpose of the response surface analyses shown in Figures 2.6-2.7 was to predict the multi-item activity of a neuron from its response to single items. We regressed the second-delay activity of a neuron on a two-factor linear model using the firing rates driven by each item in isolation as the regressors. First we determined the average firing rate of a cell for each cue during its presentation. This data represents the cell's response to a single item. We used these values to create a two dimensional grid of 16 points, one point corresponding to each combination of two items. The grid was created in this way because our hypothesis was that the combination of two images would

be linear, and if so, this is the most convenient way of visualizing the data. We then plot the actual firing rate of the cell (normalized as described above for single items) in response to each combination of two items as a point floating above this grid. All points together create a response surface which, assuming an additive model, should approximate a plane. In our regression model, the tilt of the plane and its offset were unconstrained to allow differential weighting of each item. All the cells used in the regression analyses were selective for the first image during the first cue period, the second image during the second cue period, and both images during the second delay period. We used this population because the purpose of the analysis was to use the firing rates observed during the cue periods to predict the response during the second delay period. All models were fit using least-squares.

In order to produce figure 2.7 the data was transformed to a zero to one scale by fixing the lowest firing rate at zero and the highest firing rate at one, with the two other rates linearly scaled. This normalization measure was used because we wanted a metric of selectivity that would have the same range for all cells so that we could produce a meaningful average. The response rates were similarly transformed. In the individual cell response surface plots the values between the data points were linearly interpolated.

RESULTS

Visual responsiveness

A total of 222 lateral prefrontal neurons were recorded from the left hemispheres of two monkeys during performance of the two-image sequence task (121 from *monkey A*, 101 from *monkey S*). Most of the neurons showed a significant change in activity relative to baseline activity, during one or more of the task epochs (206/222 or 92.8%, 112 from *monkey A* and 94 from *monkey S*; two-tailed *t* tests, evaluated at $P < 0.05$). In any single epoch, many neurons were responsive (128/222 or 57.7% during the first cue period; 150/222 or 67.6% during the first delay period; 159/222 or 71.6% during the second cue period; and 142/222 or 64.0% during the second delay period).

Selectivity for single images

To identify single neurons whose activity varied with the cue images, a two-factor ANOVA (one factor for each cue, evaluated at $P \leq 0.05$) was performed on the average activity of each neuron over each epoch (see Methods). A majority of neurons (163/222, 73.4%, 100 and 63 in *monkeys A* and *S*, respectively) showed activity that varied significantly with the identity of at least one of the images during at least one trial epoch. Table 1 shows the incidence of selectivity for each epoch; from one-third to one-half of all neurons showed selectivity during a given epoch.

Simultaneous selectivity for both images during the second delay period

A main interest was to determine whether information about each of the two images was maintained in a separate population of PFC neurons or, instead, whether the memories of both images were somehow combined on the single-neuron level. We found examples of both, but the majority of image-selective neurons exhibited activity that depended on the identity of both images held in memory.

During the second, two-item, delay, over half of the recorded neurons (132/222, 59.5%) showed activity that varied significantly with one or both cue images. There was no obvious topography for neurons selective for the first or the second image (Figure 2.3A). The majority of these neurons (78/132, 59.1%) showed activity that depended on both images ('two-image' neurons). Fewer neurons were selective for only the first (17/132, 12.9%) or only the second (37/132, 28.0%) image during the second delay (Figure 2.3B). Some two-image neurons (48/78, 61.5%) showed initial selectivity for the first image during the cue presentation or the first delay, which was subsequently modulated by selective activity for the second image following its presentation, resulting in activity driven by both images during the second delay period. Other two-image neurons (30/78, 38.5%) only began to show selectivity following the appearance of the second cue image; they did not show any selectivity during and immediately after presentation of the first image in the sequence (the first cue and first delay epochs, ANOVA, $P > 0.05$). These neurons, therefore, required the presence of both images in memory in order to show any selective activity, and are most likely examples of the sequence neurons that have been previously reported (Averbeck et al., 2003; Ninokura et al., 2003).

The relationship of two-image activity to individual-image activity

Given that the activity of the majority of single PFC neurons reflected both images, we naturally wondered whether this ‘two-image’ activity bore any simple relationship to the neural activity elicited by the individual images in isolation. For example, was a neuron’s activity for the two images a simple addition of its activity associated with each single image? The following analyses will address this issue. At this point, it is important to note one of the crucial features of our experimental design: image presentation was balanced. That is, across a set of trials in which a given image was a cue, it was followed (when it was cue 1) and preceded (when it was cue 2) by each and every image with equal frequency. Thus, when we sort trials by a given image as say, the first cue, any neural selectivity seen is directly attributable to that image because the influence of the images used as the other cue is factored out across trials.

A single neuron with image-selective activity is shown in Figure 2.4. When the trials are grouped according to the identity of the first cue (Figure 2.4A), activity during its presentation (i.e. during the first cue epoch) is strongest for a particular image (image ‘D’). This selective activity is maintained through the first delay period immediately after the first cue presentation. This pattern of activity is revealed in a plot of the neuron’s average, normalized activity associated with each image when it was used as the first cue during both the first cue presentation and the first delay (Figure 2.4B; see METHODS for details). The fact that these curves are the same shape indicates that the neuron’s image preferences during cue presentation and during the subsequent delay period are very similar. Likewise, when we sort the trials according to the identity of the second cue (Figure 2.4C), it is apparent that this neuron’s activity is also strongest for

image D when it appears as the second cue during both the second cue period and the subsequent delay period. The corresponding average activity plot is similar to that for the first cue (data not shown). But note this neuron's pattern of selectivity during the second delay period when the trials are grouped according to the identity of the first cue (Figures 2.4A and D). Its activity varies with the identity of the first cue but in quite a different way than seen earlier in the trial. During delay 2, rather than image D eliciting a relatively high degree of activity as it did during delay 1, the activity corresponding to image D is now lower than when other images were the first cue. This effect is seen clearly when the average activity associated with each first cue image is plotted during cue 1 and compared with the first cue activity during delay 2 (Figure 2.4D). The relatively high level of activity to image D as cue 1 does not simply carry over into the second delay or add to the activity elicited by cue 2 in a straightforward way, at least in this example cell. Again, because image presentation is balanced in our experimental design, the activity differences seen in the second delay are directly attributable to the first cue image; any image-selectivity directly attributable to the second cue is factored out.

This change in first cue-driven activity following the addition of a second cue to memory did not simply reflect differences in neural representation during sensory stimulation (cue presentation) versus memory (delay epochs). Indeed, as can be seen in Figure 2.4A and 2.4B, this neuron's image preference during the delay epochs was virtually identical to its image preference during the immediately preceding cue. This was also true across the entire population of neurons. Figure 2.5A shows the average normalized activity of all neurons selective for the first image during the both the first

cue epoch and following delay. The concept behind this figure is the same as that for Figure 2.4A, except that Figure 2.4A was for an individual neuron, whereas Figure 2.5A is the average of the population of image-selective neurons. This figure clearly indicates that, at a population level, relative image preferences established during cue presentation are maintained into the following delay period.

We performed a regression analysis to further examine the correspondence between cue-1 related activity during cue presentation and the subsequent delay (Figure 2.5B). We fit a linear model to each cell ($\text{response} = \alpha + \beta * \text{cue1}$), using its activity during the first cue period as the regressor and the activity during the first delay period as the response (see METHODS). A positive β indicates that the response of the cell during the first delay period varies directly with its response during cue presentation. All β s were tested for a significant difference from zero (two-tailed t test, $p \leq 0.05$). Most cells in the population had β s greater than zero, and the distribution of β s was significantly greater than zero (one-tailed t test, $p \leq 0.05$). These results lend support to the conclusion that image preferences during the first delay period are very similar to image preferences during cue presentation.

Across the population of neurons, there were a variety of changes in neural selectivity as a result of adding a second cue to memory; some neurons invert their preferences (as did the neuron of Figure 2.4), some changed in a different non-obvious fashion, a few maintained their preference. This is reflected in Figure 2.5C, which shows the average normalized activity across all neurons that showed significant image selectivity for the first cue during both the first cue and the second delay periods. While the average activity during presentation of the first cue shows clear selectivity for that cue

(the data are sorted by this cue), the average activities during the second delay for the images appearing as that cue are relatively equal - the line is flat - when sorted by the identity of the first cue. This does not mean that the image selectivity has disappeared; rather, the averaging of different neurons with different changes in selectivity due to addition of a second cue has created a population average that is flat. It is worthwhile noting here that the proportion of neurons selective for the first cue image is roughly comparable during the first and second delay periods (36.9% in the first delay period and 42.8% in the second delay period, Table 1), and only cells that were selective for images during this delay period went into this analysis, so we are sure that the selectivity for the first cue has not disappeared – it has just changed form.

In order to gain a better understanding of how two images are simultaneously represented in one population of neurons, we attempted to predict the delay period activity of cells driven by both images from the activity of the cells driven by a single image. To do this, we fit a linear response surface model to the data (see METHODS for details) using the activity driven by each cue in isolation as the regressors:

$$\text{response} = \alpha + \beta_1 * \text{cue1} + \beta_2 * \text{cue2}.$$

This model describes a plane in three dimensional space (an example plane is shown in Figure 2.6A); the height of the plane at each point is the predicted response. α , β_1 , and β_2 are the offset and two slope parameters determined by the regression; cue1 is the response to the image used as cue 1 during its presentation; cue2 is the response to the image used as cue 2 during its presentation. This is perhaps the most intuitively plausible

model; it would simply require the more recent memory to be linearly combined with the original memory. This model was free to vary in three parameters (α , β_1 , and β_2), and therefore was capable of finding the best weighted sum of the two cues, not just a straightforward addition. We fit all of the image-selective cells in the population (see METHODS for details of cell selection) using this model, and discovered that this model did a fair job at describing the cells' actual responses. Based on the R^2 values, the model explained on average about 10% of the variance of the activity of these cells; there were some cells that were very well fit by the model, with an R^2 approaching 80%. Obviously there is a lot of variation in the activity of these cells that is not explained by the presence of the two cues in memory. This is probably a result of the fact that prefrontal cells are capable of carrying information about several task-relevant pieces of information simultaneously, including the contingencies of the task at hand. This issue will be discussed more completely in Chapter 3.

When we fit each cell to this model, we found that the distribution of β_1 s was centered on zero (Figure 2.6B), and that the mean of the distribution was not significantly greater than zero (one-tailed t test, $p > 0.05$). There are many cells with significant slopes, as would be expected given that all of the cells show image selectivity for the first cue during this delay period. However, there are just as many negative slopes as positive slopes, indicating that these cells have changed their image preferences as a result of the addition of the second item to memory. This result is consistent with that shown in Figure 2.5C, and indicates that, while the memory of the first image is preserved after the addition of a second item to memory, it has changed form.

The mean of the distribution of β_{2s} , on the other hand, is significantly greater than zero (one-tailed t test, $p \leq 0.05$). There are many cells with significant slopes, but most of these cells have a positive slope. This result is also consistent with that discussed previously, and indicates that the delay activity representing the second cue is very similar in image preference to the activity seen during second cue presentation. These results taken together suggest that it is only partially possible to predict the two-item delay activity from the activity driven by cues in isolation. The two-item delay activity is most similar to that of the most recently represented cue, and is very different from that of older cues.

Although this model has given us a good sense of how these cells are encoding two items in memory, it is informative to look at actual response surfaces obtained from individual cells without fitting them to a model. To do this, we first transformed the activity of the cues in isolation and the second delay period activity to a 0-1 scale. This made it easier to compare the response surfaces of different cells. The axes are as before, but the surface is now the actual average firing rates for each combination of two images. Examples of actual response surfaces for two individual cells are shown in Figure 2.7A and B. Figure 2.7A shows a cell that has a maximal response when the best first item and the best second item are used as the sequence of items. This is the type of cell we would have expected to see if the cells were simply adding their responses to each item to produce a combined response. The cell shown in Figure 2.7B is interesting because it shows the best possible response when the worst first image is used in combination with the best second image. These cells are both representative of the population effect. They both exhibit the maximal firing rate for the previously determined ‘best’ second image.

Also, their preference for which item was presented first is variable; one cell prefers the original best first image, while the other cell has changed its first-image preference.

We computed the population average response surface simply by averaging all the individual response surfaces together (Figure 2.7C). Only cells selective for both images during the second delay were used in this analysis. The net surface shows an interesting characteristic: there is a large net positive slope in the Cue 2 direction, and no net slope in the Cue 1 direction. This means that the representation of the second cue during the second delay period is quite faithful to the representation we would find if it was presented in isolation. Also, since the selectivity profiles of the cells for the first cue have been changed so dramatically in each of the cells, when they are averaged together they produce a surface with zero tilt. This means that the first cue is not represented faithfully at all. These results are completely consistent with those previously presented. They indicate that the newest item in memory is represented like it would be if it were the only item in memory, while the representation of the older item has changed significantly.

As a further test of the assertion that image-selectivity changes in a non-obvious fashion with the addition of a new cue image to memory, we plotted the normalized response (see METHODS) of each neuron to its best first image during the presentation of the first cue versus its activity to that same image during either the first delay (black dots and line, Figure 2.8A) or during the second delay (black dots and line, Figure 2.8B). As a point of reference, we also plotted the response of each neuron to its best first cue image during the presentation of the first cue versus its activity to the best image as determined by the activity during that delay period (large grey dots). It is important to

remember that these may not be the same image – the best image in the first case was always the best image defined during the presentation of the first cue, while the best image in the second case was defined independently during each epoch. The purpose of the grey dots was to provide a comparison of how correlated across epochs each neuron’s activity might be if we defined each neuron’s preference independently for each epoch; it essentially provides a “best case scenario” for how well correlated activity could be across the epochs.

Figure 2.8A shows the correlation between the response to each best first image during the first cue and the first delay periods. A high response to the best image during cue presentation is correlated with a high response to the same image during the first delay period, indicating that image preferences are maintained during this delay period. The same effect with approximately the same slope is seen when the best image is determined independently for each epoch, giving strength to this argument. Also, there is a lot of overlap between the grey dots and the black dots in this figure, indicating that the same image was most-preferred across the two epochs. Overall, 54% of neurons prefer the same image during the first delay period as during the first cue period, which is an impressive alignment, given that the four images were chosen randomly, and firing rates to different images may be very similar. Figure 2.8B shows a comparison between the first cue period and the second delay period. Here, a different result is seen. The response to the best image during the first cue period is not correlated with its activity to that same image during the second delay period. This is to be expected if the image rankings are not being preserved during the second delay period. It should also be noted that in this figure there is not a lot of overlap between the grey dots and the black dots,

indicating that the preferred image during the first cue period is rarely the preferred image during the second delay period (in our sample, the two images were the same for 25% of the neurons, which is what would be expected by chance). Again, because image presentation is balanced, these analyses indicate that the activity related to the first cue image is not preserved in a straightforward way across addition of a second image to memory.

DISCUSSION

We have found that when two images held in short-term memory, they are both reflected in the activity of single prefrontal neurons. These findings are consistent with the hypothesis that multiple items are stored in working memory using a single population of neurons, and less compatible with an alternative model that posits that separate memories are stored in separate ‘boxes’ in the brain analogous to addresses in computer memory. Further, we found that there was not a straightforward relationship between the neural activity corresponding to a single item and neural activity after a second image was added to memory. The addition of the second item had a profound impact on the representation of the first, in most cases changing the preferred first item of the cell. However, the representation of the second item was very similar to its representation when presented in isolation. Also, we found two different categories of neurons: a large number of ‘superposition’ neurons that showed selectivity for single images in isolation as well as for two images in memory together, and a relatively small number of ‘sequence’ neurons that only began to show selective activity after the second image was presented.

These findings support those models that suggest that short-term memories of multiple items overlap in neural populations. We will discuss these results in relation to recent developments in the theoretical modeling of multiple memories and place them in the context of prior neurophysiological studies.

Relationship to prior neurophysiological studies

Recently, there have been several studies that have investigated the representation of sequences of items in the prefrontal cortex. Typically in these studies an animal is sequentially presented with a number of items to remember, and the activity related to the entire sequence is studied during the pre-choice delay period. These studies have found that single neurons selective for specific sequences of images, spatial locations, and movement sequences (Barone and Joseph 1989; Ninokura et al. 2003; Shima and Tanji 2000). Several studies have also reported the existence of neurons selective for rank order, a signal which may be used in the creation of sequence-selective neurons (Averbeck et al. 2002; Carpenter et al. 1999; Ninokura et al. 2004). In order to respond highly selectively to specific sequences, the neurons found in these studies are necessarily combining several memories in a non-straightforward fashion, which is consistent with the results reported here; if all possible sequences are to be represented, it is necessary that a cell change its image preference for at least one of the items. Our results extend these studies by addressing the question of how the memory trace for a single item is modified when new items are loaded into memory. We have shown here that the majority of neurons with dual image selectivity are created from neurons with ordinary image-selective delay activity. These neurons do not just fire selectively after a particular sequence is seen; they fire selectively after a single image is seen, and then change their firing pattern to accommodate an additional image.

Comparison to computational studies and models of multi-item working memory

The ability to store multiple items in memory has been explored through several different computational lines of research, and has led to the creation of neural network models that are able to store more than one item in memory. There have been two general classes of model created to address these issues, and they differ in whether storage is defined as the encoding of multiple items through the long-term modification of synaptic weights (an analog of long-term memory), or as the simultaneous activation of several delay circuits (an analog of short-term memory). The first class is typified by the Amari-Hopfield type of model (Amari 1972; Hopfield 1982), composed of a network of binary neurons, which can store a number of memories in the synaptic matrix. This type of model is particularly useful with respect to long-term memory storage of multiple memories, but does not address the issue of several memories simultaneously stored in a short-term buffer (as represented by delay activity), the issue under investigation in this study.

The second class of models, however, directly speaks to these issues, and we will therefore give a broad overview of these models and how the current results fit into this context. Several groups have approached this problem, and although the specific systems under study are quite different, the solutions that they propose are strikingly similar.

Amit and colleagues have created a network of excitatory and inhibitory integrate-and-fire neurons that is capable of simultaneous delay activity for up to six images in working memory (Amit et al. 2003; Yakovlev et al. 2004). Each image is represented by a distinct population of neurons, each with a “perfectly sharp” tuning curve; in other words, if a

neuron responds selectively to a particular image, it has a complete lack of response to all other images. Since this model assumes that a population of neurons corresponds to each image, multiple images are represented by the activation of multiple, corresponding populations of neurons. This model is perfectly additive, since memories for different images do not directly interfere with each other.

Another approach has been that of Tanaka, who created a neural network for the simultaneous representation of several spatial targets (Tanaka 2002a, b). This is also an integrate-and-fire network, and it relies on a topographically organized spatial map with hills of activity representing the memory of particular spatial locations. Although in this model the neurons are not perfectly sharply tuned, as they are in the Amit model, the same effect is realized because the model represents multiple spatial memories with non-overlapping hills of activity. This type of network does a very good job at storing multiple memories when they are distant on the spatial map, but breaks down when the memories are close enough in space to interfere with each other. Multiple memories can only be stored with high accuracy by completely separate subpopulations within these networks; when memory traces begin to overlap, fidelity is compromised. This model, as a result, is also completely additive.

These two models are similar in that they store memories for different images using distinct populations of neurons. They are also similar in that new memories are added to the network in an additive way. The results that we have reported in this chapter indicate that it will be necessary to tweak models such as these to incorporate the idea that, in most cases, single neurons respond simultaneously to more than one item. Prefrontal cells rarely have a strong response to one image and a baseline response to all

other images; in most cases, a spectrum of firing rates is observed for a large number of images. Given that this is the case, there will necessarily be an overlap between representations of different items. This is difficult to reconcile with models that require completely separate populations of neurons for each image/spatial location. Curti, Amit and colleagues (Curti et al. 2004) have begun to address these issues through the creation of a more realistic spiking network model that incorporates neurons that respond selectively for more than one image.

In conclusion, we have shown that the primate lateral prefrontal cortex exhibits signals related to the maintenance of multiple items in working memory. We have found that a single population of neurons is capable of coding two images, and that the signal related to a newer image is overlaid on the signal related to an older image, in the process dramatically changing the older signal. These two signals are combined in a partially predictable way. It remains to be shown how each memory can be reliably read out and reconstructed based on such a population code.

FIGURES

Figure 2.1. Behavioral task. The monkey was presented with a sequence of two images, which consisted of one sample cue, an intervening delay period, a second sample cue, and a second delay period. This was followed by the presentation of a test sequence which had the same temporal structure as the first. If this test sequence matched the sample sequence, the monkey was rewarded for releasing a lever during the presentation of the second matching test cue. If the test sequence was not an exact match, the monkey was required to continue grasping the lever until a match sequence appeared. A match sequence always appeared immediately following a nonmatch test sequence. See METHODS for further information.

Figure 2.2. Test trial types. *A*: Three types of nonmatching sequences were used to ensure that the monkey was correctly remembering the entire sequence. The sample sequence shown in the top row is followed by a test sequence that has a nonmatching first cue and a matching second cue. This type of sequence tested the monkeys' memory of the first cue – it would be impossible to respond correctly on this type of trial if only the second cue was being held in memory. The sample sequence shown in the middle row is followed by a test sequence that has a matching first cue and a nonmatching second cue. This type of sequence was used to test the monkeys' memory of the second cue. The sample sequence shown in the bottom row was followed by a sequence composed of the same cues, but presented in the reverse order. This type of sequence was used to test the monkeys' memory for the cue order. *B*: Behavioral performance. The monkeys

performed well on all three types of test sequence. The percent correct for each type of test sequence is shown; error bars represent the 95% confidence interval around the mean. The accuracy rate was 91% for the first condition (first cue), 85% for the second condition (second cue), and 95% for the third condition (order). Chance performance was 50% for each condition.

Figure 2.3. *A:* Anatomical location of recording sites and image-selective neurons in both monkeys. X and O, recording sites at which neurons selective for image 1 or image 2 during the second delay period were found, respectively. Black •, locations at which no image selective neurons were encountered. Multiple neurons were recorded at many locations. There was no obvious topography to task-related neurons. *B:* Relative proportions of neurons selective for only image 1, both images, or only image 2 during the second delay period. Area is to scale.

Figure 2.4. *A:* Activity of a single prefrontal cell, trials grouped according to which item appeared as the first cue. *B:* Normalized response of this cell to the first cue during both the first cue period and the first delay period (see METHODS). Curves with similar slopes indicate a similar selectivity profile in both epochs. *C:* Activity of the same cell, now grouped according to which item appeared as the second cue. *D:* Normalized response of the cell during both the first cue period and the second delay period for the first cue. Dissimilar curves indicate that the cells selectivity profile for the first cue has changed with the addition of a second item to memory.

Figure 2.5. *A:* The same analysis shown in Figure 2.4B, averaged over all cells in the population. *B:* The histogram shows the distribution of β across the population of image-selective cells. Colored bars are cells with significant β , $p \leq 0.05$. The mean of the distribution was significantly greater than zero, $p \leq 0.05$. *C:* The same analysis shown in Figure 2.4D, averaged over all image-selective cells in the population.

Figure 2.6. *A:* An additive model was used to fit the cell's response to all combinations of two items. An example of one possible response surface is shown to illustrate the model. *B:* The histogram shows the distribution of the values of β_1 obtained when every image selective cell was fit with the above model. Colored bars are significant values of β , $p \leq 0.05$. The mean of the distribution was not significantly greater than zero, $p > 0.05$. *C:* The distribution of the values of β_2 . The mean of the distribution was significantly greater than zero ($p \leq 0.05$).

Figure 2.7. *A and B:* Two actual response surfaces from two cells are shown. *C:* The average response surface of all cells in the population. A response surface was created for each image-selective cell in the population, as illustrated in Figure 2.7A and B. These surfaces were averaged together to produce an average response surface.

Figure 2.8. *A:* The correlation between the response to the best first image during the first cue and the first delay periods. Black dots, response of each neuron to its best image during cue 1 versus its response to that same image during the first delay period. Black line, linear fit to these points. Grey dots, response of each neuron to its best image during

cue 1 versus its response to its best image during the first delay period. Grey line, linear fit to these points. 54% of these neurons retain their image preference into the first delay period. *B*: The correlation between the response to the best first image during the first cue and the second delay periods. Black dots, response of each neuron to its best image during cue 1 versus its response to that same image during the second delay period. Black line, linear fit to these points. Grey dots, response of each neuron to its best image during cue 1 versus its response to its best image during the second delay period. Grey line, linear fit to these points. 25% of these neurons retain their image preference into the second delay period.

FIGURE 2.1

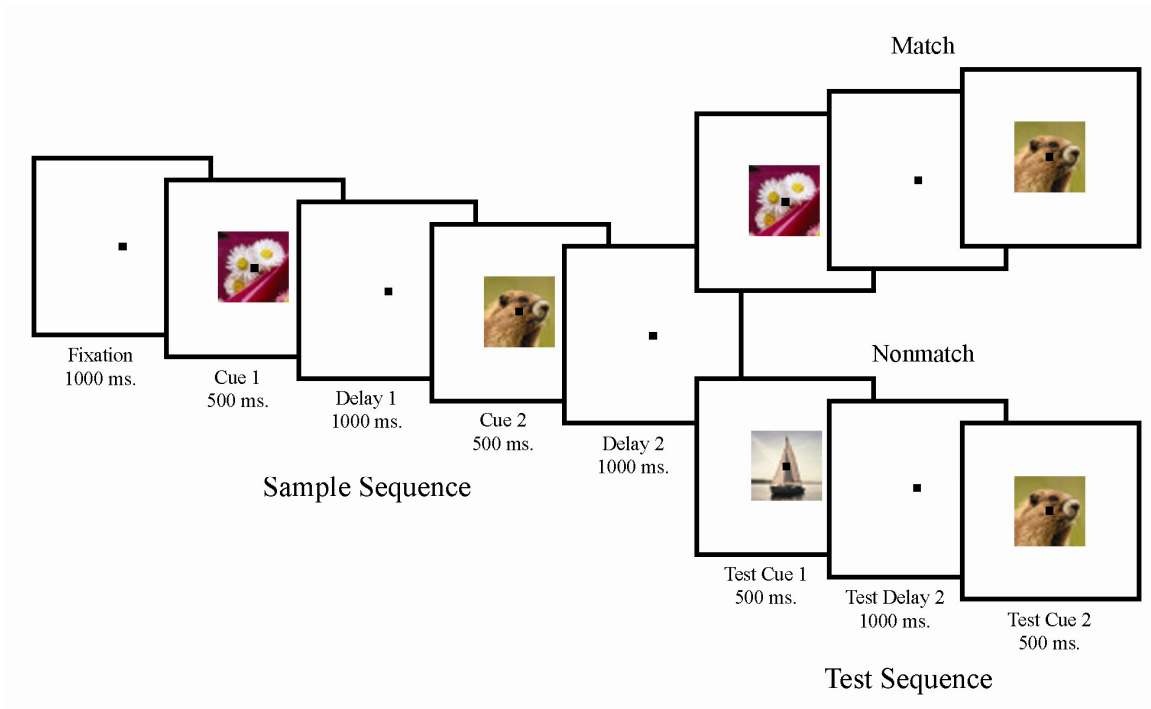


FIGURE 2.2

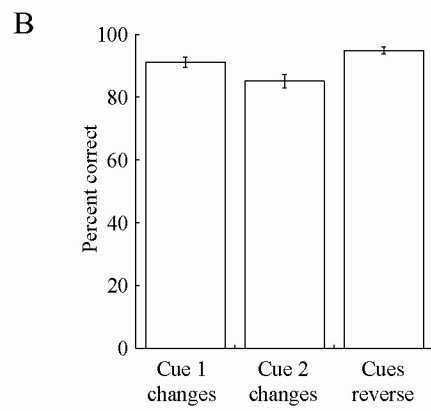
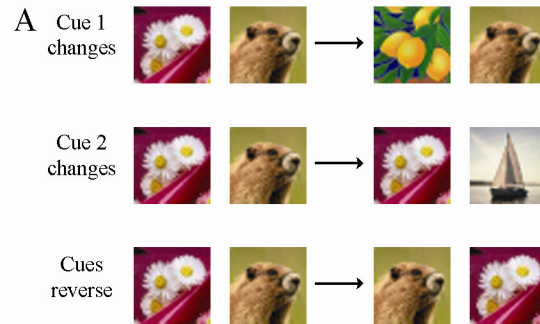


FIGURE 2.3

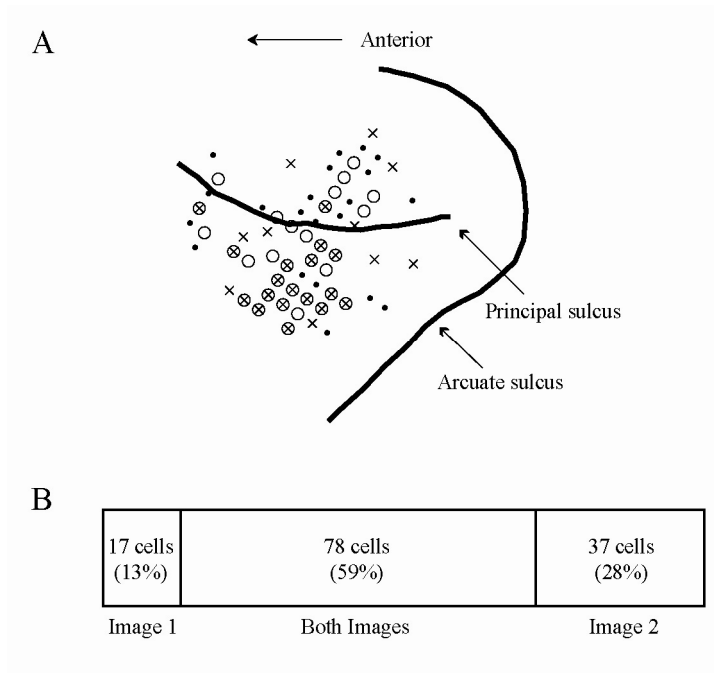


FIGURE 2.4

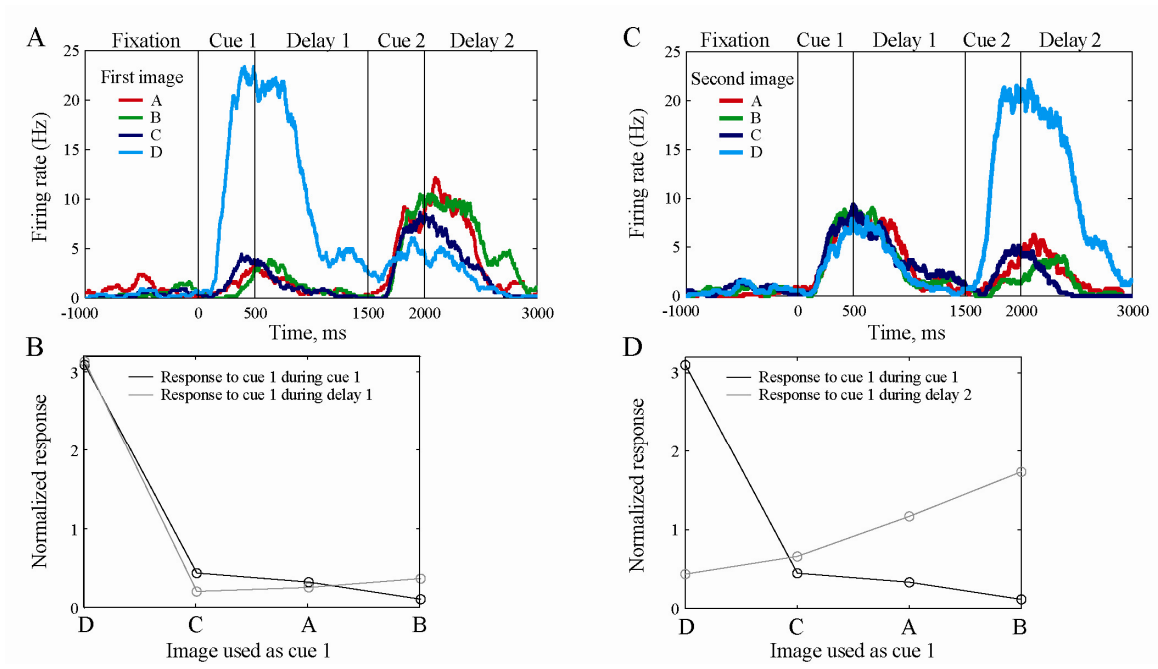


FIGURE 2.5

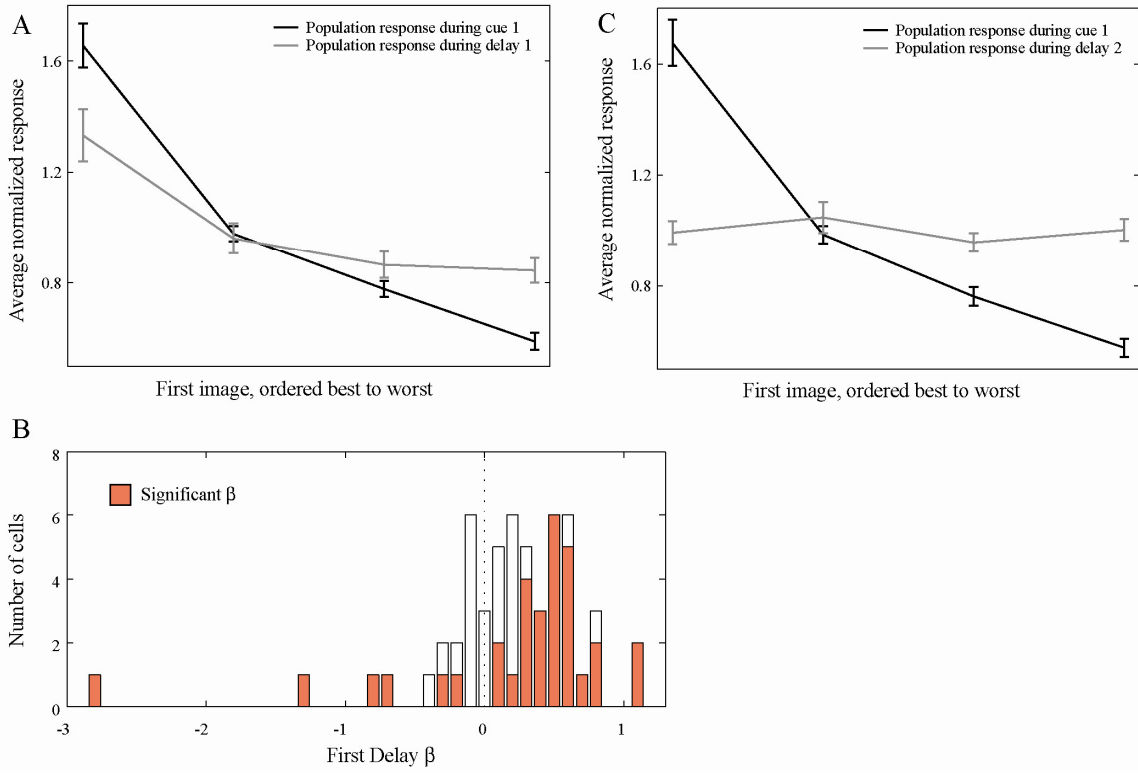


FIGURE 2.6

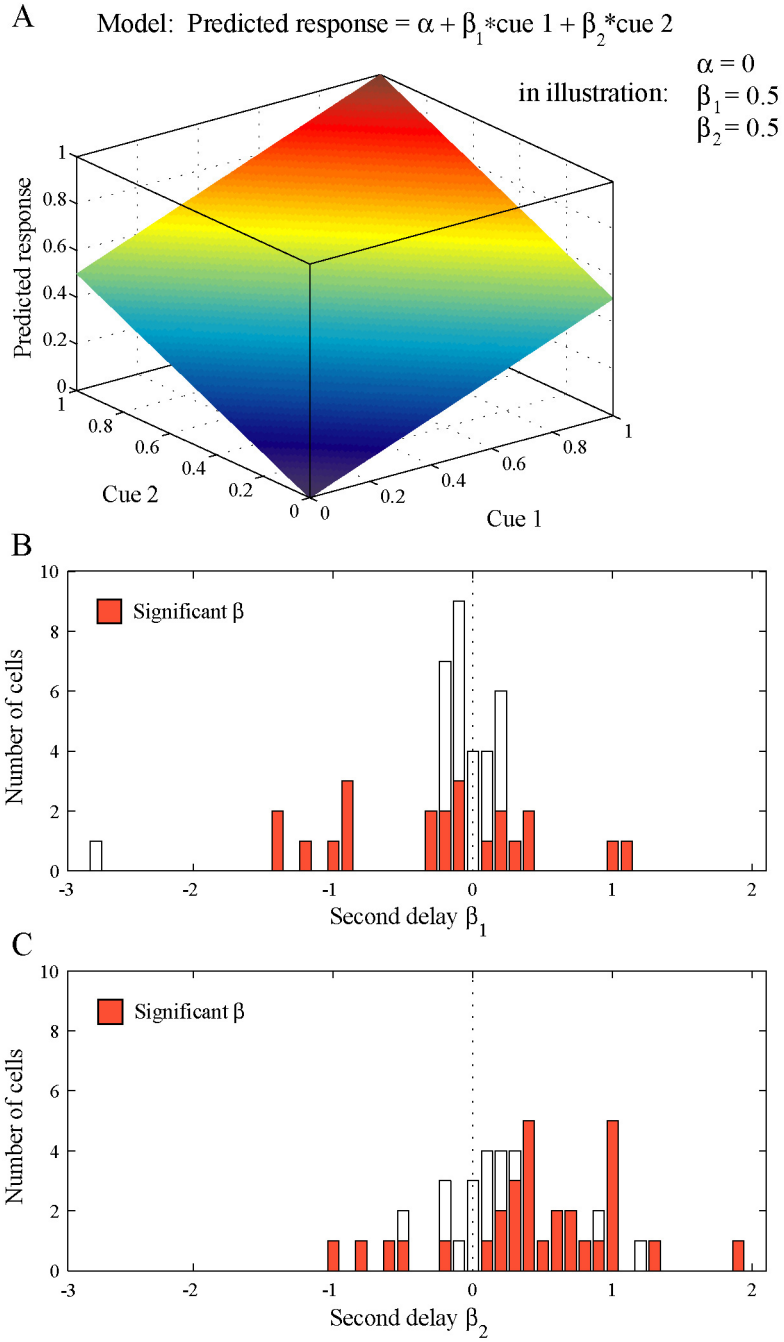


FIGURE 2.7

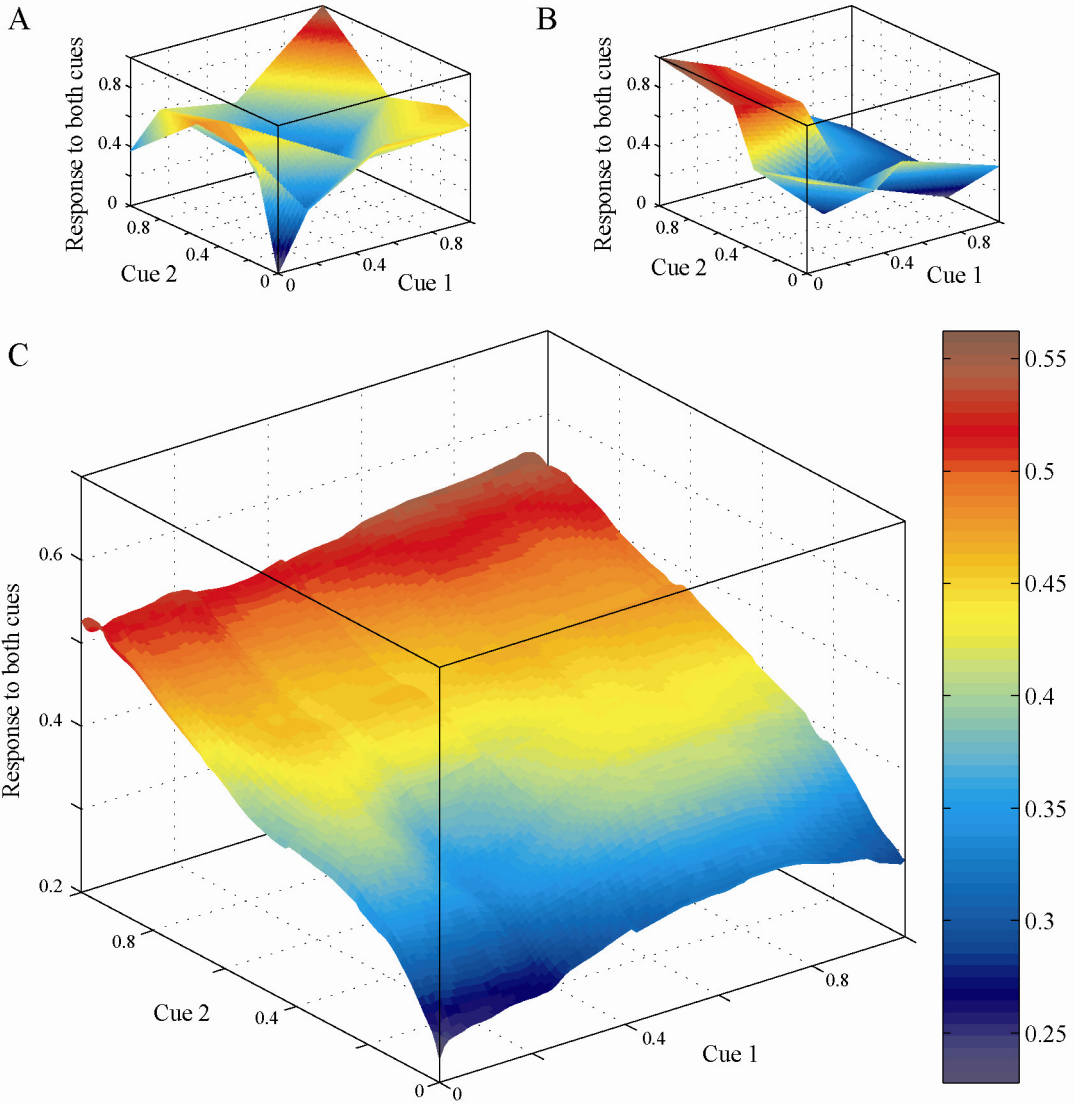


FIGURE 2.8

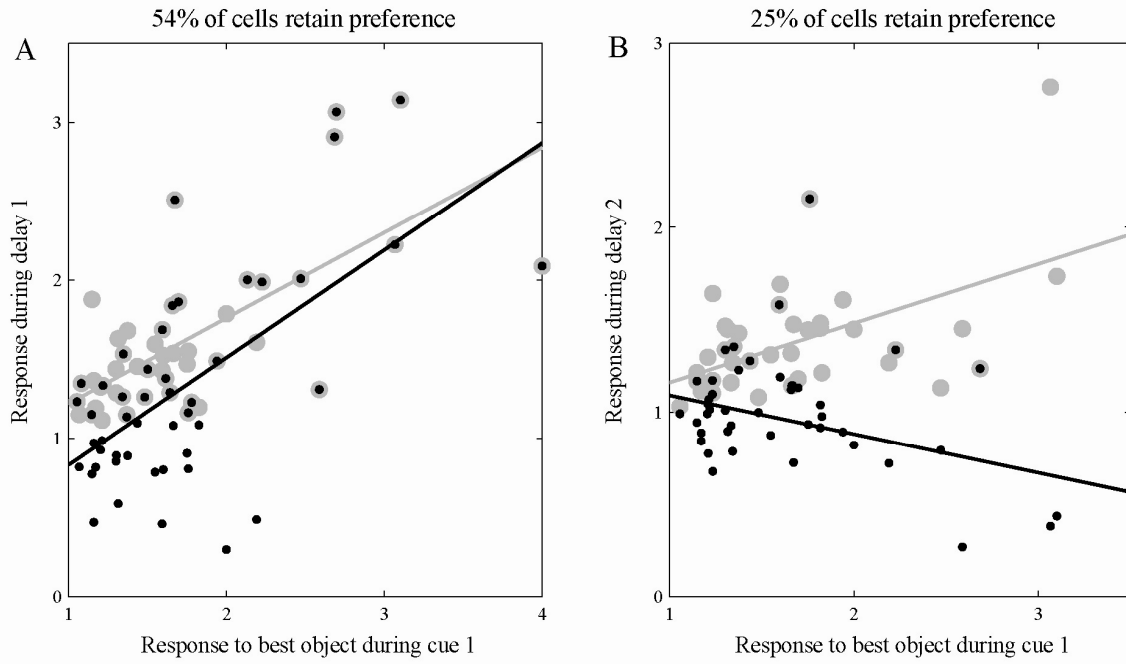


TABLE 2.1

TABLE 1. *Percentage of neurons that were selective for the first or second object that was presented, during either the first cue, the first delay, the second cue, or the second delay epoch.*

	Epoch				
	First Cue	First Delay	Second Cue	Second Delay	Any
First object selectivity					
Neurons showing effect	76	82	81	95	144
Percentage of 222	34.2	36.9	36.5	42.8	64.9
Second object selectivity					
Neurons showing effect			96	115	135
Percentage of 222			43.2	51.8	60.8
Selectivity for either object					
Neurons showing effect			107	132	163
Percentage of 222			48.2	59.5	73.4
Selectivity for both objects					
Neurons showing effect			70	78	116
Percentage of 222			31.5	35.1	52.3
First object X Second object					
Neurons showing effect			60	63	95
Percentage of 222			27.0	28.4	42.8

REFERENCES

- Amit DJ, Bernacchia A, Yakovlev V (2003) Multiple-object working memory--a model for behavioral performance. *Cereb Cortex* 13:435-443.
- Averbeck BB, Chafee MV, Crowe DA, Georgopoulos AP (2003) Neural activity in prefrontal cortex during copying geometrical shapes. I. Single cells encode shape, sequence, and metric parameters. *Exp Brain Res* 150:127-141.
- Braver TS, Cohen JD, Nystrom LE, Jonides J, Smith EE, Noll DC (1997) A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage* 5:49-62.
- Fiez JA, Raife EA, Balota DA, Schwarz JP, Raichle ME, Petersen SE (1996) A positron emission tomography study of the short-term maintenance of verbal information. *J Neurosci* 16:808-822.
- Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol* 61:331-349.
- Fuster JM (1973) Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *J Neurophysiol* 36:61-78.
- Fuster JM, Alexander GE (1970) Delayed response deficit by cryogenic depression of frontal cortex. *Brain Res* 20:85-90.
- Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. *Science* 173:652-654.
- Goldman PS, Rosvold HE (1970) Localization of function within the dorsolateral prefrontal cortex of the rhesus monkey. *Exp Neurol* 27:291-304.
- Goldman PS, Rosvold HE, Vest B, Galkin TW (1971) Analysis of the delayed-alternation deficit produced by dorsolateral prefrontal lesions in the rhesus monkey. *J Comp Physiol Psychol* 77:212-220.

- Gross CG, Weiskrantz L (1962) Evidence for dissociation of impairment on auditory discrimination and delayed response following lateral frontal lesions in monkeys. *Exp Neurol* 5:453-476.
- Inoue M, Mikami A, Ando I, Tsukada H (2004) Functional brain mapping of the macaque related to spatial working memory as revealed by PET. *Cereb Cortex* 14:106-119.
- Jensen O, Lisman JE (1996) Novel lists of 7 +/- 2 known items can be reliably stored in an oscillatory short-term memory network: interaction with long-term memory. *Learn Mem* 3:257-263.
- Jonides J, Smith EE, Koeppe RA, Awh E, Minoshima S, Mintun MA (1993) Spatial working memory in humans as revealed by PET. *Nature* 363:623-625.
- Kubota K, Niki H (1971) Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J Neurophysiol* 34:337-347.
- Leung HC, Gore JC, Goldman-Rakic PS (2002) Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *J Cogn Neurosci* 14:659-671.
- Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390:279-281.
- McCarthy G, Blamire AM, Puce A, Nobre AC, Bloch G, Hyder F, Goldman-Rakic P, Shulman RG (1994) Functional magnetic resonance imaging of human prefrontal cortex activation during a spatial working memory task. *Proc Natl Acad Sci U S A* 91:8690-8694.
- Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16:5154-5167.
- Miller GA (1956) The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 63:81-97.
- Mishkin M (1957) Effects of small frontal lesions on delayed alternation in monkeys. *J Neurophysiol* 20:615-622.

- Mishkin M, Manning FJ (1978) Non-Spatial Memory after Selective Prefrontal Lesions in Monkeys. *Brain Research* 143:313-323.
- Muller NG, Machado L, Knight RT (2002) Contributions of subregions of the prefrontal cortex to working memory: Evidence from brain lesions in humans. *Journal of Cognitive Neuroscience* 14:673-686.
- Ninokura Y, Mushiake H, Tanji J (2003) Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *J Neurophysiol* 89:2868-2873.
- Orlov T, Yakovlev V, Amit D, Hochstein S, Zohary E (2002) Serial memory strategies in macaque monkeys: behavioral and theoretical aspects. *Cereb Cortex* 12:306-317.
- Passingham RE (1975) Delayed Matching in Rhesus-Monkeys with Selective Prefrontal Lesions. *Brain Research* 85:190-191.
- Petrides M, Alivisatos B, Meyer E, Evans AC (1993) Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proc Natl Acad Sci U S A* 90:878-882.
- Sands SF, Wright AA (1980) Primate memory: retention of serial list items by a rhesus monkey. *Science* 209:938-940.
- Sompolinsky H, Kanter II (1986) Temporal association in asymmetric neural networks. *Physical Review Letters* 57:2861-2864.
- Swartz BE, Halgren E, Fuster JM, Simpkins E, Gee M, Mandelkern M (1995) Cortical metabolic activation in humans during a visual memory task. *Cereb Cortex* 5:205-214.
- Swartz KB, Chen SF, Terrace HS (1991) Serial learning by rhesus monkeys: I. Acquisition and retention of multiple four-item lists. *J Exp Psychol Anim Behav Process* 17:396-410.
- Swartz KB, Chen S, Terrace HS (2000) Serial learning by rhesus monkeys: II. Learning four-item lists by trial and error. *J Exp Psychol Anim Behav Process* 26:274-285.

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

CHAPTER 3

Task-Specific Representation of Multiple Items in the Primate Prefrontal Cortex

INTRODUCTION

One of the main functions of the prefrontal cortex is to guide the transformation of a stimulus into a behavioral response (White and Wise, 1999; Asaad et al., 2000; Miller and Cohen, 2001). This transformation is highly dependent on the behavioral context, or the “rules of the game”, and is therefore quite flexible. It is often the case that a stimulus might be associated with several different responses, depending on context. For example, a ringing phone might trigger different behaviors depending on the environment; at home, the appropriate response to this stimulus would be to answer the phone, while this response at a restaurant would be inappropriate.

Damage to the prefrontal cortex in both humans and monkeys produces impaired behavior that is consistent with the idea presented above. An example of this is the effect of prefrontal damage on patients performing the Wisconsin Card Sorting Task, a task which assesses the ability to adapt to changing context. In this task, subjects are required to sort cards into different piles; which pile a card ends up in is dependent on the rule currently in effect, which changes periodically. Therefore, a given card is associated with more than one pile during the course of an experiment, and the subject must keep the current rule in mind to perform well on this task. Humans are impaired on this task when they have sustained damage to the prefrontal cortex (Milner, 1963), and monkeys are similarly impaired on simpler analogs of this task (Dias et al., 1997).

There is also evidence that this function of the prefrontal cortex has a neurophysiological correlate. Neural activity corresponding to a given stimulus has been shown to vary depending on behavioral context, such as the upcoming motor response

(Asaad et al., 1998), which part of a stimulus is attended (Rainer et al., 1998), what type of task the animal is performing (Asaad et al., 2000), and which rule the animal is using to solve a task (Wallis et al., 2001).

An idea proposed by Goodale and Milner (Goodale and Milner, 1992) meshes nicely with the results of the studies discussed above. The primate neocortex has traditionally been divided into two separate processing streams, dorsal cortex being responsible for spatial vision, and ventral cortex responsible for object vision (Ungerleider and Mishkin, 1982). Goodale and Milner proposed an alternative division between the dorsal and ventral streams, based on observations from their own research. In their new model, the division would instead be between vision-for-action, handled by dorsal cortex, and vision-for-perception, handled by ventral cortex. Vision-for-action, thought to be the more primitive system, would be responsible for tasks such as catching a ball, or putting a card into a slot – tasks that involve quick coordination between what you see and what you do. The vision-for-perception system, on the other hand, thought to be a more recent development, would perhaps allow us to perceive events in the world, think about them, and decide what to do in a more leisurely way. There have been many studies over the past decade that have lent support to this hypothesis. For example, patients with damage to their dorsal cortex can have difficulty producing accurate grasping or orienting movements toward visual objects, even while they can accurately describe these objects (Perenin and Vighetto, 1988). Similarly, patients with damage to their ventral cortex have problems describing objects that they are able to accurately grasp and orient towards (Milner et al., 1991).

The above distinction between two visual systems, one for action and one for perception, says something far more than first meets the eye. The authors are essentially arguing that it doesn't make sense for the brain to have a general purpose representation of a stimulus when it could potentially be used for so many different things, and the resulting possible actions shunted through different motor systems. It would perhaps be more efficient to have different representations of a stimulus that depend on behavioral context, or how the stimulus will be used:

As soon as we direct a motor act towards an object an entirely different set of constraints applies. We can no longer rely on the perception system's 'general purpose' representation... Directing a saccadic eye movement, for example, will demand different transformations of visual input to motor output from those required to direct a manual grasping movement. The former will involve coordinate systems centered on the retina and/or head, while the latter will involve shoulder and/or wrist centered coordinates. While it is theoretically possible that a highly sophisticated 'general-purpose' representation could accommodate such transformations, such a possibility seems unlikely and unnecessary. (Goodale and Humphrey, 1998)

What is the role of the prefrontal cortex in this alternative model of visual processing? Does it provide a space for a general purpose representation in short-term memory of the stimuli that seem most useful, stored there temporarily for the purpose of

directing upcoming actions? Or is the representation of items in the PFC dependent on how this information will be used and what actions will be taken?

In the what/where model, the prefrontal cortex is thought to be the cortical terminus of both the object and the spatial streams, and, in a rough sense, where the two streams finally converge. For example, single cells that represent the memory of both an object and a spatial location have been found in the PFC (Rao et al., 1997). It is clear that the PFC receives abundant input from both streams, and many of its cells are selective for the spatial and non-spatial characteristics of visual input, among other things. However, if one of the principal functions of the PFC is to produce context-appropriate behavior, as argued above, then perhaps it might also be interesting to examine how it represents these visual pieces of information when they are going to be used for different purposes. For example, it would be interesting to look, within the PFC, at the distinction between visual information used for action and visual information used for perception, or the distinction between visual information used to direct a saccadic eye movement and visual information used to guide a grasping movement. The PFC maps stimuli onto actions. It is also a short-term memory buffer. How do these functions interact?

One of the questions that we are particularly interested in is whether or not the prefrontal delay activity representing images is dependent on how they will be used. To address this question, we trained two monkeys to remember a sequence of two items and report this information in one of two ways. The first variant of the task used a bar release as the behavioral report, and required the monkeys to release a bar when they viewed a sequence of two items that matched the original sample sequence. This version was called the bar-release task. The second version, called the eye-movement task, required

that the monkeys report their memory of the sequence of items using a sequence of two saccadic eye movements to the matching items in an array of images.

The objective of this second set of experiments is to determine the effect of the rule (type of behavioral report) on the neural representation of the memory of a sequence of items. In particular, we are interested in how this change in task contingencies affects the strength of representation of each individual item. In previous work (Chapter 2) we investigated the issue of how multiple memories are encoded in prefrontal delay activity. In that study we focused on the issue of stimulus selectivity preferences independent of the notion of strength of stimulus encoding. In this study, we focus primarily on the strength of stimulus representation, not the particular images that are preferred by a given cell. We are particularly interested in investigating how the relative image strengths evolve over time depending on how and when the information is going to be used.

EXPERIMENTAL DESIGN AND METHODS

Subjects

The subjects were two rhesus monkeys, *Macaca mulatta*, one male and one female, weighing 6.0 and 6.5 kg. Eye movements were monitored and stored using an infrared eye-tracking system (ISCAN, Burlington, MA). Using previously described methods (Miller et al., 1993), monkeys were implanted with recording chambers and with a head bolt to immobilize the head during neuronal recordings. All surgeries were performed under aseptic conditions while the animals were anesthetized with isoflurane. The animals received post-operative antibiotics and analgesics and were always handled in accord with NIH guidelines and the recommendations of the MIT Animal Care and Use Committee.

Bar-release sequence task

Monkeys first performed a two-item image sequence memory task (delayed-match-to-sequence, Figure 3.1) that required them to judge if two successively presented sequences of two natural images were the same. The task was administered and behavior monitored by two computers running the “CORTEX” real-time control system (<http://www.cortex.salk.edu>). The trial began when the monkeys grasped a lever and fixated a small (0.15°) white spot at the center of a CRT screen. They were required to maintain gaze within a $\pm 1.5^\circ$ square window around the fixation spot for the entire trial. After an initial 1,000 ms fixation period, an image was presented at the center of the screen for 500 ms. The image was then extinguished and was followed by a 1,000 ms

memory delay (the one-item memory delay). A second image was then presented at the same position for 500 ms and was also followed by a 1,000 ms memory delay (the two-item memory delay). The presentation of this sequence of images constituted the sample phase of the task, because the monkeys were required to remember both of these images and their order throughout the duration of the trial. The sample phase was followed by the presentation of a temporally identical test sequence, again consisting of two images presented on the screen for 500 ms each, separated by a 1,000 ms delay. If the test sequence exactly matched the sample sequence, the monkeys were required to release the lever within the 900 ms following the onset of the second test item in order to receive a juice reward. If the test sequence differed in any way from the original sample sequence (if either of the images was different, or if their order was reversed), the monkey was required to continue holding the lever until a second test sequence was presented. This second test sequence was always a match and thus required a lever release. As a result, a sequence judgment was only required for the first test sequence; the second test sequence was used so that a behavioral response would be required on every trial. This ensured that the monkeys were always paying attention. Note that with this design, the behavioral response (lever release) is not uniquely associated with a sequence (it was used to signal “match”, not a particular sequence) and, further, the monkeys could not predict whether the first test sequence would require a response. Thus any differential activity to the sample sequences could not be related to the behavioral response. 50% of all trials were match trials, and 50% were nonmatch trials. A 1,000 ms inter-trial interval followed all trials.

For each recording session four novel cue stimuli, never before seen by the animal, were chosen at random from a database of images (Corel, Ottawa, Canada). The stimuli were small, complex images about 2° by 2° in size. The images were presented on a computer screen positioned directly in front of the animal. We made no attempt to determine which features of particular images were responsible for the cells' responses; for this study, it was necessary only that different cues elicited selective activity from a number of PFC neurons. Complex images were used because they have been shown to elicit robust activity from lateral prefrontal neurons (Miller et al. 1996). Each of the four images had a 25% chance of appearing as the first cue and a 25% chance of appearing as the second cue. All combinations of two images in sequence were used, including the four sequences composed of a single image shown twice, leading to a total of 16 sequences. The design was completely balanced, in that each possible first image was followed equally often by each possible second image. The converse was also true; each possible second image was preceded equally often by each possible first image. This allowed us to disambiguate the signals related to the first and second images, and to follow each signal independently throughout the course of the trial. This design ensured that if the second image simply erased the effects of the first image (as one might expect to find in a primary sensory area), the cell would show no selectivity for the first image during the latter phase of the trial. However, if activity related to the first image was still carried by the cell, this task design would allow us to extract that signal.

Three types of nonmatching test sequences were used to ensure that the monkeys were remembering the sequence correctly (Figure 3.2A). One type of nonmatch was that in which the first image changed and the second image remained the same. This

nonmatch was used to ensure that the monkey remembered the first cue – it would be impossible to correctly respond to this type of trial if the monkey only remembered the second cue. The second type of nonmatch was a sequence in which the first image stayed the same but the second image changed. This was used to test the memory of the second image. The third type of nonmatch was that in which the same images were used, but they were presented in the reverse order. This type of nonmatch was used to ensure that the monkeys were remembering the images in the correct order. The monkeys performed well on all types of trials (Figure 3.2B, first cue 91% correct; second cue 85% correct; order 95% correct; chance on all conditions was 50%), indicating that they were remembering both items and the order in which they were presented.

Eye-movement sequence task

After recordings were completed for the bar-release sequence task, training began on an alternative version of this task. In this version the monkeys responded to the presentation of the sample sequence with two sequential eye movements (Figure 3.3) instead of reporting their memory of the sequence with a bar release. The sample sequence was identical to that seen in the bar-release task - two images were presented sequentially on the fovea for 500 ms each with an intervening delay period of 1000 ms. After the second delay period, however, the structure of the trial was different. In this new version, an array of three images was presented, two of which had just been seen in the sample sequence. These images were presented in a triangle around the fixation point at an eccentricity of 5°. The monkey was required to saccade to the images that he had just seen in the order in which they were presented. Loose time constraints were in place;

the monkey had 2000 ms to initiate a saccade to the first item. However, once the eyes had left the fixation point, the monkey was required to reach the first item within 70 ms. This was to ensure that the monkey made a saccade directly to an image without any intervening saccades. The time constraints for the second saccade were identical. In this task repeated items were not allowed, unlike the bar-release task; this was because the monkey needed to saccade to two different items sequentially. Again, four novel objects were used each day, and each could appear at either position.

Performance on the eye-movement task was somewhat worse than performance on the bar-release task, largely due to the difficulty of quickly identifying two images in the periphery of the visual field without the chance to visually inspect them. The monkeys performed better both as the number of items in the choice grid decreased, and as they moved closer to the fovea. Chance performance was at 16.7%, and the monkeys performed significantly better than chance at 63% correct averaged over all recording sessions. The monkeys' memory of the first cue and the second cue were both significantly above chance (first cue was 75% correct, chance was 33%; second cue was 70% correct, chance was 33%; given that the first cue was correct, the second cue was 83% correct, chance was 50%). The monkeys' performance on cue order was also significantly above chance (given that both items were correct, order was 85% correct, chance was 50%). Only neural data from correct trials was used in all analyses. On average, 469 correct trials were performed each day. There was no spatial bias in responses; each two-saccade path was represented with equal frequency among all completed trials.

Switching task

During this task, the above two tasks were performed on the same day. Blocks of 100 trials of each type were used, and the monkey alternated between task types. 4 blocks (occasionally 3) of each type were performed each day. On two days, blocks of 250 trials were used, and the monkey performed 2 blocks of each type. There was no explicit cue for switching behavioral response; this was obvious from context. Each day a different task (bar-release or eye-movement) was chosen as the first block of the experiment. In this version, the repeated-item trials were omitted from the bar-release part of the task. This was to ensure that the sample portion of both tasks was exactly the same under both conditions, since these trials were necessarily not included in the eye-movement task.

Performance was good on both versions of the task. During the bar-release task the monkey performed an average of 264 correct trials each day, at a performance level of 95% correct, again performing each of the three trial types significantly above chance (test of first cue: 95% correct; test of second cue: 94% correct; test of order: 97% correct; chance for all was 50%). During the eye-movement task the monkey performed an average of 265 correct trials each day, at a performance level of 76% correct, significantly above the chance level of 16.7%. Performance on the first cue and the second cue was good (first cue was 84% correct, chance was 33%; second cue was 81% correct, chance was 33%; given that the first cue was correct, the second cue was 91% correct, chance was 50%), and performance on order was good (given that both items were correct, order was 92% correct, chance was 50%). All analyses only used neural data from correct trials.

Recording technique

Electrode penetration sites and the location of the recording chambers were determined using structural magnetic resonance imaging (MRI) scans obtained prior to surgery. The recording chambers were positioned stereotaxically over the lateral prefrontal cortex of each animal, anterior to the arcuate sulcus, such that the principal sulcus and lateral prefrontal cortex were readily accessible.

Monkeys were seated in primate chairs within sound-attenuating enclosures (Crist Instruments, Damascus, MD). Their heads were restrained, and a juice spout was placed at their mouths for automated reward delivery. Recordings were made using arrays of eight independently moveable dura-puncturing tungsten microelectrodes (FHC Instruments, Bowdoinham, ME). The electrodes were advanced using custom-made screw-driven mini-microdrives (Nichols et al., 1998) mounted on a plastic grid (Crist Instruments, Damascus, MD) with 1-mm spacing between adjacent locations. Neuronal activity was amplified, filtered, and stored for off-line sorting into individual neuron records (Plexon Systems, Dallas, TX). We did not prescreen neurons for task-related activity such as visual responsiveness or stimulus selectivity. Rather, we randomly selected neurons for study by advancing each electrode until the activity of one or more neurons was well isolated, and then began data collection. This procedure was used to ensure an unbiased estimate of prefrontal activity. In any given session, we were able to simultaneously record the activity of up to 12 individual neurons (an average of 5.8 per recording session).

Analysis of neural data

Data were analyzed using custom-written routines in MATLAB (Mathworks, Natick, MA). Trials were divided into five epochs for the analysis of neural activity. The ‘fixation’ epoch consisted of the 500 ms immediately preceding stimulus onset. The ‘first cue’ epoch began 100 ms after the onset of the first cue and had a duration of 400 ms. The first 100 ms were excluded to compensate for the minimum latency of visual responses in the prefrontal cortex. The ‘one-item delay’ or ‘first delay’ epoch started 200 ms after the offset of the first cue and had a duration of 800 ms. Likewise, the ‘second cue’ epoch started 100 ms after the onset of the second cue and had a duration of 400 ms, and the ‘two-item delay’ or ‘second delay’ period started 200 ms after the offset of the second cue and had a duration of 800 ms. These epochs were chosen for simplicity. The results reported here were insensitive to the exact time windows used.

To assess the effect of each of the two cues on neural activity, a two-way ANOVA was performed for each cell on the activity during each epoch. A significant effect of the first or second cue means that activity varied significantly with the identity of the first or second cue during the analysis epoch. If the effect of one of the cues on neural activity depended on the other cue, this would produce a significant interaction between cues. All ANOVAs were evaluated at $p \leq 0.05$. All neural activity histograms were calculated with a resolution of 1 ms, and then smoothed with a boxcar window.

Figures depicting neural selectivity were created by performing ANOVAs on a sliding 200 ms time window that moved forward every 20 ms. Simple-effects ANOVAs were done for these analyses instead of two-way ANOVAs because of the presence of a large amount of interaction between the first and second cues; over 80% of cells showed

significant interaction in a standard two-way ANOVA, which had the potential to obscure selective activity. The resultant sums of squares for each ANOVA were used to estimate the percentage of variance attributable to either the first or the second cue for each cell (Sokal and Rohlf, 1995) as a function of time. All cells were then averaged together, yielding a population estimate of the average percentage of variance explained by each cue. All cells contributed to the variance component figures shown in this paper, although repeating the analysis using only image-selective cells did not alter the pattern of the results. The only effect of this modification was to increase the overall percentage of variance explained.

RESULTS

Visual responsiveness, bar-release task

A total of 222 lateral prefrontal neurons were recorded from the left hemispheres of two monkeys during performance of the bar-release sequence task (121 from *monkey A*, 101 from *monkey S*). Most of the neurons showed a significant change in activity relative to baseline (fixation) activity during one or more task epochs (206/222 or 92.8%, 112 from *monkey A* and 94 from *monkey S*; two-tailed *t* tests, evaluated at $P < 0.05$). In any single epoch, many neurons were responsive (128/222 or 57.7% during the first cue period; 150/222 or 67.6% during the one-item delay period; 159/222 or 71.6% during the second cue period; and 142/222 or 64.0% during the two-item delay period).

Strength of image selectivity, bar-release task

An issue that has not yet been addressed is how strongly each of the two items is encoded in prefrontal delay activity; the previous study only dealt with relative image preferences. As a rough estimate of how strongly each item is encoded in each of the four major epochs (both cue and both delay periods), we can determine the number of cells showing significant selectivity for each image using a two-way ANOVA. This data is shown in Table 3.1, and here we see that during both delay periods approximately one third to one half of recorded cells show selectivity for the first cue. Likewise, a similar proportion of cells show selectivity for the second cue during the second delay period, although the second item is represented somewhat more strongly than the first.

This rough estimate does not give a very clear picture of how the strength of image selectivity evolves over time. We therefore looked at much smaller time bins across the extent of the trial, and calculated the percentage of the variance of the neural activity explained by the first cue or the second cue during each of these bins (see METHODS for further explanation). This allowed us to visualize how strongly the population of neurons encoded each image at each moment in time, and made it possible for us to compare the relative strengths of the representation of each of the two items in prefrontal delay activity. The result of this analysis (Figure 3.4) shows that the strength of the first cue increased shortly after it was seen by the monkey, and then decayed as time progressed. The strength of the second cue did roughly the same thing. If we examine the relative strengths of the two cues during the second delay period, we see that the second item, most recently seen, had a stronger representation than the first item, which was presented further in the past. Alternative measures of stimulus selectivity (mutual information and ROC analysis, among others, were tried) showed an identical pattern of results. Although it appears from this figure that the percentage of variance explained by each object is relatively low (peaking at about 5%), this quantity is an average across every cell in the population. If a selective subset of cells is used for this analysis, the shapes of the curves remain the same, but the overall percentage of variance explained is greater; up to 67% of the variance of individual cells was accounted for by one of the images.

Eye-movement task

After the monkeys had completed training and all recording on the bar-release task, we trained the monkeys on a variant of that task (Figure 3.3; see METHODS for details). In this new version, the eye-movement sequence task, we presented the monkey with a sample two-item sequence identical in structure to that seen in the bar-release task. The monkey was then shown a triangular array of three images, two of which had been seen during sample presentation. He was required to make a saccadic eye movement to each of the items that he had seen during sample presentation in the correct temporal order. It should be noted that the memory demands of this task were identical to those required by the bar-release task; in both tasks, the monkeys were required to remember the identity of both items and the order in which they were presented. However, in this task the behavioral report was destined to be routed through the saccadic system rather than through the motor system used to control grasping movements. We reasoned that this might have an impact on how the sequence of items was stored in short-term memory.

Visual responsiveness, eye-movement task

A total of 177 lateral prefrontal neurons were recorded from the left hemispheres of two monkeys during performance of the eye-movement sequence task (91 from *monkey A*, 86 from *monkey S*). Most of the neurons showed a significant change in activity relative to baseline (fixation) activity during one or more task epochs (158/177 or 89.3%, 83 from *monkey A* and 75 from *monkey S*; two-tailed *t* tests, evaluated at $P < 0.05$). In any single epoch, many neurons were responsive (97/177 or 54.8% during the

first cue period; 97/177 or 54.8% during the one-item delay period; 123/177 or 69.5% during the second cue period; and 106/177 or 59.9% during the two-item delay period).

Strength of image selectivity, eye-movement task

We repeated the image strength analysis previously performed on data from the bar-release task and found a profound difference in how strongly each item was encoded during the two-item memory delay period (Figure 3.5). During the eye-movement task the image seen earlier in the sample period was maintained with a much greater strength and did not decay over time, in contrast to the result found in the bar-release task. In fact, at the end of the second delay period the representation of the item seen first was actually stronger than the representation of the second, more recently seen item. This result is robust, and does not depend on the method used to calculate strength of stimulus selectivity.

Difference in relative strength of image representations

We wanted to quantify the relative strengths of the image representations, so we looked at the difference in the percentage of variance accounted for by each of the two items within each task. We begin with the bar-release task. For each cell recorded during this task we subtracted the percent variance explained by the second object from the percent variance explained by the first object. This led to a set of difference values associated with each point in time, one value per cell. We can take the average of these values to obtain the average difference in coding strength between the first and second images at each point in time for the bar-release task; this is the blue curve in Figure 3.6.

Analogously, we can compute the same difference curve for the eye-movement task; this is the red curve in Figure 3.6. When these curves are overlaid, we see that there is an obvious difference between tasks during the second, two-item delay period. At each point in time a t test ($p < 0.05$) was computed to assess whether the two tasks were significantly different. The two curves were found to be significantly different beginning during the second, two-item delay period. During all other epochs there was no difference in relative item strengths between the two tasks. This result supports our hypothesis that differences in behavioral report can affect how information is held in working memory.

Switching task

Having found that the behavioral report can affect relative memory strengths, we were interested to see how individual cells were impacted by this change. In order to do this, we trained a monkey to flip between the two tasks using blocks of 100-250 trials (see METHODS). There was no explicit cue indicating the switch; the response was obvious by the context. We blocked the trials so that the monkey could settle into one strategy during one type of behavioral report, and then change his strategy when the task changed. We recorded the activity of 137 single PFC neurons while the monkey alternated between the two tasks.

Visual responsiveness, switching task

A total of 132 lateral prefrontal neurons were recorded from the left hemisphere of one monkey during performance of the switching task. Most of the neurons showed a

significant change in activity relative to baseline (fixation) activity during one or more task epochs (123/132 or 93.2%, two-tailed t tests, evaluated at $P < 0.05$). In any single epoch, many neurons were responsive (87/132 or 65.9% during the first cue period; 72/132 or 54.5% during the one-item delay period; 90/132 or 68.2% during the second cue period; and 97/132 or 73.5% during the two-item delay period).

Strength of image selectivity, switching task

We first established that the behavioral report had an impact on relative memory strengths when the monkey was performing both tasks in one recording session; in this case, the same population of neurons was used for the determination of cue strength in each task, bar-release and eye-movement. We found that the neurons continued to show a difference in relative memory strengths (Figure 3.7A, B), lending strength to our prior conclusion. We observed a difference in relative memory strengths that was somewhat less than that which we observed in separate recording sessions, which was perhaps due to the fact that the monkeys only had 100 trials to adopt a behavioral strategy for each task. However, the difference between tasks was still significant, as seen when we compare the differences in relative memory strength between the two tasks (Figure 3.8).

We found that we could observe individual cells changing how they represented the items when the behavioral report was changed. Figure 3.9A shows the activity of a single prefrontal cell during the performance of the bar-release task. The trials are grouped according to the identity of the first cue. In this panel, we see that during the second delay period, when the animal is remembering both items, this cell is not showing any selectivity for the first item. However, during the eye-movement task (Figure 3.9B)

the same cell shows a very different profile of activity. Now, this cell is showing a great deal of selectivity for the first item during the second delay period. This single cell has a response pattern that is dependent on the task that the animal is performing, even though the stimulus presentation and memory demands are identical. This cell is one of those in the population that contribute to the effect that we observe.

We also found examples of cells that showed reduced selectivity for the second item during the eye-movement task. An example of such a cell is shown in Figure 3.9C and D. The trials are grouped according to the second cue. As seen in the first panel, this cell shows selectivity for the second item during the second delay period, but only during the bar-release task. During the eye-movement task, this selectivity disappears, as seen in the second panel. This cell also contributes to our population effect.

Task selectivity

A large fraction of cells in the population showed differential firing rates depending on which task was being performed, independent of the issue of image selectivity strength. These cells did not necessarily show image-selective activity, but did respond preferentially during the performance of one task. Four example cells are shown in Figure 3.10A-D. The cells in panels A and C showed a higher firing rate during the second delay period for the bar-release task, while the cells in panels B and D showed a higher firing rate during this period for the eye-movement task. These cells are typical of those found in the population.

Across the population, roughly the same number of cells showed a higher firing rate for each task (Figure 3.11). Each point in this figure represents the activity of a

single cell during the second delay period. The firing rate during the bar-release task is plotted against the firing rate during the eye-movement task. The data is plotted logarithmically due to the spread of the data and the large number of points at low firing rates. Overall, 54% of cells showed a significantly different response during the second delay period depending on the task the monkey was performing.

Of these task-selective cells, we found that 79% were also image selective. An example of such a cell is shown in Figure 3.12. Panel A demonstrates that this cell responds more strongly during the second delay period when the monkey is performing the eye-movement sequence task. Panel B shows the selectivity that the cell has for the first image during the eye movement task. We see that this cell is not only encoding the task that the monkey is performing, it is also encoding the image that the monkey is remembering. Panel C shows that this cell does not exhibit image coding when the animal is performing the bar-release sequence task. This result is very interesting, because it demonstrates that the mnemonic buffer and executive control functions of the prefrontal cortex are dependent on the same network of cells.

DISCUSSION

The results that we have obtained show that the mnemonic representation of stimuli in the prefrontal cortex is highly dependent on the way the information will eventually be used. When more than one image is simultaneously remembered, each item is represented in the neural activity as a memory trace of a certain strength. Our results indicate that the relative strengths of different memory traces can change depending on the context under which the animal is performing the task. During a task that involved a passive bar release upon the presentation of a matching sequence of images, the prefrontal cortex represented multiple images as decaying memory traces; items that had been seen further back in time were represented less strongly than more recently seen items. However, during a task that required an active eye movement to each of the items in the sequence, items in memory did not decay over time. Instead, the first item either was maintained at the same level as the more recent item, or even more strongly, depending on whether the two tasks were performed on the same day or in separate recording sessions. Individual cells in the population reflected these differences, and the strength with which a single cell encoded the first or the second item was highly dependent on which task the animal was performing. We also found a large population of cells that directly represented the specific task that the animal was performing, a result consistent with previous reports (Morton, 1968; White and Wise, 1999; Asaad et al., 2000; Wallis et al., 2001; Wallis and Miller, 2003), and further confirmation that the rules guiding behavior are strongly represented in the prefrontal cortex.

This result is interesting because it shows that there is no single, canonical way to store items in working memory; the storage depends on the context. This notion runs counter to the notion that single cells in the prefrontal cortex possess defined response properties that can be relied upon to accurately convey information about what an animal is remembering. The standard model of prefrontal cortex function posits that there is a separation between various components, in particular between the central executive and the various types of storage buffers, e.g. the visuospatial sketch pad and the phonological loop (Baddeley, 1986; Lie et al., 2006; Repovs and Baddeley, 2006). However, our results indicate that this boundary is blurry at best, and probably non-existent. Not only is the maintenance of multiple items in memory deeply affected by the task the animal is performing, we have found that 79% of the task-selective neurons are also selective for images. These results can only be interpreted to mean that executive processing and mnemonic functions are deeply intertwined in the prefrontal cortex, and any model that attempts to explain prefrontal function should recognize this fact.

Our results also shed light on one of the unsolved problems of neural coding: how does the brain represent temporal order? There have been many theories proposed to explain how this might be accomplished, but one of the more popular theories relies on using the differing strength of memory traces to represent order (Konorski, 1961; Morton, 1968; Hinrichs, 1970; Bugmann and Bapi, 2000). According to this theory, as each stimulus enters memory it is represented by a trace that increases until the stimulus disappears. At this point, the strength of the trace begins to decrease, and continues to do so as time progresses. The recency of an item could therefore be determined based on the relative strengths of the items in memory. Short term memory has also been

generally found to be primarily sustained by the prefrontal cortex. If this is the primary locus of short term memory, our results cast doubt on this theory of temporal coding. The strengths of the memory traces in the prefrontal cortex are evidently quite variable, given our data, and their relative order is highly dependent on what the animal will eventually do with the stored information. It seems that the relative strengths of each of the memory traces are not enough to determine the temporal order of item presentation. It is possible that relative stimulus order is maintained in a different brain structure through the use of trace strengths; however, others have reported that, at least in the hippocampus, it is unlikely that trace strength supports the memory for stimulus order (Fortin et al., 2002).

Together, these results indicate that the processing and the maintenance of information in the prefrontal cortex cannot be thought of as distinct entities. Rather, processing has a strong impact on maintenance, as demonstrated by its effect on relative memory strengths. Also, the cells that encode the items to be remembered are the very same cells that encode task identity. Context appears to shape many aspects of the prefrontal cortex, including the most basic mnemonic functions.

FIGURES

Figure 3.1. Bar-release task. The monkey was presented with a sample sequence of two images. This sequence consisted of one sample cue, an intervening delay period (the one-item delay), a second sample cue, and a second delay period (the two-item delay). The sample sequence was followed by the presentation of a test sequence which had the same temporal structure as the first. If this test sequence exactly matched the sample sequence, the monkey was rewarded for releasing a lever during the presentation of the second matching test cue. If the test sequence was not an exact match, the monkey was required to continue grasping the lever until a match sequence appeared. A match sequence always appeared immediately following a nonmatch test sequence. See METHODS for further information.

Figure 3.2. *A:* Three types of nonmatching sequences were used to ensure that the monkey was correctly remembering the entire sequence. The sample sequence shown in the top row was followed by a test sequence that had a nonmatching first cue and a matching second cue. This type of sequence tested the monkeys' memory of the first cue; it would be impossible to respond correctly on this type of trial if only the second cue was being held in memory. Analogously, the sample sequence shown in the middle row is followed by a test sequence that has a matching first cue and a nonmatching second cue. This type of sequence was used to test the monkeys' memory of the second cue. The sample sequence shown in the bottom row was followed by a sequence composed of the same cues, but presented in the reverse order. This type of sequence

was used to test the monkeys' memory for the cue order. *B*: Behavioral performance. The monkeys performed well on all three types of test sequence. The percent correct for each type of test sequence is shown; error bars represent the 95% confidence interval around the mean. The accuracy rate was 91% for the first condition (first cue), 85% for the second condition (second cue), and 95% for the third condition (order). Chance performance was 50% for each condition.

Figure 3.3. Eye-movement task. The monkey was presented with a sample sequence of two images identical to that shown in Figure 3.1-1. This sample sequence was immediately followed by the presentation of an array of three images around the fixation point, two of which had been shown as part of the sample sequence. The monkey was required to saccade to the two previously seen images in the correct order. See METHODS for further information.

Figure 3.4. Relative image strengths, bar-release task. Dark blue curve: percent variance explained by the first cue, averaged across the population of neurons. Light blue curve: percent variance explained by the second cue, averaged across the population of neurons. During this task, the population of neurons encoded the second cue more strongly than the first during the pre-test delay period.

Figure 3.5. Relative image strengths, eye-movement task. Dark red curve: percent variance explained by the first cue, averaged across the population of neurons. Light red curve: percent variance explained by the second cue, averaged across the population of

neurons. During this task, the population of neurons encoded the first cue more strongly than the second during the pre-test delay period, which is the opposite of the result found in the bar-release task.

Figure 3.6. Difference in relative image strengths, bar-release task vs. eye-movement task. Blue curve: percent variance explained by the first cue minus percent variance explained by the second cue, bar-release task. Red curve: percent variance explained by the first cue minus percent variance explained by the second cue, eye-movement task. Shaded grey areas are statistically significant, $p < 0.05$.

Figure 3.7. Blocks of bar-release and eye-movement task, interleaved, same neurons. *A*: Relative image strengths, bar-release task blocks. Dark blue curve: percent variance explained by the first cue, averaged across the population of neurons. Light blue curve: percent variance explained by the second cue, averaged across the population of neurons. During this task, the population of neurons encoded the second cue more strongly than the first during the pre-test delay period. *B*. Relative image strengths, eye-movement task blocks. Dark red curve: percent variance explained by the first cue, averaged across the population of neurons. Light red curve: percent variance explained by the second cue, averaged across the population of neurons. During this task, the population of neurons encoded the first cue as strongly as the second during the pre-test delay period.

Figure 3.8. Difference in relative image strengths, bar-release task vs. eye-movement task, interleaved blocks, same neurons. Blue curve: percent variance explained by the first cue minus percent variance explained by the second cue, bar-release task. Red curve: percent variance explained by the first cue minus percent variance explained by the second cue, eye-movement task. Shaded grey areas are statistically significant, $p \leq 0.05$.

Figure 3.9. Two single neuron examples. *A,B*: The first panel shows the activity of a cell during the bar-release task. Trials are grouped according to the identity of the first cue. The second panel shows the activity of the same cell during the eye-movement task. Trials are again grouped according to the identity of the first cue. This cell shows enhanced first-cue-related activity during the eye-movement task. *C,D*: The first panel shows the activity of a different cell during the bar-release task. Trials are grouped according to the identity of the second cue. The second panel shows the activity of the same cell during the eye-movement task. Trials are again grouped according to the identity of the second cue. This cell shows reduced second-cue-related activity during the eye-movement task.

Figure 3.10. Four example cells showing task-dependent differences in firing rate. Panels A and C show cells that show enhanced firing rate during the second delay period during the bar-release task. Panels B and D show cells that show enhanced firing rate during the second delay period during the eye-movement task.

Figure 3.11. Population distribution of cells showing task-dependent differences in firing rate. Each point in this figure represents the activity of one cell during the second delay period. Average firing rate during the bar-release task is plotted against average firing rate during the eye-movement task. Cells that show significantly different firing rates between tasks are plotted using + or x. 54% of cells showed a significant difference between tasks.

Figure 3.12. *A:* A single cell that responds differentially depending on which task is being performed. *B:* The same cell during the eye-movement task, with the trials grouped according to which image was presented as the first cue. This cell shows task-dependent and image-dependent coding. *C:* The same cell during the bar-movement task, with the trials grouped according to which image was presented as the first cue.

FIGURE 3.1

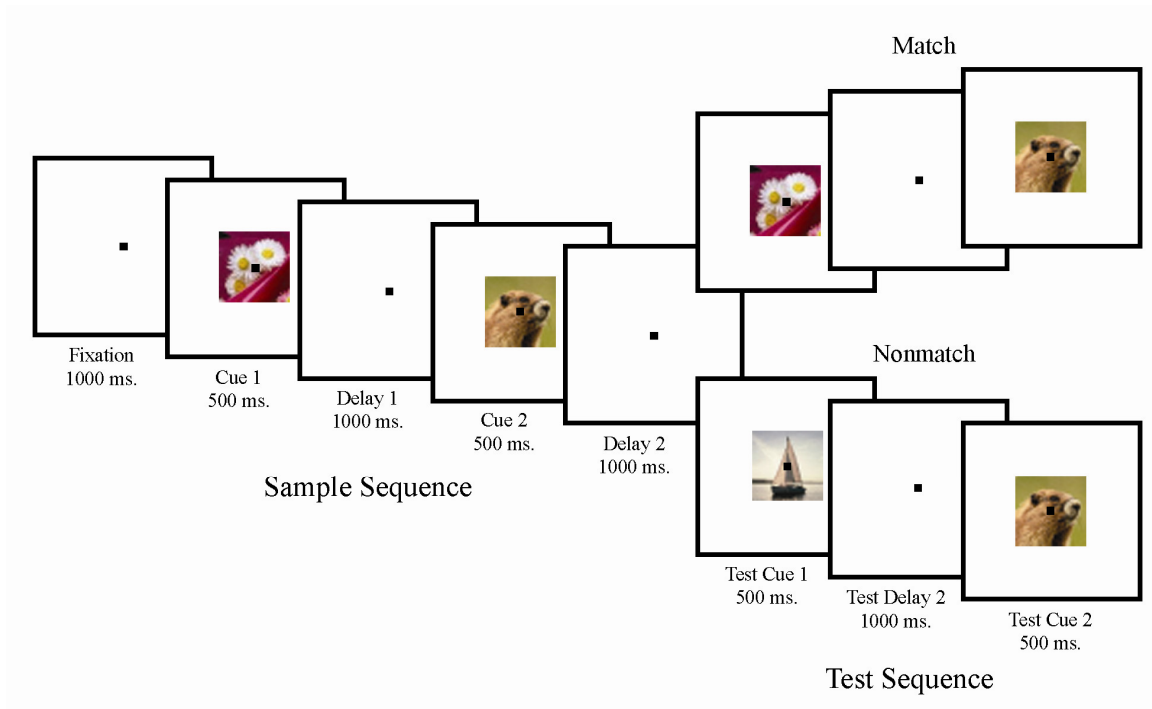


FIGURE 3.2

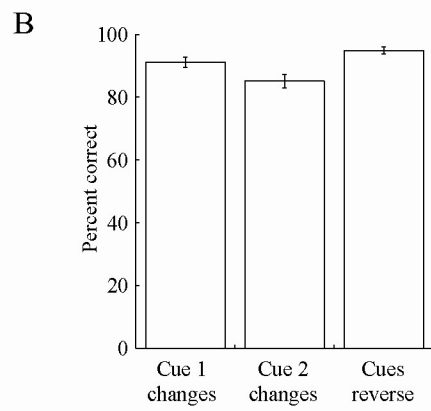


FIGURE 3.3

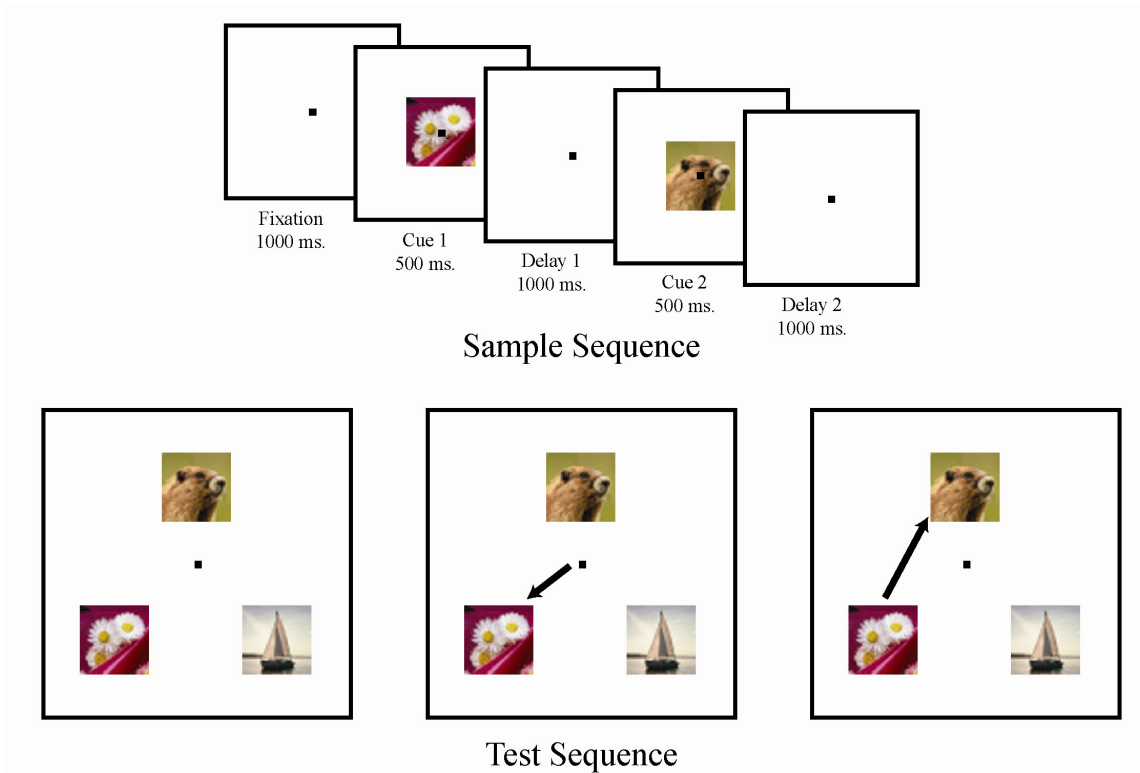


FIGURE 3.4

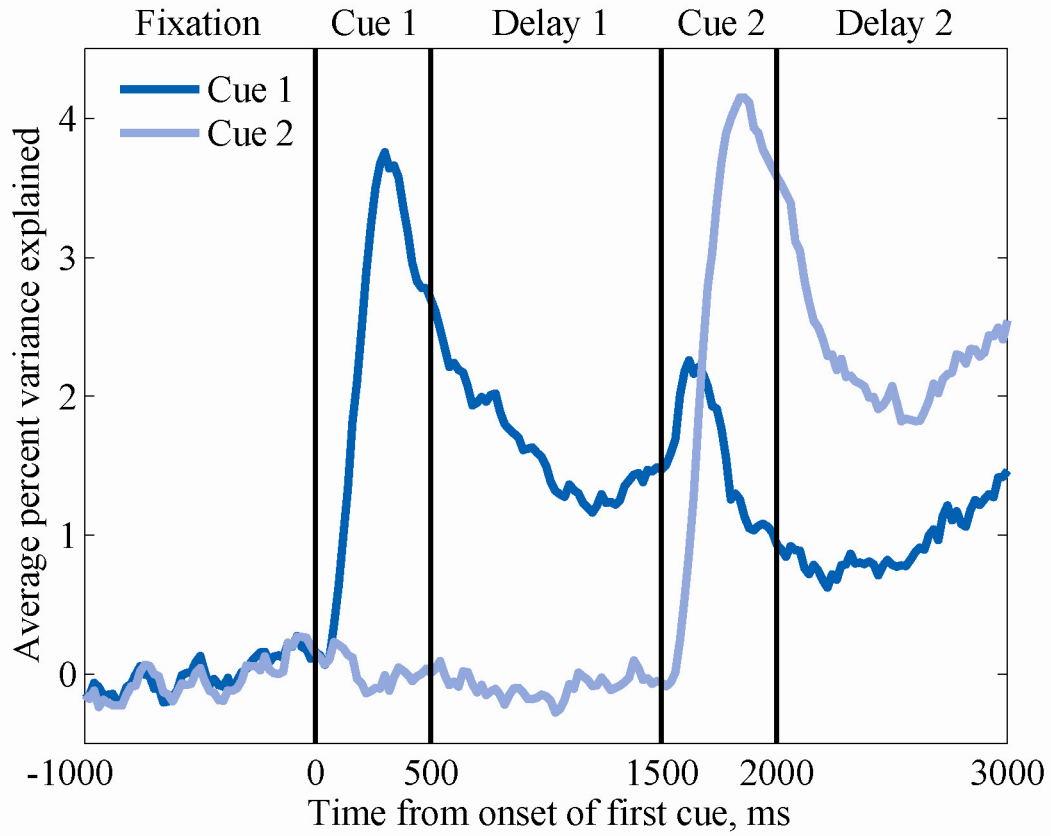


FIGURE 3.5

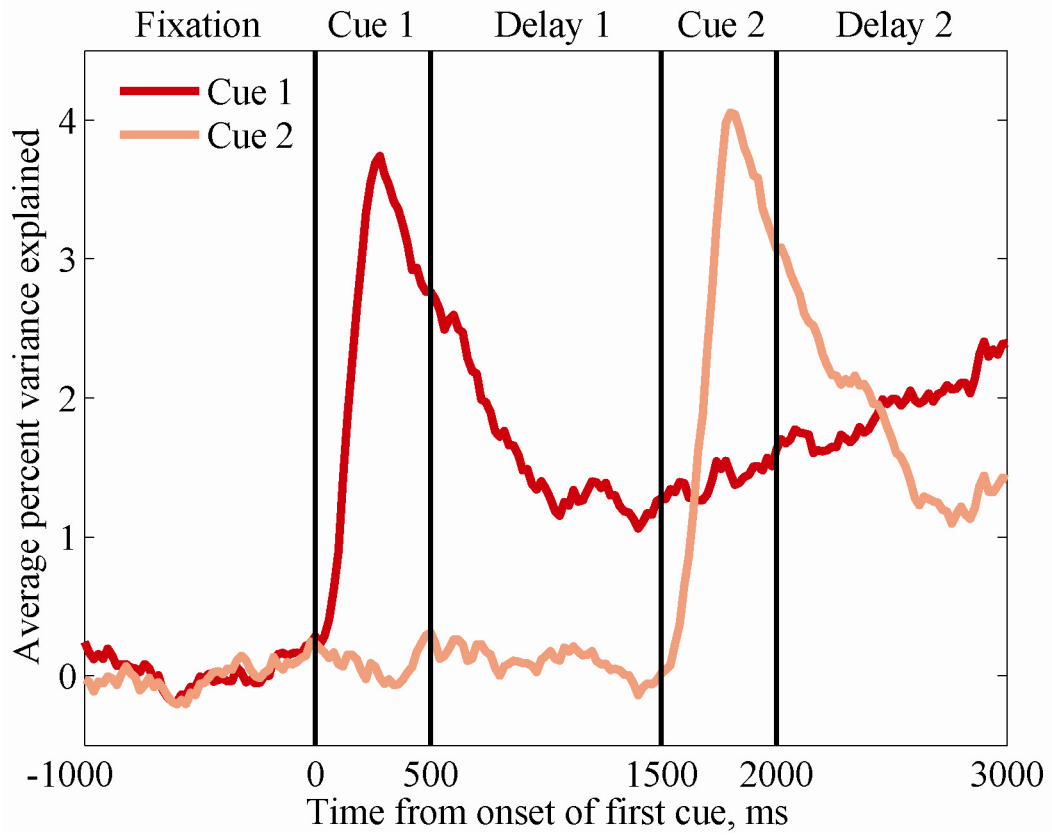


FIGURE 3.6

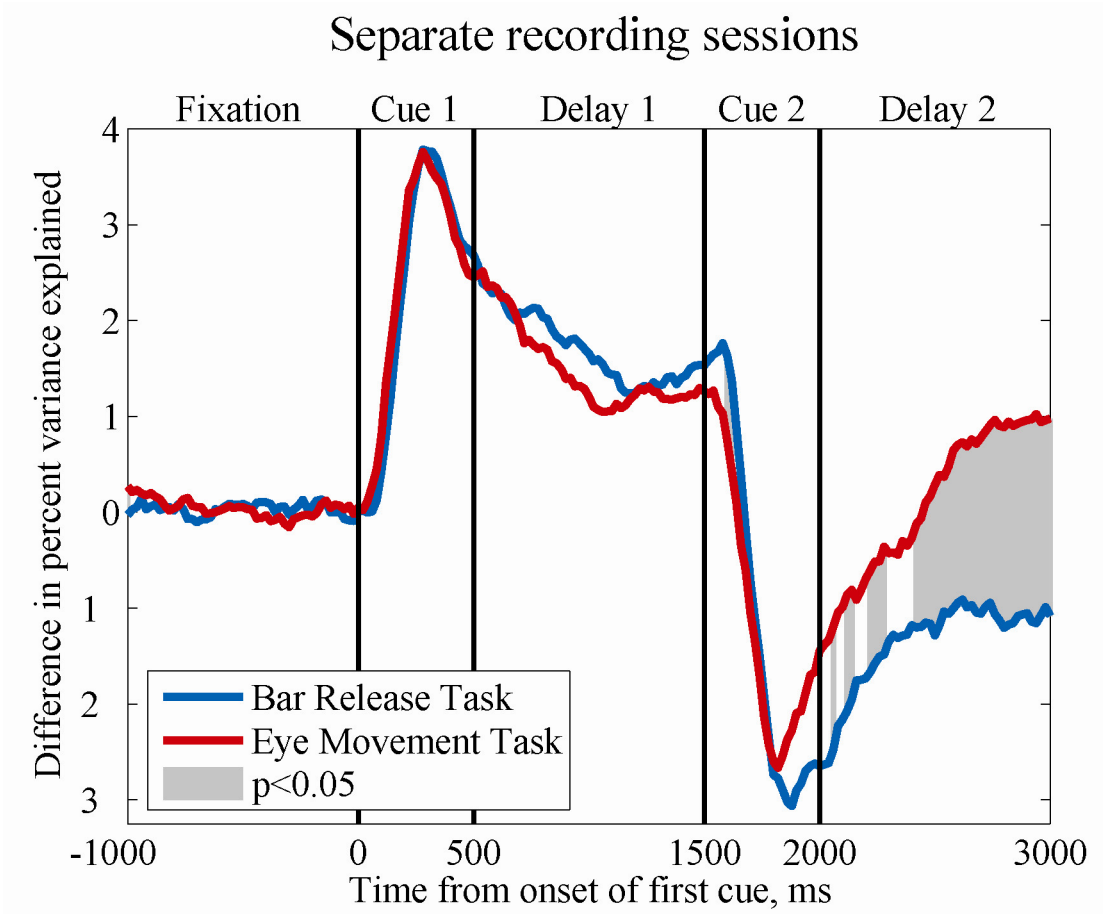


FIGURE 3.7

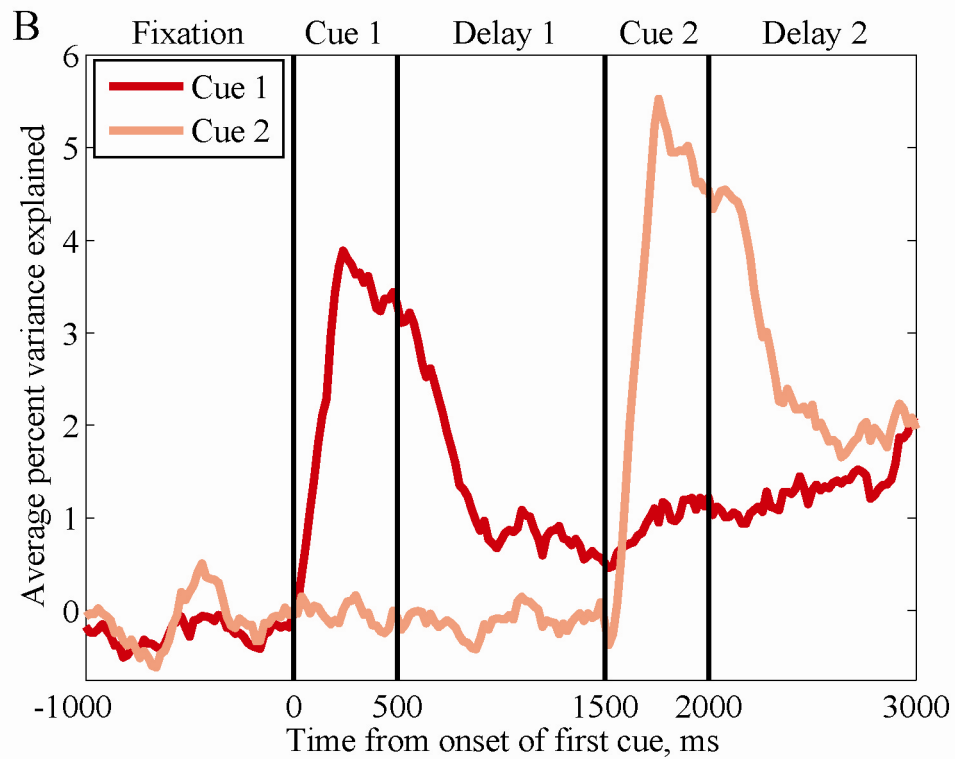
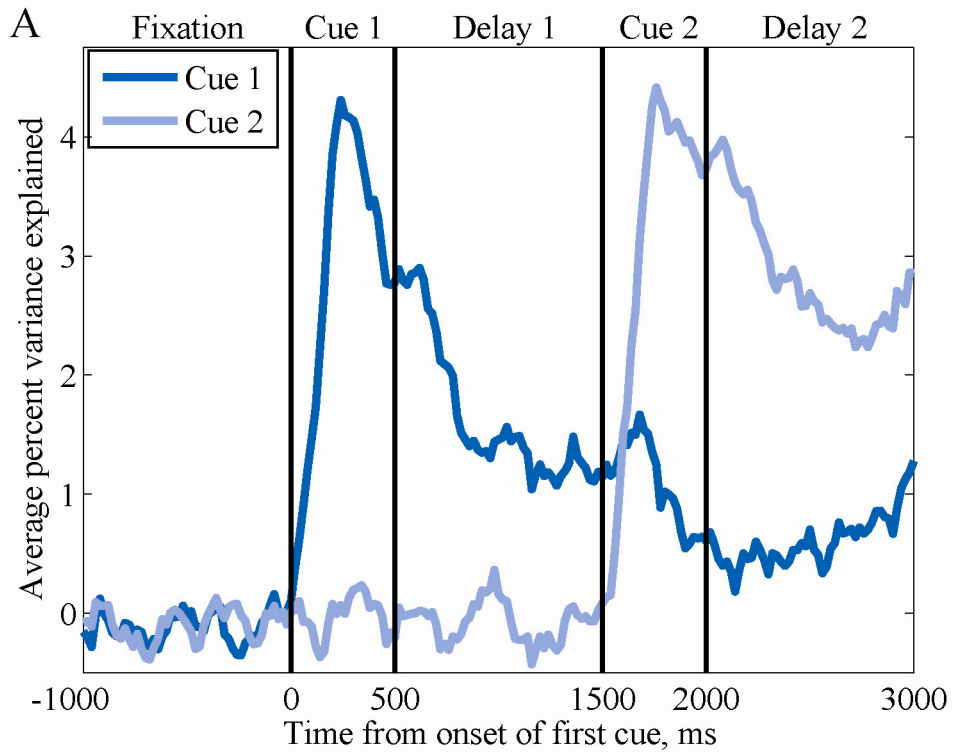


FIGURE 3.8

Interleaved blocks within same recording session

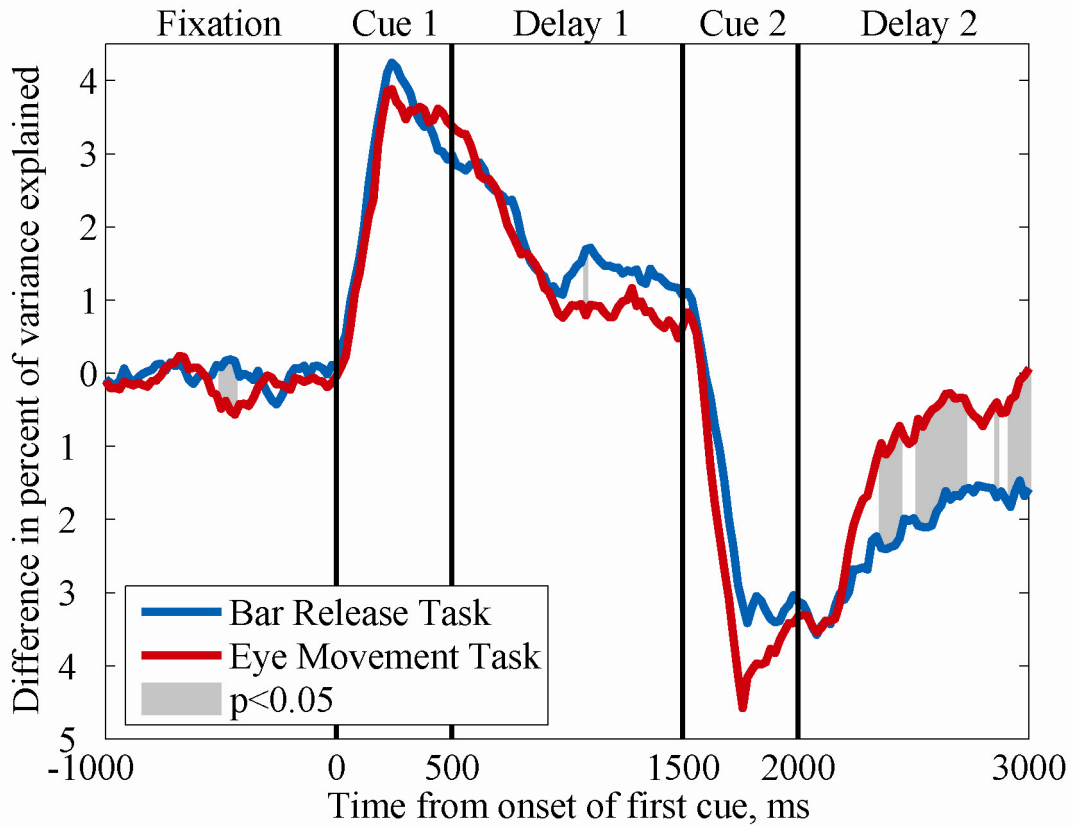


FIGURE 3.9

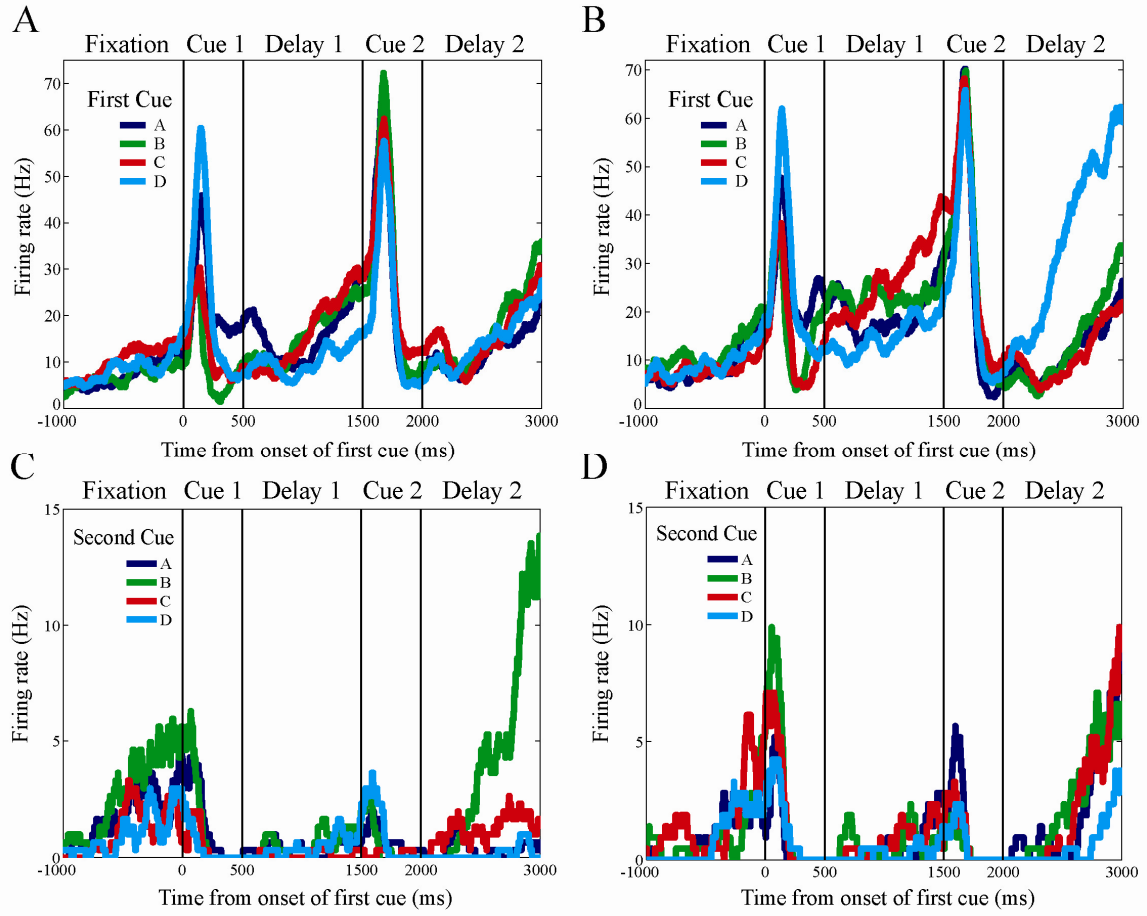


FIGURE 3.10

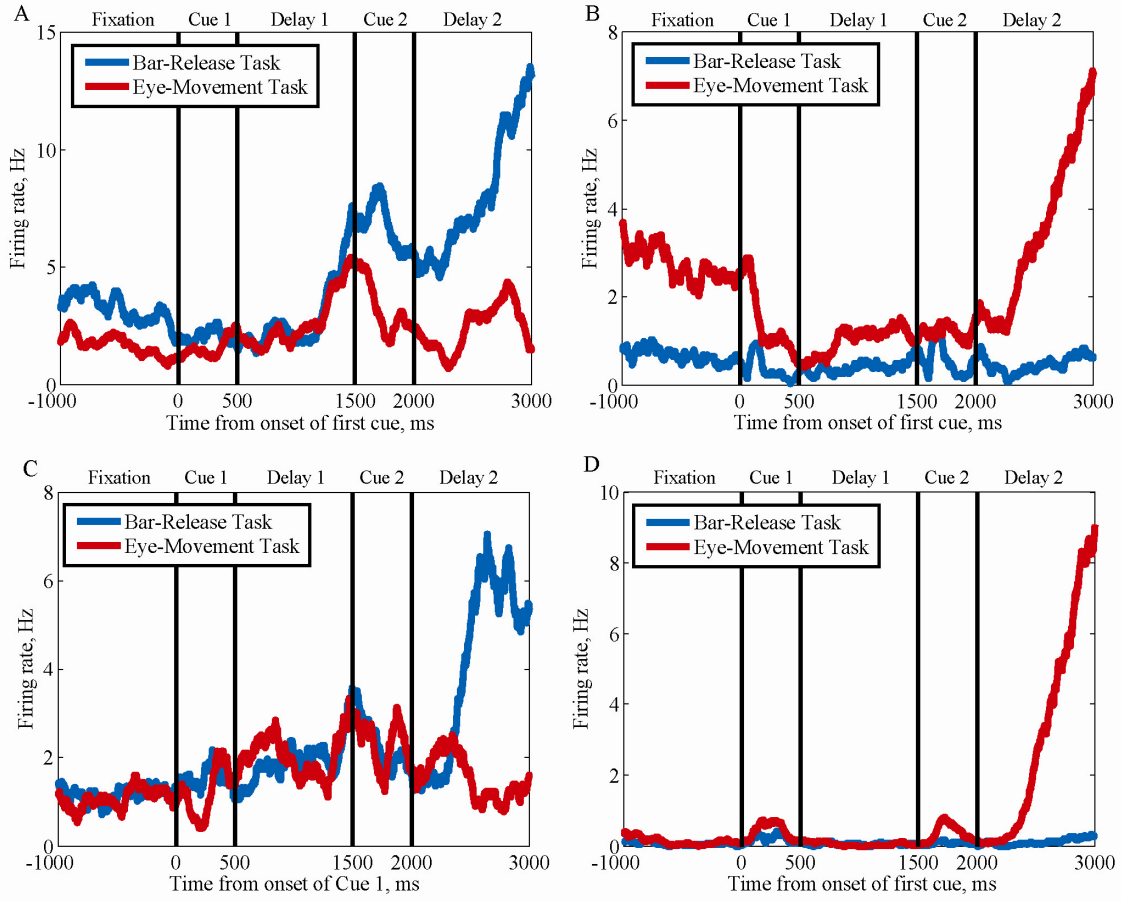


FIGURE 3.11

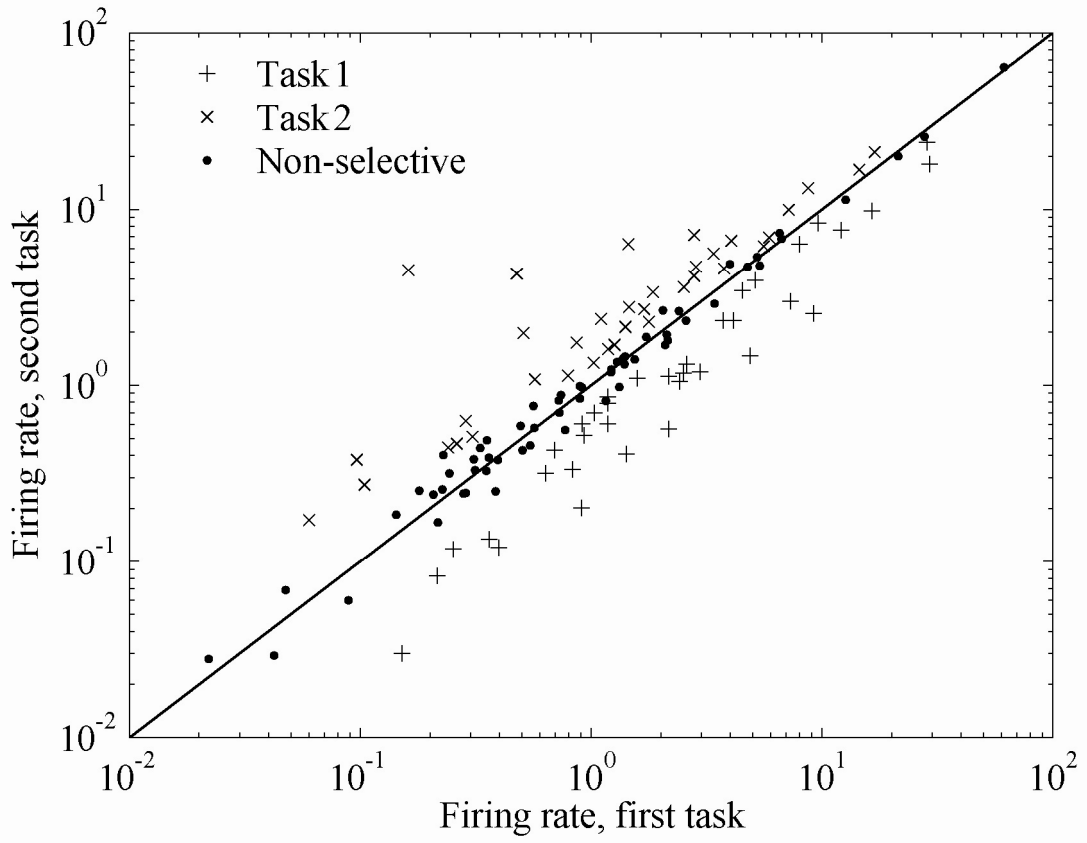


FIGURE 3.12

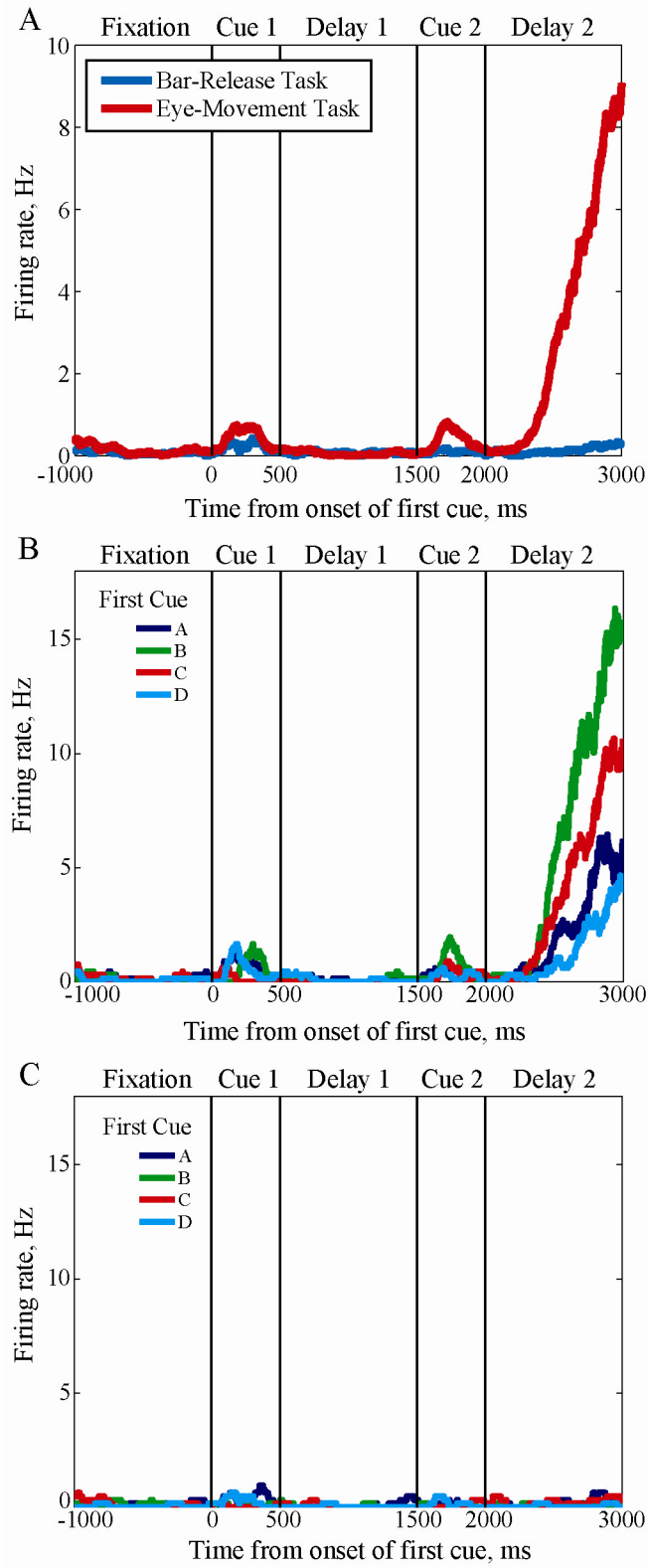


TABLE 3.1

TABLE 1. *Percentage of neurons that were selective for the first or second object that was presented, during either the first cue, the first delay, the second cue, or the second delay epoch.*

	Epoch				
	First Cue	First Delay	Second Cue	Second Delay	Any
First object selectivity					
Neurons showing effect	76	82	81	95	144
Percentage of 222	34.2	36.9	36.5	42.8	64.9
Second object selectivity					
Neurons showing effect			96	115	135
Percentage of 222			43.2	51.8	60.8
Selectivity for either object					
Neurons showing effect			107	132	163
Percentage of 222			48.2	59.5	73.4
Selectivity for both objects					
Neurons showing effect			70	78	116
Percentage of 222			31.5	35.1	52.3
First object X Second object					
Neurons showing effect			60	63	95
Percentage of 222			27.0	28.4	42.8

REFERENCES

- Asaad WF, Rainer G, Miller EK (1998) Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21:1399-1407.
- Asaad WF, Rainer G, Miller EK (2000) Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 84:451-459.
- Baddeley AD (1986) *Working Memory*. New York: Oxford University Press.
- Bugmann G, Bapi RS (2000) Modelling relative recency discrimination tasks using a stochastic working memory model. *Biosystems* 58:195-202.
- Dias R, Robbins TW, Roberts AC (1997) Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin Card Sort Test: restriction to novel situations and independence from "on-line" processing. *J Neurosci* 17:9285-9297.
- Fortin NJ, Agster KL, Eichenbaum HB (2002) Critical role of the hippocampus in memory for sequences of events. *Nat Neurosci* 5:458-462.
- Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. *Trends Neurosci* 15:20-25.
- Goodale MA, Humphrey GK (1998) The objects of action and perception. *Cognition* 67:181-207.
- Hinrichs JA (1970) A two-process memory-strength theory for judgment of recency. *Psychological Review* 77:223-233.
- Konorski J (1961) The physiological approach to the problem of recent memory. In: *Brain Mechanisms and Learning* (Delafresnaye JF, ed). London: Oxford University Press.
- Lie CH, Specht K, Marshall JC, Fink GR (2006) Using fMRI to decompose the neural processes underlying the Wisconsin Card Sorting Test. *Neuroimage*.

- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167-202.
- Miller EK, Li L, Desimone R (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J Neurosci* 13:1460-1478.
- Milner AD, Perrett DI, Johnston RS, Benson PJ, Jordan TR, Heeley DW, Bettucci D, Mortara F, Mutani R, Terazzi E, et al. (1991) Perception and action in 'visual form agnosia'. *Brain* 114 (Pt 1B):405-428.
- Milner B (1963) Effects of different brain lesions on card sorting. *Arch Neurol* 9:90-100.
- Morton J (1968) Repeated items and decay in memory. *Psychonomic Science* 10:219-220.
- Nichols AM, Ruffner TW, Sommer MA, Wurtz RH (1998) A screw microdrive for adjustable chronic unit recording in monkeys. *J Neurosci Methods* 81:185-188.
- Perenin MT, Vighetto A (1988) Optic ataxia: a specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. *Brain* 111 (Pt 3):643-674.
- Rainer G, Asaad WF, Miller EK (1998) Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* 393:577-579.
- Rao SC, Rainer G, Miller EK (1997) Integration of what and where in the primate prefrontal cortex. *Science* 276:821-824.
- Repos G, Baddeley A (2006) The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience* 139:5-21.
- Sokal RR, Rohlf FJ (1995) *Biometry, Third Edition*. New York: W.H. Freeman and Company.
- Ungerleider LG, Mishkin M (1982) Two Cortical Visual Systems. In: *Analysis of Visual Behavior* (Ingle DJ, Goodale MA, Mansfield RJW, eds). Cambridge, MA: MIT Press.

Wallis JD, Miller EK (2003) From rule to response: neuronal processes in the premotor and prefrontal cortex. *J Neurophysiol* 90:1790-1806.

Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953-956.

White IM, Wise SP (1999) Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res* 126:315-335.

CHAPTER 4

Conclusions

CONCLUSIONS

The representation of multiple items in memory

Through the course of these studies, we have taken a comprehensive look at the representation of multiple items in short term memory. We began with the basic question of how a single brain region, the prefrontal cortex, might represent more than one item at a time. First, we established that single neurons in the prefrontal cortex are capable of simultaneously representing multiple items. When a monkey is holding two items in memory, the delay period activity of a single cell is modulated by both of these items. Most of these cells also show selective activity when the monkey is only holding a single item in memory; the second item, when it is seen and added to memory, modulates the cell's activity to accommodate its representation. It seems that the brain does not represent items in separate 'boxes' in memory, analogous to an address in computer memory; rather, a distributed network of cells is responsible for maintaining multiple items, with individual cells participating in the maintenance of more than one item.

We have also found that when a cell represents more than one item in memory, the addition of a second item to memory profoundly changes the representation of the first item. The representation of the newest item in memory is very similar to the representation of an item in isolation, while the representations of older items in memory are very different.

Our results are complimentary to those obtained by groups that have concentrated on the representation of whole sequences of items in memory. There have been several reports of cells that represent the memory of sequences of movements, spatial positions,

and objects (Shima and Tanji, 2000; Averbeck et al., 2003; Ninokura et al., 2003; Averbeck et al., 2006). These results are very interesting, because they show the result of a computation that is probably designed to facilitate storage of temporal epochs in working memory. However, the question remains of how it is possible to create these cells from cells that are selective for single images. We are reporting here for the first time the existence of a population of cells that is capable of encoding both individual items and multiple items, which may be an intermediary stage in this computation.

Task-dependent representation of multiple items

The results of our second and third sets of experiments demonstrated that the way in which a monkey reported the memory of a sequence of items had a strong impact on the way that the items were represented in memory. If the monkey was using a bar release to indicate its memory, each item in memory was represented by a trace that decayed in strength as time progressed. However, if the monkey was using a sequence of eye movements to indicate its memory, the strength of each item did not decay with time. Instead, the representation of each item was maintained at a roughly equivalent level. If the monkey had sufficient time to settle into a new strategy for the eye movement task, the pattern of results even reversed: there was a stronger representation of the older item. When we had a monkey perform both of these tasks on the same day, we were able to record the activity of individual neurons that were consistent with the population results. We found, for example, individual neurons that did not represent the first item in the bar-release task, but did represent it in the eye-movement task. We also found neurons that represented the second item in the bar-release task, but that did not represent it in the eye-

movement task. Consistent with previous reports (White and Wise, 1999; Asaad et al., 2000; Wallis et al., 2001; Wallis and Miller, 2003), we also found a large population of cells that represented which task the monkey was performing. We found that a large fraction of these task-selective cells were modulated by the images that the monkey was holding in memory.

These results are interesting because they indicate that the various components of working memory may not be as separate as they appear to be at first. Working memory assumes the existence of a central executive that is separate from the short term storage buffers, the visuospatial sketch pad and the phonological loop. We have found, however, that how an animal is going to use a piece of information can have a profound impact on its neural representation. We also found cells that simultaneously represented task and item identity. We interpret these results to mean that the central executive and the storage buffers are not separate, and in fact probably involve the same networks of cells.

Future directions

There are several ways that these experiments could be extended to shed more light on the issues that we have discussed here. First, it would be interesting to repeat the first experiment with a broader sampling of possible objects. We would obtain more insight from the response surface graphs if they were created with more points in order to more fully sample the representation space.

Also, it would be interesting to train animals to remember more than two items, although this may not be possible. There were several interesting effects that were not discussed here because we had only two data points: one or two items in memory. If we

had three or more, we could begin to say something interesting about how memory capacity is represented at the single neuron level. As it stands, the trends that we found but did not report could be attributed to other causes.

One future direction that we intend to pursue is to look for interesting multi-item memory and/or temporal order signals in other brain areas. In particular, we would like our next target to be the hippocampus. This area is a particularly interesting one because many decades of study have shown that its cells are specialized for encoding relationships between stimuli. This is true for spatial stimuli (the hippocampus as cognitive map) as well as for other stimuli that are not as naturally associated, such as odors or visual stimuli (Nadel, 1991; Dusek and Eichenbaum, 1997; Wirth et al., 2003). The hippocampus is also interesting because of a hypothesis that the signal for temporal order might be found there. John Lisman and colleagues have proposed a model that suggests that the theta and gamma rhythms found in the hippocampus might be the key to understanding how we remember the relative timing of events (Jensen and Lisman, 1996). There are roughly 7 gamma cycles per theta cycle, a number that is spookily reminiscent of the idea that we can hold 7 ± 2 items in working memory. The theory contends that one item can be represented in each gamma cycle, for a total of 7 memories per theta cycle. This theory has not yet been directly verified, so finding support for this hypothesis in monkey hippocampus would revolutionize our understanding of how the brain represents temporal order.

Conclusions

In conclusion, we have found that a single network of cells represents multiple items in the primate prefrontal cortex, and it does so in a way that is not easily predictable from the response to individual items in isolation. Exactly how these items are stored is highly dependent on context, and we have found single cells that represent both the task that the animal is performing as well as the stimuli that it is keeping in mind. Executive control appears to use the same network of cells as mnemonic storage in the prefrontal cortex.

REFERENCES

- Asaad WF, Rainer G, Miller EK (2000) Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 84:451-459.
- Averbeck BB, Sohn JW, Lee D (2006) Activity in prefrontal cortex during dynamic selection of action sequences. *Nat Neurosci* 9:276-282.
- Averbeck BB, Chafee MV, Crowe DA, Georgopoulos AP (2003) Neural activity in prefrontal cortex during copying geometrical shapes. I. Single cells encode shape, sequence, and metric parameters. *Exp Brain Res* 150:127-141.
- Dusek JA, Eichenbaum H (1997) The hippocampus and memory for orderly stimulus relations. *Proc Natl Acad Sci U S A* 94:7109-7114.
- Jensen O, Lisman JE (1996) Novel lists of 7 +/- 2 known items can be reliably stored in an oscillatory short-term memory network: interaction with long-term memory. *Learn Mem* 3:257-263.
- Nadel L (1991) The hippocampus and space revisited. *Hippocampus* 1:221-229.
- Ninokura Y, Mushiake H, Tanji J (2003) Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *J Neurophysiol* 89:2868-2873.
- Shima K, Tanji J (2000) Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *J Neurophysiol* 84:2148-2160.
- Wallis JD, Miller EK (2003) From rule to response: neuronal processes in the premotor and prefrontal cortex. *J Neurophysiol* 90:1790-1806.
- Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953-956.
- White IM, Wise SP (1999) Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res* 126:315-335.

Wirth S, Yanike M, Frank LM, Smith AC, Brown EN, Suzuki WA (2003) Single neurons in the monkey hippocampus and learning of new associations. *Science* 300:1578-1581.